Title
The Spatial Segregation of Ethnic and Demographic Groups: Comparative Evidence from Stockholm and San Francisco

Authors
Harsman, Bjorn
Quigley, John M.

Publication Date
1992-05-01

Peer reviewed

Institute of
Business and
Economic Research

University of
California at
Berkeley

CENTER FOR REAL ESTATE AND URBAN ECONOMICS

WORKING PAPER SERIES

WORKING PAPER NO. 92-209

THE SPATIAL SEGREGATION OF ETHNIC AND DEMOGRAPHIC GROUPS: COMPARATIVE EVIDENCE FROM STOCKHOLM AND SAN FRANCISCO

By

BJORN HARSMAN
JOHN M. QUIGLEY

WALTER A. HAAS SCHOOL OF BUSINESS

CENTER FOR REAL ESTATE AND URBAN ECONOMICS
UNIVERSITY OF CALIFORNIA AT BERKELEY
Kenneth Rosen, Chair

The Center was established in 1950 to examine in depth a series of major changes and issues involving urban land and real estate markets. The Center is supported by both private contributions from industry sources and by appropriations allocated from the Real Estate Education and Research Fund of the State of California.

INSTITUTE OF BUSINESS AND ECONOMIC RESEARCH
Richard Sutch, Director

The Institute of Business and Economic Research is an organized research unit of the University of California at Berkeley. It exists to promote research in business and economics by University faculty. These working papers are issued to disseminate research results to other scholars. The authors welcome comments; inquiries may be directed to the author in care of the Center.

THE SPATIAL SEGREGATION OF ETHNIC AND DEMOGRAPHIC GROUPS: COMPARATIVE EVIDENCE FROM STOCKHOLM AND SAN FRANCISCO*

by

Bjorn Harsman
Royal Institute of Technology
Stockholm

and

John M. Quigley
University of California
Berkeley

May 1992

WORKING PAPER NO. 92-209

* A preliminary and less complete version of this paper was presented at the Sodertorn Summer Institute, Huddinge, Sweden in August 1990 and at the European meetings of the RSA in August 1990. We are grateful to Lata Chatterjee and Folke Snickars for comments and criticism. Computational assistance was provided by Roger Bernow and Scott Hacker for computational assistance. Quigley's research was supported by the Center for Real Estate and Urban Economics and by the Transportation Center, University of California, Berkeley.

Abstract

This paper compares the level of spatial segregation by race or ethnicity with the level of spatial segregation by demographic group in two metropolitan areas with similar incomes and demographic composition, but with very different racial make up. We compare census tract data for the San Francisco Bay Area for 1980, a region with six large ethnic divisions, with similar data for the Stockholm metropolitan area, a region with a much more homogeneous racial composition.

An extensive comparison of entropy measures of segregation in the two regions is presented, including for Stockholm, an analysis of spatial segregation by income class. One important finding of the analysis, replicated in two very different metropolitan regions, is that spatial segregation by race or ethnicity is unrelated to the principal economic factors which presumably underly spatial segregation by income class or demographic grouping.

I. INTRODUCTION

Even the most casual observer notices that residential patterns in American urban areas are highly segregated by race. It is only slightly less obvious that urban areas throughout the developed world are segregated by income, by household size and composition, and by other demographic characteristics. Presumably, residential segregation by sociodemographic group reflects similarity of tastes for local public goods and locational amenities and similarity in disposable income. Residential segregation by race and ethnic group may reflect the same phenomenon. It may also reflect the outcomes of a discriminatory market in which minority households have less access to the entire housing stock or in which minority households feel less threatened by choosing to reside in close proximity to one another.

Disentangling "natural" segregation by sociodemographic group from that which arises from prejudice is no easy task. Yet the distinction is important, at least in the American context, to interpreting trends in segregation. In previous work (Miller and Quigley, [1990]), we compared the pattern of spatial segregation by race and household type in 1970 and 1980 for the San Francisco Bay Area, concluding that levels of spatial segregation by race declined slightly during the decade, and that levels of segregation by household type declined more substantially. That work also indicated that only a small fraction of segregation by race could be "explained" by the prior segregation of households by household type. The socioeconomic forces which led to spatial clustering of different types of households "explained" practically none of the spatial segregation of races in the San Francisco Bay Area in 1970 or in 1980.

This paper provides a more abstract benchmark for assessing these conclusions by presenting a similar analysis of spatial segregation by sociodemographic group over time in a racially and culturally homogeneous society. The analysis concentrates on residential patterns in Stockholm, as reported in special census tabulations for 1975 and 1985. To facilitate comparisons with previous work, we also use an entropy index to measure segregation. However, we derive the measure in an alternate way that permits a more general comparative treatment of segregation.

We investigate the level of spatial segregation by type of household, by income, and by ethnicity using an identical methodology and consistent definitions for 1975 and 1985. We also compare these results to those obtained for San Francisco in 1980 and which are based on almost identical definitions of household type.

In many ways San Francisco and Stockholm exhibit a similar pattern of spatial and demographic development (See Harsman and Quigley [1991] for a more detailed discussion).

Both regions have a well defined central core, and both regions have high average incomes, with considerable growth in nonmanufacturing employment. A principal difference is the ethnic makeup of populations. San Francisco has large and growing populations of hispanic, black, and Asian households. Although Stockholm does show an increase in the fraction of non-Swedes and non-European households, it is from a very small base. By any international standard, Stockholm is ethnically homogeneous.

II. SEGREGATION

Consider households' choices to reside at various points in an urban area, where conflicting choices are resolved by some impersonal mechanism. Now if the conditional distribution of households by demographic group in space differs from the unconditional distribution, the population may be said to be segregated. Of course, if conflicting choices are resolved by a price mechanism, then differences in income or wealth among groups may lead to this segregation. Similarly, if prices are regulated and households are assigned to dwellings according to waiting time in a public queue, differences in age or recency of migration may be reflected in segregation. (To a first approximation the price mechanism is used to ration access in San Francisco, and the public queue is used to ration access in Stockholm.)

Also, systematic differences in those factors which affect preferences for location, for example family size and household composition, can similarly give rise to residential segregation by demographic group.

More generally, the problem of defining and measuring segregation can be formulated as follows. Let

(1) $p = (p_1, p_2, \ldots p_n)$

be a vector describing an observed distribution over subareas of a given household or ethnic group, $\Sigma_i p_i = 1$. let

(2) $q = (q_1, q_2, \ldots, q_n)$

be a vector describing the distribution of all households over the same subareas, $\Sigma_i q_i = 1$. Define a measure of segregation $I(p,q)$ that indicates how the conditional distribution, p deviates from the "reference" distribution, q,

(3) $$\begin{aligned} I(p,q) &= \Sigma_i\, p_i \log(p_i/q_i) \\ &= \Sigma_i\, p_i \log(1/q_i) - \Sigma_i\, p_i \log(1/p_i). \end{aligned}$$

This measure is a weighted sum of the logarithmic difference between the two distributions in each of the subareas. The fractions given by the conditional distribution are used as weights. When the conditional distribution p equals the reference distribution q, there is no segregation, and $I(p,q) = 0$.

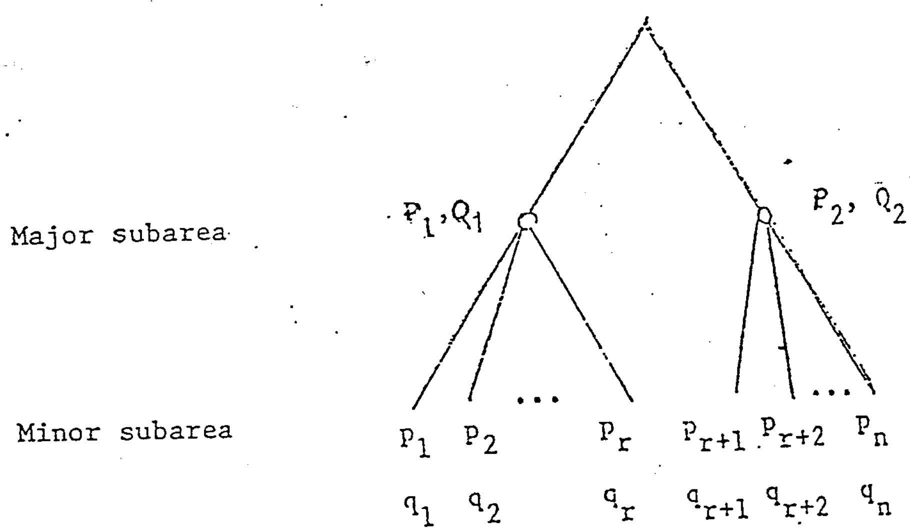
The second term in the expression on the far right is the entropy of the conditional distribution, and the first term is the entropy of the reference distribution, using the weights given by the conditional distribution.

This definition of I(p,q) was used by Shannon [1948] to measure the information obtained when it is first assumed that q is the distribution characterizing a certain phenomenon and subsequently it is learned that the phenomenon is characterized by the distribution p. The analogy to segregation is straightforward; we first assume that the spatial distribution of a certain household type is the same as the distribution of other households over subareas, q. Then we learn that the conditional distribution is p. The measure of segregation I(p,q) can be interpreted as value of the information gained when the level of segregation is measured directly.

Hobson [1969] has shown that this definition of I(p,q) is unique up to a positive constant under five general conditions. The first three state that I(p,q) should be a continuous function of all its variables, that it should be independent of the order in which the states, or subareas, are numbered, and that I(q,q) = 0. According to the fourth condition, I(p,q) should increase when the shape of p tends to be more peaked or the shape of q tends to be less peaked. The last condition states that I(p,q) should be independent of the number of states or subareas in the following sense. Assume that the r first and the n-r last states are aggregated according to Figure 1.

Figure 1

Hierarchy of Subareas



With this notation, the condition can be stated as:

$$(4)\quad I(p_1,\ldots,p_n;q_1,\ldots q_n) = I(P_1,P_2;Q_1,Q_2) + P_1 I(p_1/P_1,\ldots,p_r/P_1;q_1/Q_1,\ldots,q_r/Q_1) + P_2 I(p_{r+1}/P_2,\ldots,p_n/P_2;q_{r+1}/Q_2,\ldots,q_n/Q_2) \quad .$$

Thus, in general, segregation can be measured by partitioning space into larger and smaller subareas. $I(p,q)$ is the weighted sum of the segregation measures calculated within each subarea, and the weights are the population fractions in each of the subareas.

This definition implies that the presence of segregation can be tested by partitioning the area geographically and by examining the demographic composition of each subarea. Two problems arise with this approach to analyzing either the presence or the degree of segregation. One is that, as the sizes of the subareas increase, the same physical area appears less segregated. In the limit, when the subarea subsumes everything, the metropolitan area must be "integrated." A second is that the way in which the area is partitioned can affect conclusions about the presence of segregation or the degree of segregation. For example, a checker-board pattern of residential occupancy by subarea can give rise to extreme differences in residential occupancy by subarea or to identical measures of the racial composition of subareas, depending only upon how the checker-board is partitioned.

Despite the potential importance of this partitioning problem, any empirical analysis of patterns of residential segregation must ultimately begin with counts of individuals or households by predetermined geographical areas: census tracts or perhaps block faces. For both the U.S. and for Sweden, census tracts were established to have stable boundaries, and were designed to be relatively homogeneous areas with respect to population characteristics, economic status and living conditions. Census tracts for these two regions are of roughly the same size. Any measure of segregation is conditional upon the prior partitioning of the urban area into these geographical subareas.

Let n_{it} be the number of individuals of group i residing in census tract t. Thus $n_{*t} = \Sigma_t n_{it}$ is the total number of individuals residing in t, and $n_{i*} = \Sigma_t n_{it}$ is the total number of individuals of type i in the entire area. Similarly, $n_{**} = \Sigma_i n_{i*} = \Sigma_t n_{*t}$ is total population. Let $\rho_{it} = n_{it}/n_{*t}$ be the number households of category i as a fraction of all households in census tract t, $\omega_t = n_{*t}/n_{**}$ be the number of households in census tract t as a fraction of all households, and $\omega_i = n_{i*}/n_{**}$ be the number of households of type i as a fraction of all households.

Comparing the demographic composition in tract t with that in the whole region, the index of segregation can be computed as

$$(5) \qquad I_t(\rho_{it}, \omega_i) = \Sigma_i \, \rho_{it} \log(\rho_{it}/\omega_i) \quad .$$

By summing I_t over all subareas, using the population fractions, ω_t, as weights we get the following measure for the level of segregation in the region:

$$(6) \qquad I = \Sigma_t \, \omega_t \, I_t \quad .$$

As noted below, these measures of segregation are equivalent to those defined by Theil [1972] and used by Miller and Quigley [1990]. In comparison with other commonly used measures of segregation, the entropy index is the only one which satisfies properties of symmetry, continuity and full additivity (See Allison [1978], or White [1983]. See also Schnare [1980]).[1]

III. THE DATA

As noted above, this analysis of spatial segregation is based upon data from the San Francisco Bay Area (The "San Francisco-San Jose-Oakland Consolidated Metropolitan Statistical Area") which includes nine counties and five Metropolitan Statistical Areas (MSA's) and the Stockholm Metropolitan Area (Stockholm County) which includes the

---

1 One difficulty with this index remains. The entropy index does not overcome the ambiguity arising from the arbitrary way that a region is partioned into subareas. In principle, this could be addressed by considering the demographic composition in all surrounding tracts when judging the extent to which a given tract is segregated. For example, define a non-increasing distance function $f(d_{tk})$ which assumes the value 1 when the distance between t and k, $d_{tk} = 0$, and approaches 0 when d_{tk} approaches infinity. Now compute the accessibility, m_{ik} to households of category i in subarea t by using the distance function as a weighting factor (Weibull [1976] presents a thorough discussion of the theoretical properties of this distance or accessibility weighted measure).

$$(N-1) \quad m_{it} = \Sigma_k f(d_{tk}) n_{it}$$

Weighting by (m_{it} / m_{i*}) in equation (5) instead of (n_{it} / n_{i*}) would produce a segregation index robust to the partioning of the region into subareas. Lack of data precludes utilizing measures of this type in this analysis.

central city, an inner ring, and the suburban fringe. The San Francisco analysis is based upon census tract data for 1980, consisting of 1079 census tracts. The Stockholm analysis is based upon 806 census tracts defined identically for 1975 and 1985. Figures 2 and 3 present, in schematic terms, the two metropolitan regions.

Table 1 summarizes the demographic groupings available for San Francisco in 1980. For the nine county region as a whole, some 72 percent of the population is classified as white, 9 percent is Hispanic, 7 percent black and 6 percent is Asian.

The classification of the population into household types is straightforward.[2] The seven major types of household include traditional husband-wife families with and without children, single adults living alone, by sex, single parent households, by sex, and non family households containing two or more adults. Note that Asian, Hispanic, and "other" households are far more likely to involve married couples with children than is true for white, black, or native American

2 According to U.S. Census conventions, the population is counted by family and by household. Families are defined on the basis of relationships; households are defined on the basis of living quarters. Households are of two basis types. Family households include two or more related persons living together. Non-family households are persons living alone or sharing living quarters with persons to whom they are not related.

FIGURE 2

The San Francisco Bay Area

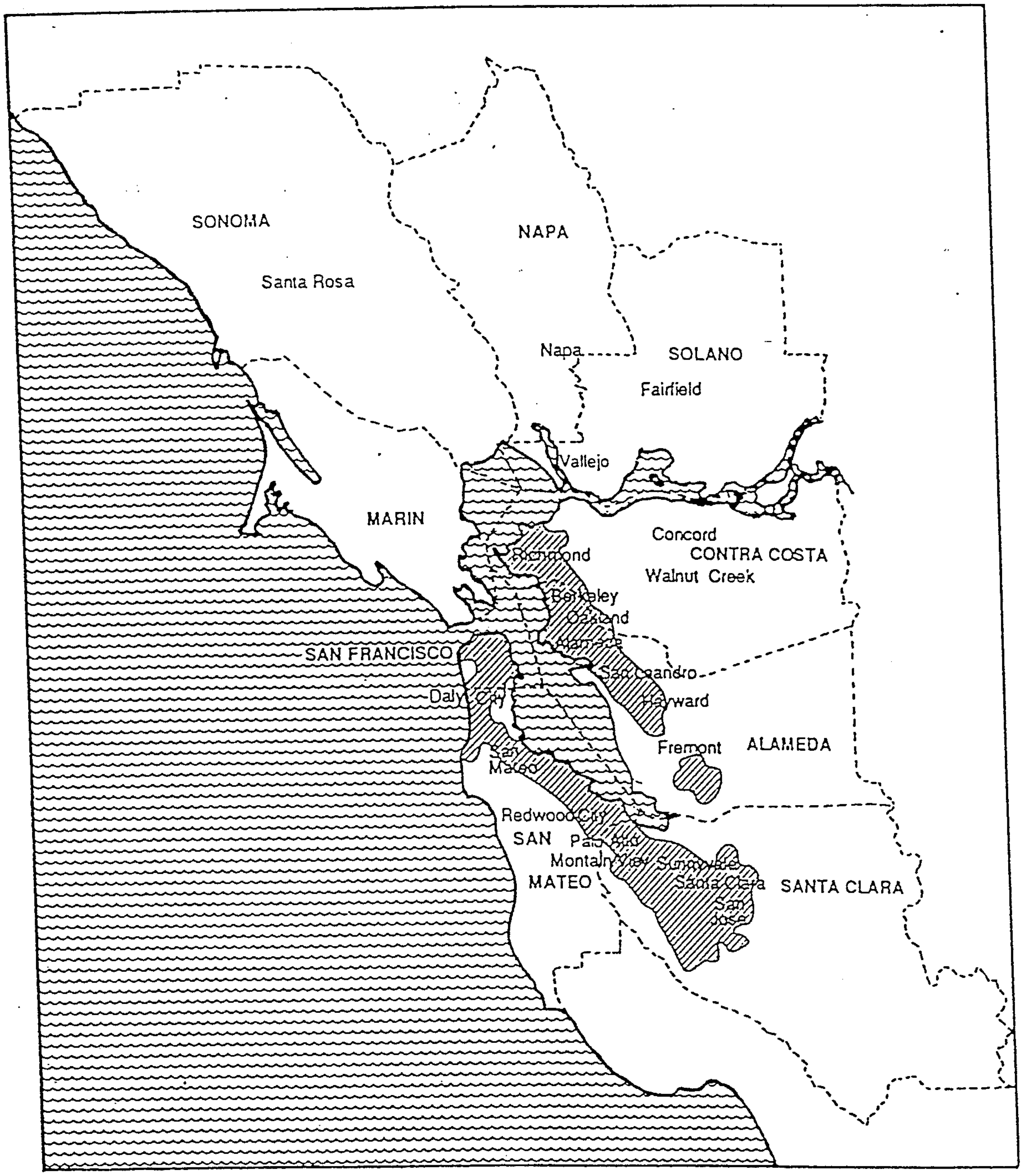


FIGURE 3

The Stockholm Metropolitan Area

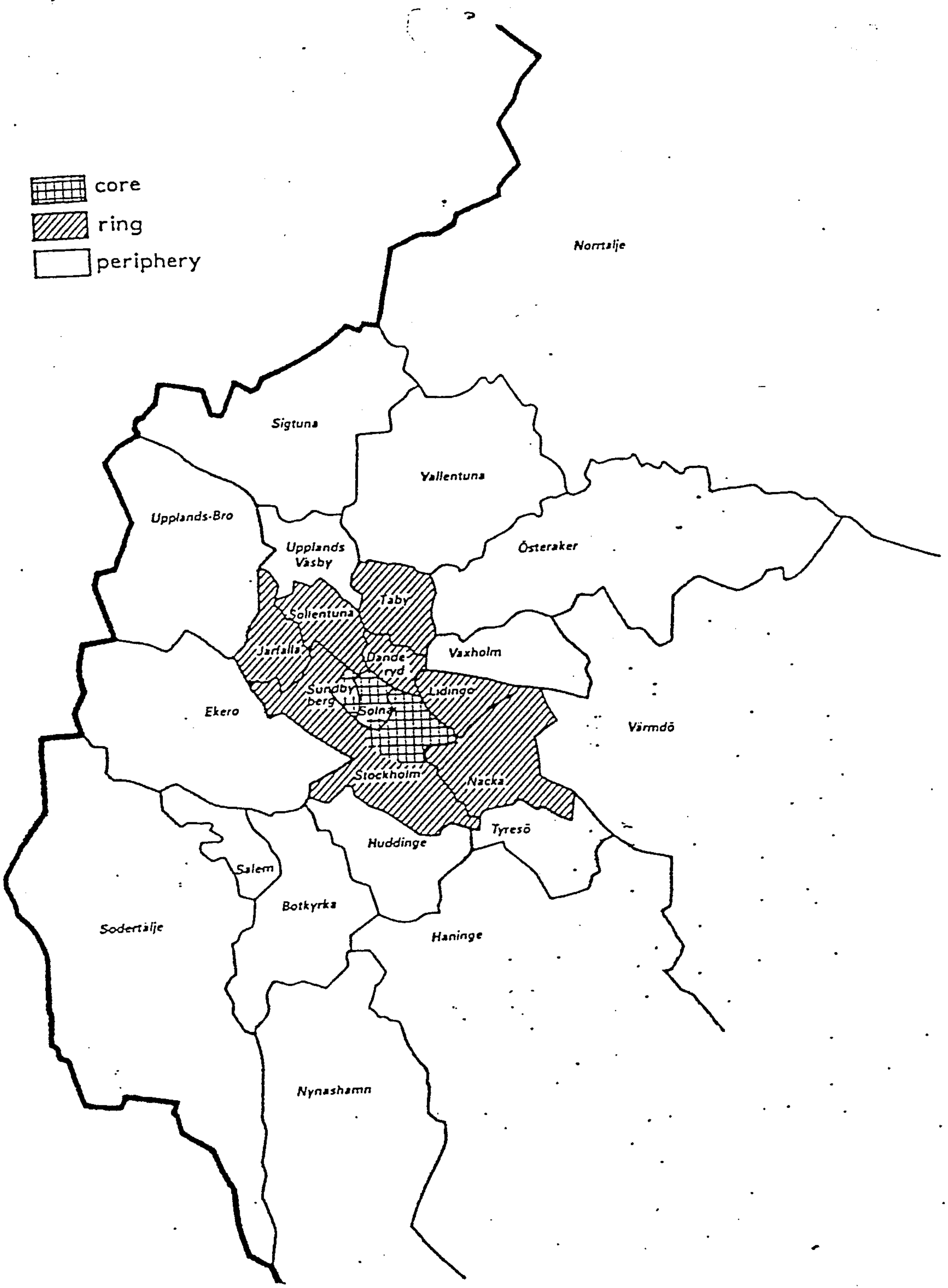


TABLE 1

Household Type by Ethnicity
San Francisco Bay Area, 1980

Household Type	White	Black	Native	Asian	Hispanic	Other	Total
Family							
(Married Couple)							
With Children	331,493	28,834	2,314	44,208	70,110	31,657	508,616
No Children	427,324	25,863	1,582	29,979	36,530	20,243	541,521
Male Householder							
(Unmarried)							
With Children	16,900	5,612	2,623	3,592	3,574	3,757	36,058
No Children	19,821	3,410	161	3,505	4,209	1,715	32,821
Female Householder							
(Unmarried)							
With Children	66,317	27,706	1,116	4,628	17,346	6,861	123,974
No Children	45,005	9,679	397	4,914	7,937	2,973	70,904
Non-Family	524,036	53,845	2,786	29,160	41,741	21,184	672,752
Total	1,430,895	154,949	10,979	119,986	181,447	88,390	1,986,646

Source: See Miller and Quigley [1990] for details.

Note: Male and female "household" classes may include other adults.

households.[3] Also, black households are three times more likely to be made up of an unmarried female head with children than is the case for other groups. Among households with children, 45 percent of black households are headed by single women, compared to 16 percent for all other groups. 27 percent of all the households in the Bay Area are white, non-family households. Only 22 percent of all households are white married couples with children. Married couples of all races with children account for only 27 percent of Bay Area households.

Tables 2 and 3 present comparable information for the Stockholm metropolitan area for 1975 and 1985. As far as possible, households are classified in a similar fashion. Household types include two adults with and without children (who together accounted for 47 percent of the Stockholm metropolitan area population in 1985), single men and women with children, single individuals, and a residual category "other." Ethnic information is available in three categories: Swedish (in which all adults in the household are Swedish citizens); "mixed" (in which one of the adults is a Swedish citizen), and not Swedish (in which no adult in the household is a Swedish citizen). In 1985 almost 89 percent of the population lived in households containing at least one Swedish citizen, a slight decline from 91 percent in 1975. The

3 Race is defined by the race of the "householder," generally the adult cited first by the census respondent.

TABLE 2

Household Type by Ethnicity and by Income; Ethnicity by Income Stockholm Metropolitan Area, 1975

	Ethnicity				Income		
Household Type	Swedish	Mixed	Not Swedish	Total	Low	Medium	High
Two Adults							
No Children	126,660	13,605	7,489	147,754	16,227	75,052	56,475
With Children	97,730	16,498	11,814	126,042	4,099	66,981	54,962
Single Female With Children	19,760	0	4,222	23,982	10,199	13,581	202
Single Male With Children	2,074	0	462	2,536	326	1,862	348
Single person	212,013	0	24,718	236,731	109,049	120,515	7,167
Others	92,446	7,914	12,732	113,092	5,640	41,147	66,305
Total	550,683	38,017	61,437	650,137	145,540	319,138	185,459

		Income		
Ethnicity	Total	Low	Medium	High
Swedish	550,683	128,264	268,868	153,551
Mixed	38,017	2,076	17,701	18,240
Not Swedish	61,437	15,200	32,569	13,668
Total	650,137	145,540	319,138	185,459

Source: Unpublished tabulations provided by the Central Bureau of Statistics, Stockholm.

TABLE 3

Household Type by Ethnicity and by Income; Ethnicity by Income
Stockholm Metropolitan Area, 1985

Household Type	Ethnicity: Swedish	Ethnicity: Mixed	Ethnicity: Not Swedish	Total	Income: Low	Income: Medium	Income: High
Two Adults							
No Children	164,528	25,252	15,940	205,720	9,046	64,653	102,021
With Children	96,801	22,162	17,665	136,628	5,089	54,591	76,948
Single Female With Children	20,062	0	6,344	26,406	13,829	12,233	344
Single Male With Children	2,907	0	788	3,695	704	2,587	404
Single person	263,367	0	37,373	300,740	133,922	158,560	8,258
Others	41,997	9,660	4,935	56,592	6,326	36,346	13,920
Total	589,662	57,074	83,045	729,781	168,916	358,970	201,895

Ethnicity	Total	Income: Low	Income: Medium	Income: High
Swedish	589,662	138,781	290,799	160,082
Mixed	57,074	3,223	26,086	27,765
Not Swedish	83,045	26,912	42,085	14,048
Total	729,781	168,916	358,970	201,895

Source: Unpublished tabulations provided by the Central Bureau of Statistics, Stockholm.

Swedish data also includes a cross classification by income group, in three categories. This feature of the data is discussed in more detail below.

Altogether, the San Francisco data for 1980 includes 42 demographic categories (6 racial groups by 7 household types); the Stockholm data for 1975 and 1985 includes 54 demographic categories (6 household types by 3 ethnic groups by 3 income categories).

IV. RESIDENTIAL SEGREGATION BY DEMOGRAPHIC GROUP

Households in these two metropolitan regions are partitioned into ethnic categories and household types and into a several spatial groupings, for example central city or suburban location.

Let M_s be the set of census tracts in some spatial aggregation s, such as the central city.

Thus

(7) $$W_s = \Sigma_{t \epsilon M_s} \omega_t$$

is the number of households in aggregation s as a fraction of area total, and ρ_{is} is the proportion of households of type i in region s

(8) $$\rho_{is} = \Sigma_{t \epsilon M_s} (\omega_t / W_s) \rho_{it} \quad .$$

Following (5), the segregation of a tract t in relation to region s is

$$(9) \quad I_t = \Sigma_i \rho_{it} \log(\rho_{it} / \rho_{is})$$

The average segregation in region s equals

$$(10) \quad I_s = \Sigma_t (\omega_t/W_s) I_t$$

Inserting (10) into (11) and using (8) we get

$$
\begin{aligned}
(11) \quad I_s &= \Sigma_t (\omega_t/W_s) \Sigma_i \rho_{it} \log(\rho_{it}/\rho_{is}) \\
&= \Sigma_t \Sigma_i (\omega_t/W_s) \rho_{it} \log(1/\rho_{is}) - \Sigma_t (\omega_t/W_s) \Sigma_i \rho_{it} \log(1/\rho_{it}) \\
&= \Sigma_i \rho_{is} \log(1/\rho_{is}) \qquad - \Sigma_t \rho_{is} \Sigma_i \rho_{it} \log(1/\rho_{it}) \\
&= H_s - \overline{H}_s
\end{aligned}
$$

The first term on the last line of (11) is the entropy of the distribution over household categories at the regional level, H_s. The second term is a weighted average, $\overline{H}_s$, of the entropies of the tracts, H_t in the same region, where the population shares of the tracts are used as weights.

The value of I_s is merely the difference, $H_s - \overline{H}_s$. Since $I_s = 0$ when $\rho_{it} = \rho_{is}$, H_s is the maximum value that the average regional entropy, $\overline{H}_s$ can attain. Larger values of I_s, i.e., greater levels of segregation, are manifest as reductions in the average entropy. From the relationship between ρ_{it} and ρ_{is} in (8), it is clear that the maximum value of I_s is obtained when $\overline{H}_s$ attains its minimum value. If the number of tracts is equal to or larger than the number of household categories, $\overline{H}_s = 0$ and, equivalently, the maximum value of $I_s = H_s$. This means that $I_s/\text{Max } I_s = I_s/H_s$ is a relative measure of segregation. This measure of segregation is identical to the relative reduction of entropy measure introduced by Theil [1972] and used by Miller and Quigley [1990].

The previous discussion deals with classifications in one dimension, say racial categories. The extension to the bivariate case -- the joint distribution of ethnic category, e, and household type, h -- is straightforward.

As before,

(12) $\rho_{eht} = n_{eht}/n_{**t}$

is the number of households of group e and housing type h as a fraction of all households in census tract t. The probabilities of the two marginal distributions are

(13) $\rho_{e*t} = \Sigma_h \rho_{eht}$

$\rho_{*ht} = \Sigma_e \rho_{eht}$,

and the entropies of these distributions are

(14) $H(e)_t = \Sigma_e \rho_{e*t} \log (1/\rho_{e*t})$

$H(h)_t = \Sigma_h \rho_{*ht} \log (1/\rho_{*ht})$

$H(e,h)_t = \Sigma_e \Sigma_h \rho_{eht} \log (1/\rho_{eht})$.

$H(e|h)_t = \Sigma_h \rho_{*ht} \Sigma_e \{\rho_{eht}/\rho_{*ht}\} \log \{\rho_{*ht}/\rho_{eht}\}$.

Consider the difference between the joint distribution, ρ_{eht} and the distribution given by the product of the two marginal distributions, $[\rho_{e*t}\ \rho_{*ht}]$.

(15) $I(e,h)_t = \Sigma_e \Sigma_h \rho_{eht} \log(\rho_{eht} / [\rho_{e*t}\ \rho_{*ht}])$.

This expression can be interpreted as the information obtained if it is initially assumed that the bivariate distribution in each tract equals the product of the corresponding marginal distributions. When the latter two distributions are independent, then $\rho_{eht} = [\rho_{*ht}\rho_{e*t}]$ and $I(e,h)_t = 0$.

Another interpretation of $I(e,h)_t$ is obtained by rewriting (15). After some manipulation,

(16) $I(e,h)_t = \Sigma_h \rho_{*ht} I(\{\rho_{eht} / \rho_{*ht}\}, \rho_{e*t})$

$I(\{\rho_{eht} / \rho_{*ht}\}, \rho_{e*t})$ measures the average difference between the conditional distribution $\{\rho_{eht} / \rho_{*ht}\}$ and the marginal distribution ρ_{e*t}. By symmetry,

(17) $I(e,h)_t = \Sigma_e \rho_{e*t}\ I(\{\rho_{eht}/\rho_{e*t}\}, \rho_{*ht})$

This measure of the difference between the two distributions given by $\{\rho_{eht} / \rho_{e*t}\}$ and ρ_{*ht} can be interpreted as the average demographic segregation caused by the ethnic segregation.

Using (14) and (16), equation (15) can be further interpreted as

$$(18) \quad I(e,h)_t = H(e)_t + H(h)_t - H(e,h)_t$$
$$= H(e)_t - H(e|h)_t$$
$$= H(h)_t - H(h|e)_t$$

The segregation measure is the sum of the marginal entropies minus the joint entropy. As noted in the second and third lines, it can also be expressed as a marginal entropy minus the relevant average conditional entropy, $H(e|h)_t$ and $H(h|e)_t$ respectively.

The marginal and conditional entropies in the second and third lines of equation (18) are equal when $I(e,h)_t = 0$. Hence the marginal entropies are upper bounds of the conditional entropies. The lower bounds of the conditional entropies are equal to zero. Hence the maximum of $I(e,h)_t$ is given by $H(e)_t$ and $H(h)_t$, respectively. An index of relative segregation is obtained by dividing $I(e,h)_t$ by $H(e)_t$ and $H(h)_t$, respectively.

V. COMPARATIVE RESULTS

Table 4 compares, for 1980, the household type and racial entropy of the geographic components of the San Francisco Bay Area with the maximum entropy possible. The table gives the values of I_s normalized by H_s for the five MSA's in the San Francisco Bay area and the three regions in the Stockholm

TABLE 4

Indices of Residential Segregation by Ethnicity, Household Type and Income for Stockholm and San Francisco

	Ethnicity	Household Type	Income
San Francisco Bay Area			
Year: 1980			
Entire Region	22.43%	8.19%	NA
Central City	23.22	8.53	
Oakland	25.16	8.49	
San Jose	12.06	6.36	
Santa Rosa	8.73	2.94	
Napa	13.25	5.16	
Stockholm Metropolitan Area			
Year: 1975			
Entire Region	4.31%	9.54%	6.94%
Central City	2.00	4.68	4.18
Ring	3.46	7.14	6.62
Suburbs	5.39	5.70	5.56
Year: 1985			
Entire Region	5.80	8.58	8.21
Central City	2.37	3.53	3.10
Ring	5.34	8.07	9.95
Suburbs	6.89	5.58	7.45

Note: Table entries are $(100\ I_S)/H_S$ where H_S is the maximum entropy possible each geographical region.

NA: Not available.

County. The first column presents the index of ethnic segregation (six races are used for San Francisco), and the second presents the index of segregation by demographic group. The third (only available for Stockholm) presents the measure of segregation by income class. These indexes are interpreted in the following way. Considering the San Francisco Bay Area, the maximum racial entropy in the region is 0.978, which would be obtained if each and every census tract had the racial composition of the region as a whole -- that is, if each tract had the racial proportions indicated by the last line of Table 1. The actual racial entropy of the region is lower, 0.759, due to the segregation of races. The reduction in entropy due to racial segregation is 0.219 or 22.43 percent of the maximum.

Taking the 5 MSA's individually, the maximum racial or ethnic entropy is largest in San Francisco and Oakland, the two MSA's with the smallest fractions of white households. The measures of segregation are also largest in these two MSA's, 25.16 percent and 23.33 percent respectively. The least segregated MSA is clearly Santa Rosa, but it is also the one with the smallest non white population.

The table presents similar information for the Stockholm Metropolitan area for 1975 and for 1985. The reduction in entropy caused by segregation by ethnic group is much smaller, 5.80 percent in 1985, but the segregation index increased considerably during the decade 1975-1985. The level of segregation also appears to be higher in the suburban areas. The level of ethnic segregation is 4 or 5 times greater in San Francisco than Stockholm, but of course the definitions of the ethnic groups are quite different.

Column 2 of table 4 presents analogous information on the segregation of households by demographic type within these two metropolitan regions. For the San Francisco region as a whole, the maximum entropy is 1.485, which would be obtained if each census tract had a distribution of household types identical to that reported in the last column of Table 1. The maximum entropy by household type is a good bit larger than the racial entropy, reflecting in part the more equal classification of households into groups. For the San Francisco region, segregation by household type reduces actual entropy to 1.363 or by 8.19 percent. Thus, for San Francisco racial segregation is about two and a half times more intense than segregation by demographic group. When the entropy measures are disaggregated by MSA, the results are similar. The index of segregation varies from 2.9 percent in the Santa Rosa MSA to 8.5 percent in the Oakland and San Francisco metropolitan areas. In contrast, the index of racial segregation varies from 8.7 percent in Santa Rosa to 23.2 percent in Oakland and 25.2 percent in San Francisco.

The results presented for Stockholm indicate that the level of spatial segregation by demographic type is somewhat greater than in San Francisco. In 1985 the maximum entropy is 1.381 for the region as a whole. The actual entropy level is 1.263, i.e., a reduction by 8.58 percent.

In particular, the spatial segregation of households by demographic type is less in the central city of Stockholm than in San Francisco, but the level of segregation is far more intense in the inner suburbs ringing Stockholm than in the suburban counties surrounding San Francisco. In general, there has been a modest decline in segregation by household type in the Stockholm metropolitan area during the decade 1975-1985, with the sharp exception of the inner ring.

The third column presents, for Stockholm only, the level of segregation estimated by income class. Income segregation is less pronounced than is segregation by household type, but segregation has increased sharply in the inner ring and in the suburbs of Stockholm during the period 1975-1985.

Table 5 compares the conditional and unconditional distributions by ethnic group and household type for the various geographical components of the San Francisco Bay Area and of greater Stockholm. The first column gives the average value of $I(e,h)_t/H(e)_t$ and the second of $I(e,h)_t/H(h)_t$ for the various subregions in the Bay Area and Stockholm. As we have indicated, the entries in the table have a convenient

TABLE 5

Proportionate Differences in Conditional and Unconditional Entropies by Ethnicity and Household Type for Stockholm and San Francisco

	Ethnicity	Household Type
San Francisco Bay Area		
Year: 1980		
Entire Region	8.30%	4.62%
Central City	8.34	4.65
Oakland	9.82	5.72
San Jose	6.76	3.95
Santa Rosa	7.56	2.46
Napa	7.22	3.98
Stockholm Metropolitan Area		
Year: 1975		
Entire Region	3.00%	7.88%
Central City	3.84	10.38
Ring	2.80	7.49
Suburbs	2.64	6.59
Year: 1985		
Entire Region	4.57	9.89
Central City	5.75	12.98
Ring	4.46	9.69
Suburbs	3.90	8.19

Note: For column 1, table entries are $[H(e)+H(h)-H(e,h)]/H(e) = I(e,h)/H(e)$.

For column 2, table entries are $[H(e)+H(h)-H(e,h)]/H(h) = I(e,h)/H(h)$.

interpretation. Suppose the spatial distribution of household types in the metropolitan region is governed by "economic forces." Under these circumstances, recognizing the known and prior spatial distribution of household types explains only a small fraction of the observed segregation of household by race or ethnic group. For San Francisco, only 8.3 percent of the racial segregation observed could be attributed to segregation by household type arising from economic forces.

For Stockholm in 1975 the fraction is even smaller. Only about 3 percent of the segregation of households by ethnic group could be "explained" by the segregation of households by household type. The fraction has risen substantially during the decade 1975-1985 however.

From column 2 only about 4.6 percent of the spatial segregation of household types in San Francisco could be explained by the prior segregation of households by race. For the largest central cities of San Francisco and Oakland, the upper limit is less than 6 percent.

For Stockholm a much larger fraction of spatial segregation by household type could be explained by the prior segregation of households by ethnic group. Moreover, the fraction has grown considerably during the decade 1975-1985.

Despite the many differences in the metropolitan areas, the principal results are similar: Only a small fraction of segregation by household type can be explained by a prior segregation of households by race or ethnic type. An even smaller fraction of segregation by race can be explained by economic forces leading to a clustering by demographic group.

Table 6 indicates, for Stockholm only, the influence of income class. Consider again the three bivariate distributions ρ_{eh*t}, ρ_{e*it} and ρ_{*hit} and the three marginal distributions ρ_{e**t}, ρ_{*h*t} and ρ_{**it}. The table compares two of the bivariate distributions with the corresponding marginal distributions. Each cell in the table is obtained by dividing the average value of $I_s = \Sigma_t(w_t/W_s)I_t$ by the maximum value of I_s. The first column reports

$$(19) \quad I(e,i)_t = \Sigma_i \rho_{**it}\, I(\{\rho_{e*it}/\rho_{**it}\},\rho_{e**t}) \quad ,$$

where the maximum value of $I(e,i)_t = H(e)_t$. The ratio between them indicates the extent of ethnic segregation which could be "explained" by income segregation. The values in the second, third and fourth column are given by $I(e,i)_t$, $I(h,i)_t$, and $I(h,i)_t$ divided by the corresponding maximum values of $H(e)_t$, $H(i)_t$, and $H(h)_t$, respectively.

As indicated in the first two columns, practically none of the segregation of households by ethnic group can be explained by income segregation, and none of the segregation by income group can be explained by ethnic segregation. In contrast, a large and growing fraction of segregation by

TABLE 6

Proportionate Differences in Conditional and Unconditional Entropies for Stockholm Metropolitan Area

	By Ethnicity and Income		By Household Type and Income	
	Ethnicity	Income	Household Type	Income
Year: 1975				
Entire Region	2.19%	1.14%	13.20%	18.10%
Central City	2.70	1.14	11.76	16.18
Ring	2.16	1.16	13.40	19.28
Suburbs	1.84	1.11	12.09	18.23
Year: 1985				
Entire Region	2.80	1.71	15.18	20.01
Central City	3.62	1.80	16.37	18.31
Ring	2.81	1.76	15.27	20.83
Suburbs	2.28	1.55	14.28	20.32

Note: For column 1, table entries are $[H(e)+H(i)-H(e,i)]/H(e) = I(e,i)/H(e)$.
For column 2, table entries are $I(e,i)/H(i)$.
For column 3, table entries are $I(h,i)/H(h)$.
For column 4, table entries are $I(h,i)/H(i)$.

household type can be explained by segregation by income class. A larger and growing fraction of segregation by income class can be explained by patterns of segregation by household type.

Table 7 presents further disaggregation. The three columns present the segregation indexes obtained when comparing the conditional distributions $\{\rho_{ehit}/\rho_{*hit}\}$, $\{\rho_{ehit}/\rho_{e*it}\}$ and $\{\rho_{ehit}/\rho_{eh*t}\}$ with the marginal distributions ρ_{e**t}, ρ_{*h*t} and ρ_{**it}, respectively.

Column 1 indicates the fraction of observed segregation by ethnic group which could be explained by the prior segregation of households by household type and income. The extent to which segregation of ethnic groups is explicable by these other forces is rather small, but it is growing. In contrast, according to column 2, the extent to which segregation by household type is explicable by the prior segregation of households by income class and ethnicity is much larger, and it is growing. As indicated in column 3, about a fifth of the observed segregation of households by income level is explicable by the pattern of household occupancy by ethnicity and household type.

VI. SUMMARY AND CONCLUSIONS

This paper considers residential segregation by ethnic group, household type, and income class for the Stockholm

TABLE 7

Proportionate Differences in Conditional and Unconditional Entropies by Ethnicity, Household type, and Income for Greater Stockholm

	Ethnicity	Household Type	Income
Year: 1975			
Entire Region	9.99%	16.17%	19.20%
Central City	12.47	17.81	17.06
Ring	9.49	16.14	20.86
Suburbs	8.85	14.30	19.60
Year: 1985			
Entire Region	11.95	19.40	21.27
Central City	15.00	21.41	19.31
Ring	11.69	19.35	22.09
Suburbs	10.35	- 18.08	21.80

Note: For column 1, table entries are $[H(e)+H(q)-H(e,q)]/H(e) = I(e,q)/H(e)$, where q is the set of household type-income categories.
For column 2, table entries are $I(h,r)/H(h)$, where r is the set of ethnicity-income categories.
For column 3, table entries are $I(i,z)/H(i)$, where z is the set of ethnicity-household type categories.

metropolitan area. By relying upon special census tabulations, the analysis is replicated for 1975 and 1985 using identical definitions. The results indicate that spatial segregation by ethnic group is small, but it is growing. Spatial segregation by household type is larger, and has declined slightly in Stockholm with the exception of the inner suburban ring. Segregation by income class is slightly less pronounced than is segregation by household type, but it is growing especially outside the central city of Stockholm.

Very little of the segregation by ethnicity can be explained by the prior segregation of households by household type or income class or by the joint distribution by household type and income class. Very little of the spatial segregation by household type can be explained by the prior segregation of households by ethnicity. A larger fraction can be explained by the distribution of households by income class.

Some of these results can be compared directly with patterns of segregation in the San Francisco Bay Area observed in 1980. The extent of segregation by household type is somewhat larger in Stockholm than in San Francisco, with reduced levels of segregation in the city of Stockholm offset by increased demographic segregation in the near suburbs. For both cities, only a small fraction of the observed pattern of racial or ethnic segregation can be explained by the pattern of segregation by household type. Similarly, only a small fraction of observed segregation by household type can be explained by the residential patterns of ethnic or racial groups.

Even though ethnic segregation is defined very differently for San Francisco and for Stockholm, it is tempting to attribute the low level of segregation in Stockholm to Swedish housing policy which rations residential locations by queue. As indicated in the Appendix, this conjecture is probably false.

Evidently the forces which give rise to segregation by demographic group are somewhat stronger in Stockholm than in San Francisco. In both metropolitan regions, the forces that give rise to segregation by demographic group are quite independent of the forces giving rise to segregation by racial or ethnic group. In Stockholm, the segregation of households by income class does explain a substantial fraction of the observed segregation by household type, but it explains almost none of the observed segregation by ethnicity.

In each of these very different metropolitan regions, spatial segregation by race or ethnicity seems unrelated to spatial segregation by income class or demographic grouping.

Appendix

A policy that rations rental housing by a queue and which supplies municipally owned rental housing could, of course, be used to promote the integration of ethnic groups or household types. Table A1 provides some evidence on this issue. It reports the simple correlations, across census tracts, between one of the segregation indexes and a measure of government activity in housing supply. Simple correlations are reported between the measure of segregation by ethnic group and the fraction of dwellings in multi-family structures. There is essentially no correlation between these measures. However, the correlation between the level of segregation and the fraction of dwellings in non-profit, municipally-owned, structures is much larger.

This positive correlation between segregation and the extent of non-profit (state subsidized) housing suggests that housing policy might, in fact, be one cause of increased ethnic segregation.[4]

4 It has been reported elsewhere, for example, that almost all dwellings in the most intensely segregated areas, in Stockholm as well as other large Swedish cities, are owned by non-profit companies under municipal control.

TABLE A1

Simple correlation coefficients for Stockholm, 1975 and 1985 (806 Census tracts)

		Index of Segregation by Ethnic Group	
		1975	1985
Fraction of dwellings in multi-family structures	1975	0.12	-
	1985	-	0.05
Fraction of dwellings in municipally-owned structures	1975	0.34	-
	1985	-	0.40

REFERENCES

Allison, Paul D., "Measures of Inequality," **American Sociological Review**, Vol 43, December 1978, pp 865-880.

Harsman, Bjorn and John M. Quigley, **Housing Markets and Housing Institution: An International Comparison**, Boston: Kluwer, 1991.

Hobson, A., "A New Theorem of Information Theory," **Journal of Statistical Physics**, #1, 1969, pp 383-391.

Miller, Vincent P. and John M. Quigley, "Segregation by Racial and Demographic Group: Evidence from the San Francisco Bay Area," **Urban Studies**, Vol 27, #1, 1990, pp 3-21.

Schnare, Ann B., "Trends in Residential Segregation by Race: 1960-1970," **Journal of Urban Economics**, Vol 7, 1980, pp. 293-301.

Shannon, C. E., "A Mathematical Theory of Communication," **Bell System Technical Journal**, #27, 1948, pp 379-423 and 623-656.

Struyk, Raymond and M. Turner, "Exploring the Effects of Racial Preferences on Urban Housing Markets," **Journal of Urban Economics**, Vol 19, 1986, pp. 131-147.

Theil, Henri, **Statistical Decomposition Analysis**, Amsterdam: North Holland Publishing Co., 1972.

Weibul, Jorgan W., "An Axiomatic Approach to the Measurement of Accessibility," **Regional Science and Urban Economics**, Vol 6, 1976, pp 357-379.

White, Michael J., "The Measurement of Spatial Segregation," **American Journal of Sociology**, Vol 88, #5, March 1983, pp 1008-1018.