Economics Letters 30 (1989) 237–243
North-Holland

ON THE MEASUREMENT OF EMPLOYMENT SEGREGATION

Jacques G. SILBER

Bar Ilan University, 52100 Ramat Gan, Israel

Received 15 December 1988
Accepted 21 February 1989

A new index of occupational segregation is proposed, which is related to Gini's Concentration Ratio. An illustration based on French data (1987) indicates that similar conclusions are drawn whether the famous dissimilarity index [Duncan and Duncan (1955)] or the one proposed here is used, but the new index might have some advantage over the one suggested by the Duncans.

1. Introduction

In a recent letter, Butler (1987) has proposed to use the Pietra and Gini indices when measuring employment segregation. The purpose of the present letter is to show, firstly, that another kind of index called the *G*-segregation index may be borrowed from the income inequality literature and adapted to the study of segregation, secondly, that Butler's indices were wrongly defined, but that if correct weights are used, his Pietra index is in fact the Duncans' famous dissimilarity index while his Gini index is simply the *G*-segregation index. An empirical illustration based on French data indicates that similar conclusions may be drawn whether the Duncans' or the G-segregation index is used.

2. Income inequality indices and the measurement of employment segregation

In a recent study of income inequality change in the U.S.A. [Berrebi and Silber (1987b)] it has been proven that the Gini Index of Inter-States Income Inequality (concentration) could be written as

$$J_G = e'Gs, \tag{1}$$

where s is a column vector giving the share of each state in the U.S. National Income (the shares are ordered by decreasing value of the average income), e' is a row vector giving the corresponding population shares of each state and G is a square matrix, called the G-matrix, whose elements g_{ij} are equal to 0 when $i = j$, to -1 if $j > i$, and to $+1$ if $j < i$.

Expression (1) may be extended to measure any kind of inequality. Let F_i and M_i be respectively the number of female and male workers in occupation i and let s_i^{F} be the share of occupation i in total female employment so that

$$s_i^{\mathrm{F}} = (F_i/F) = [F_i/(F_i + M_i)][(F_i + M_i)/(F + M)][(F + M)/F], \tag{2}$$

where $F = \Sigma_i F_i$ and $M = \Sigma_i M_i$.

Expression (2) may also be written as

$$s_i^{\mathrm{F}} = (f_i/f)w_i, \tag{3}$$

where $f_i = F_i/(F_i + M_i)$ is the share of female employment in sector i, $f = F/(F + M)$ the share of female employment in the whole economy and $w_i = (M_i + F_i)/(M + F)$ the employment share of occupation i.

The ratio $h_i = f_i/f$ is in fact an index of the feminization of occupation i and it may be seen that s_i^{F} may be written as

$$s_i^{\mathrm{F}} = h_i w_i. \tag{4}$$

Combining (1) and (4) it is possible to define an index J_{F} of the diversity (inequality) in the degree of feminization of the various occupations as

$$J_{\mathrm{F}} = w_h' G s_h^{\mathrm{F}}, \tag{5}$$

where s_h^{F} is the column vector of the ranked shares s_i^{F} (ranked by decreasing value of the indices h_i) and w_h' the row vector of the shares w_i (the latter are also ranked by decreasing value of the indices h_i).

It is evidently possible to define in a similar way an index J_{M} of the diversity in the degree of male employment intensity of the various sectors. J_{M} would be written as

$$J_{\mathrm{M}} = w_k' G s_k^{\mathrm{M}}, \tag{6}$$

where s_k^{M} would be the column vector of the ranked shares s_i^{M} of occupation i in total male employment, and w_k' the ranked vector of the shares w_i, both vectors s_k^{M} and w_k' being ranked by decreasing value of a male employment intensity index k_i defined as $k_i = (1 - f_i)/(1 - f)$.

Expressions (5) and (6) may be extended to define any kind of concentration, once it is assumed that the vector w' refers to expected shares whereas s^{F} and s^{M} measure actual shares.

It is in fact proposed to base the measurement of occupational segregation on the comparison of two concentration curves. The first one would correspond to the set of the cumulative values of w_i and s_i^{F} (ranked by increasing value of h_i). It can then be proven that the area between this curve and the diagonal is equal to half the value of J_{F} in (5). The second concentration curve would be defined by the set of the cumulative values of w_i and s_i^{M}, both being ranked also by increasing value of h_i. Whereas the first curve is a graphical representation of the diversity in the degree of feminization of the various occupations, the second one does not necessarily refer to the degree of diversity of male employment intensity because the index of feminization h_i (and not the index k_i of male employment intensity) was the occupation classification criterion used in deriving this second curve (and nothing guarantees that both classifications would have been identical). But the difference between the areas lying between these concentration curves and the diagonal should give an indication of the amount of occupational segregation. Using the notations previously defined, it can be seen that this difference may be written as

$$S = w_h' G \left(s_h^{\mathrm{F}} - s_h^{\mathrm{M}}\right). \tag{7}$$

It is proposed to use this index S which will be called henceforth the G-segregation index as a measure of the degree of employment segregation in the economy.

In (7), the subscript h indicates that the shares w_i, s_i^F, and s_i^M have been ranked by decreasing value of the feminization index h_i. It can be shown that if the indices k_i of male intensity had been used as reference, the expression $w_k'G(s_k^M - s_k^F)$ would have been also equal to S, this being a consequence of the fact that $k_i = (1 - f_i)/(1 - f)$.

3. Comparing the G-segregation index with other measures of segregation

Most studies of occupational, residential or school segregation have been based on the use of the by now famous dissimilarity index D [Ducan and Duncan (1955)], which is defined as

$$D = \sum_i |(M_i/M) - (F_i/F)|/2. \tag{8}$$

The principal shortcoming of the dissimilarity index D, as indicated by Butler (1987), is that it gives equal weight to each occupation, regardless of its relative size (share in total employment). Butler (1987) has therefore advocated the use of two alternative segregation indices, which both take into account the relative importance of each occupation. His proposal is also to apply to the analysis of segregation, concepts which are widely used in income inequality studies. The variable which would play the role of income is the ratio M_i/F_i to male of female employees in each occupation. The segregation measures suggested are the Pietra (relative mean deviation) and Gini indices defined respectively [Butler (1987)] as

$$P = \sum_i w_i |l_i - \bar{l}|/2\bar{l}, \tag{9}$$

$$I = \sum_i \sum_j w_i w_j |l_i - l_j|/2\bar{l}, \tag{10}$$

where $l_i = (M_i/F_i)$ and $\bar{l} = \sum_i w_i l_i$.

In an illustration based on the 1950, 1960, and 1970 U.S. censuses, Butler shows that the ranking of occupational segregation by industry is very sensitive to the choice of the segregation index (D, P, or G). Some of the correlations between the indices are even negative, a result which is seldom obtained in income inequality studies when comparing the value given for example by Theil's, Atkinson's or Gini's indices.

Such a puzzling result might well be due to the fact that the average $\bar{l}$ which is used in regression (9) and (10) is not equal to the average male/female ratio M/F in the economy. It is therefore doubtful that his P and I indices may be called Pietra and Gini indices. However, if instead of using the weights w_i (the share of occupation i in total employment) the shares s_i^F of occupation i in total female employment are used as weights, it can be seen that

$$\hat{l} = \sum_i s_i^F l_i = \sum_i (F_i/F)(M_i/F_i) = \left(\sum_i M_i\right)/F = M/F. \tag{11}$$

Butler's proposal (1987) should therefore be corrected by defining the Pietra and Gini indices of segregation as

$$P' = \sum_i s_i^F |l_i - \hat{l}|/2\hat{l}, \tag{12}$$

$$I' = \sum_i s_i^F s_j^F |l_i - l_j|/2\hat{l}. \tag{13}$$

It will now be proven that P' is in fact the Duncans' index D whereas I' is equal to the G-segregation index defined in section 2.

Combining (11) and (12) one derives that

$$P' = \sum_i [(F_i/F)\,|(M_i/F_i) - (M/F)|]/2(M/F)$$

$$\leftrightarrow P' = \sum_i [(F_i/F)\,|((M_i/F_i)/(M/F)) - l|]/2$$

$$\leftrightarrow P' = \sum_i |(M_i/M) - (F_i/F)|/2 = D.$$

Since the index I' defined in (13) is a Gini index, it may be written also, using the notations adopted in (1) as

$$I' = s_i^{F\prime} G\left((F_i l_i)/\left(\sum_i F_i l_i\right)\right). \tag{14}$$

In (14), the vectors on both sides of the G-matrix G have been symbolised by their ith element in order to simplify the notations. Combining (2), (13), and (14) one derives that

$$I' = (F_i/F)G(M_i/M). \tag{15}$$

On the other hand, S, as defined in (7), may be written also, using the definitions of w_i, s_i^F and s_i^M as

$$S = [(M_i + F_i)/(M + F)]\,G[(F_i/F) - (M_i/M)]. \tag{16}$$

In (16), the vectors on both sides of the G-matrix have also been represented by their ith element. Expression (16) may also be written as

$$S = [(M_i/M)(M/(M+F)) + (F_i/F)(F/(M+F))]\,G[(F_i/F) - (M_i/M)]$$

$$\leftrightarrow S = (M/(M+F))[(M_i/M)G(F_i/F) - (M_i/M)G(M_i/M)]$$

$$+ (F/(M+F))[(F_i/F)G(F_i/F) - (F_i/F)G(M_i/M)],$$

since $M/(M+F)$ and $F/(M+F)$ are constant.

It should be noted, however, that, given the definition of the G-matrix G,

$$(F_i/F)G(F_i/F) = (M_i/M)G(M_i/M) = 0,$$

so that S may be simplified as follows:

$$S = (M/(M+F))[(M_i/M)G(F_i/F)] - (F/(M+F))[(F_i/F)G(M_i/M)].$$

It may be observed also that, by definition of the G-matrix, $(M_i/M)G(F_i/F) = -(F_i/F)G \times (M_i/M)$ so that S is finally written as

$$S = -(F_i/F)G(M_i/M), \tag{17}$$

since $(M/(M+F)) + (F/(M+F)) = 1$.

In (17), the vectors representing the elements (F_i/F) and (M_i/M) have been ranked, as indicated earlier, by decreasing value of the index h_i. In (15), the elements (F_i/F) and (M_i/M) have been ranked by decreasing value of the ratio (M_i/M). Since $h_i = (f_i/f) = (F_i/(F_i + M_i))/(F/(F+M))$, the order of the elements h_i is in fact that of the ratios $F_i/(F_i + M_i) = 1/(1 + (M_i/F_i))$, and is therefore the opposite of the order of the ratios (M_i/F_i). Since it can be easily proven that by

Table 1
Occupational segregation indices by area of residence, level of education, age or duration of work (France, 1987). [a]

Criteria	Duncan index	G-segregation index
(I) *Area of residence* [b]		
– Rural areas	0.439	0.619
– Urban areas with less than 20000 inhabitants	0.515	0.679
– Urban areas with 20000 to 200000 inhabitants	0.528	0.669
– Cities with more than 200000 inhabitants (not including Paris)	0.503	0.650
– Paris (metropolitan area)	0.500	0.627
– France (total)	0.485	0.644
(II) *Level of education* [b]		
– Unspecified diploma	0.439	0.597
– No diploma or 5 years of schooling only	0.398	0.576
– B.E.P.C. (9 years of schooling)	0.371	0.497
– C.A.P. or B.E.P. [c]	0.586	0.726
– Baccalaureat (12 years, of schooling)	0.456	0.577
– University degree (B.A. or B.Sc.)	0.434	0.500
– University degree (M.A. or Ph.D.)	0.309	0.398
– Presently studying	0.533	0.631
– All categories	0.452	0.602
(III) *Age* [b]		
– 15 to 24	0.606	0.712
– 25 to 39	0.495	0.649
– 40 to 49	0.484	0.653
– 50 to 59	0.455	0.629
– 60 or more	0.382	0.516
– Total	0.485	0.644
(IV) *Hours of work* [b]		
– Part-time, less than 15 hours per week	0.459	0.572
– Part-time, 15 to 29 hours per week	0.383	0.506
– Part-time, 30 hours or more per week	0.445	0.542
– Full-time, less than 30 hours per week	0.211	0.267
– Full-time, 30 to 34 hours per week	0.336	0.482
– Full-time, 35 to 39 hours per week	0.528	0.644
– Full-time, 40 hours or more per week	0.399	0.551
– Unspecified schedule	0.321	0.457
– Total	0.452	0.601

[a] *Data Sources*: Tables PA04, PA14, GE0603, FORM05 in Collections de l'I.N.S.E.E. (1987).
[b] 31 occupations were distinguished in the classification by area of residence and age, whereas there were only 17 occupational categories in the classification by level of education and hours of work.
[c] These are diplomas certifying the acquisition of a specific profession (e.g. carpenter) whereas the B.E.P.C. is a general diploma, at the end of the ninth grade.

reversing the order of the vectors on both sides of the G-matrix, the scalar obtained in the multiplication will have the opposite sign, it can be concluded that

$$S = I' = (F_i/F)G(M_i/M), \tag{18}$$

where (F_i/F) and (M_i/M) are ranked by decreasing value of (M_i/M).

The G-segregation index defined in section 2 is therefore equal to the Gini index of the (male/female) ratios l_i, where the weights used in the computation of this index are the shares of each occupation in total female employment.

It is suggested to use in measuring employment segregation, either the Duncans' index which is in fact a Pietra index (relative mean deviation) of the male/female ratios l_i or the G-segregation index which is in fact a Gini index of these ratios. Since it has been proven [Berrebi and Silber (1987a)] that the Gini index is a weighted relative mean deviation, one expects to find probably similar results when using the Duncans' or the G-segregation index. This is indeed what the illustration presented in the next section will show.

4. An illustration: Occupational segregation in France in 1987

In table 1, occupational segregation indices have been computed, based on a Labor Force Survey taken in France in 1987. Data on male and female employment by occupation were classified by area of residence, level of education, age or numbers of hours of work per week. Two segregation indices have been computed: the dissimilarity index and the G-segregation index which, as indicated earlier, are in fact respectively Pietra and Gini indices of the male/female ratios.

As a whole, the dissimilarity index and the G-segregation index lead to similar conclusions. Occupational segregation is lowest in rural areas and highest in urban areas of medium size (20 000 to 200 000 inhabitants). Occupational segregation is lowest among individuals with M.A. or Ph.D. and highest among individuals having a professional diploma (e.g., carpenters or garage mechanics). Segregation decreases also with age, although it is not clear whether this is an age or a cohort effect. Finally, segregation seems to be higher among part-time than among full-time employees but the differences are less clear once the breakdown by hours of work is more detailed.

The correlation coefficients between both indices for the four sets of data used, are respectively equal to 0.800, 0.924, 0.938, and 0.966. Given such a strong correlation it does not probably matter whether one or the other index is used but the G-segregation index might be more handy to use when a breakdown by population subgroups is desired since the use of the G-matrix facilitates such a breakdown [c.f. Silber (1989)]. Future work based on more detailed Census data and using such a breakdown should help drawing firmer conclusions concerning the usefulness of the G-segregation index.

References

Berrebi, Z.M. and J. Silber, 1987a, Dispersion, asymmetry and the Gini index of inequality, International Economic Review 28, 331–338.

Berrebi, Z.M. and J. Silber, 1987b, Regional differences and the components of growth and inequality change, Economics Letters 25, 295–298.

Butler, Richard J., 1987, New indices of segregation, Economics Letters 24, 359–362.

Collections de l'I.N.S.E.E., Demographie et Emploi, 1987, Enquete sur l'emploi de 1987, resultats detaillés, D 122.

Duncan, Otis D. and Beverly Duncan, 1955, A methodological analysis of segregation indexes, American Sociological Review 20, 210–217.

Silber, Jacques G., 1988, Growth, inequality change, permutations and the measurement of change, Mimeo. (Bar Ilan University, Ramat Gan).

Silber, Jacques G., 1989, Factor components, population subgroups and the computation of the Gini index of inequality, The Review of Economics and Statistics, to appear.