DIGANTA MUKHERJEE

ON A SYMMETRIC MEASURE OF OCCUPATIONAL SEGREGATION

(Accepted 28 June 2001)

ABSTRACT. Occupational segregation according to sex is widely prevalent in almost all countries. This paper sugge sts and characterizes a symmetric measure of segregation. This measure was originally proposed in a different situation by Jeffreys (1946). We also provide two numerical illustrations of change in segregation over time.

KEY WORDS: characterization, ILO, NSS, segregation

1. INTRODUCTION

It is well known that there exists wide differences between the earnings of men and women in almost all the countries of the world. The prime reason behind this is the fact that women are largely concentrated in a limited range of generally low paid jobs, or in household chores of no well defined pay at all. With the awareness of its existence, a literature in Economics and Sociology has grown, which focuses attention on inequality in the distribution of people across occupations. This phenomenon is popularly termed 'occupational segregation'. Though we talk about the differences between sexes; racial, national or regional differences could also be the variable of interest. The key questions here are: (i) why some groups tend to be concentrated in a subset of all occupations? and (ii) whether the degree of concentration is changing through time?

Any discussion on the phenomenon requires some yardstick to lean upon. Hence, over time a large number of measures for occupational segregation have been proposed in the literature. Duncan-Duncan (1955) defined a dissimilarity index which is the most popular till date. Moir and Selby-Smith (1979), Lewis (1982), Karmel and Maclachlan (1988) and Silber (1992) developed vari-

Social Indicators Research **57:** 1–11, 2002.

ants of this index. Butler (1987), Silber (1989), Hutchens (1991) are a few of the other significant contributions.

The present paper aims to propose a set of intuitively natural axioms which any reasonable measure of occupational segregation should satisfy. Section 2 of this paper starts with the notations to be used throughout the paper and goes on to state our axioms with some economic interpretation of them. In Section 3, we cite some popular (or not so popular) measures of segregation and discuss their reasonableness in the light of our proposed axioms.

The main goal of the paper is to characterize a symmetric measure of segregation originally proposed by Jeffreys (1946). We try to find the set of axioms that are necessary and sufficient for characterizing the proposed measure. Section 4 contains the characterization results. Section 5 gives two numerical illustrations of our measure of segregation using Indian National Sample Survey (NSS) data on four rounds and International Labour Organization (ILO) data from 12 countries over three time periods. Finally Section 6 concludes.

2. THE AXIOMS

We assume that there are n occupations in the economy with F_i (M_i) number of females (males) employed in the ith occupation, $i = 1, 2, \ldots, n$ ($n > 1$). Let $p_i = F_i/\sum_{j=1}^{n} F_j$ and $q_i = M_i/\sum_{j=1}^{n} M_j$ be the proportion of female (male) workers in the ith occupation, $i = 1, 2, \ldots, n$ (obviously $\sum p_i = \sum q_i = 1$). Henceforth we talk only in terms of the n-vectors $P = (p_1, p_2, \ldots, p_n)$ and $Q = (q_1, q_2, \ldots, q_n)$ which are occupational distributions of the two sexes. For the sake of analytical simplicity, we talk about vectors P only if $P = (p_1, p_2, \ldots, p_n)$ with $p_i > 0$ for all $i = 1, 2, \ldots, n$. That is, the admissible set of P vectors is $\Gamma_n^0 = \{P = p_1, p_2, \ldots, p_n) | 0 < p_i < 1, \sum p_i = 1\}$. We could also have used $\Gamma_n = \{P = (p_1, p_2, \ldots, p_n) | 0 \leq p_i \leq 1, \sum p_i = 1\}$, but that would lead us to non-essential analytical difficulties which we defer for later discussion. In what follows, f is a real valued function defined on (0,1). That is, f:(0,1) $\rightarrow R$. Now, a segregation measure F is a functional defined by setting

$$(1) \qquad F(P, Q) = \sum_{k=1}^{n} f(p_k, q_k) \text{ whenever } P, Q \in \Gamma_n^0,$$

satisfying certain properties which we state as the following axioms.

Finiteness (**F**): $F(P,P) < \infty$ for all $P \in \Gamma_n^0$.
Directedness (**D**): $F(P,Q) \geq F(P,P)$ for all $P \in \Gamma_n^0$, $>$ holds if $P \neq Q$.
Symmetry (**S**): $F(P,Q) = F(Q,P)$ for all $P,Q \in \Gamma_n^0$.
Anonymity (**A**): $F(P,Q) = F(\Pi P,\Pi Q)$ for all $P,Q \in \Gamma_n^0$ and all permutation matrices Π.
Transfer Principle (**T**): If for $P,Q \in \Gamma_n^0$ with $p_i > q_i$ and $p_j < q_j$, we define P' by $p'_i = p_i - e$, $p'_j = p_j + e$ for $e > 0$ sufficiently small and $p'_k = p_k$ for $k \neq i,j$. Then $F(P,Q) > F(P',Q)$.

Axiom (F) simply ensures that we are not talking in vacuum, F is finite at least in the case of identical distributions. Axiom (D) says that identical distribution is the case of least segregation. (F) and (D) was proposed by Johnson (1979). (S) ensures impartiality between the two sexes. (A) requires that the ordering of the occupations is irrelevant to the measurement of segregation (Chakravarty and Silber, 1990). Axiom (T) states that if there is a small shift of the female labor force from a female dominated occupation to a male dominated occupation, the segregation measure must decrease. (S) and (T) are due to Kakwani (1994).

We now state two **Sensitivity** axioms (Kakwani, 1994) for segregation measures for $P,Q \in \Gamma_n^0$.

Upward Sensitivity (**Se1**): If $p_i > q_i$ and $p_j > q_j$ $(j \neq i)$ with $p_j - q_j = p_i - q_i$ and $p_i + q_i < p_j + q_j$, then $F(P,Q) > F(P',Q)$ with P' defined from P as in (T).
Downward Sensitivity (**Se2**): If $p_i > q_i$ and $p_j > q_j$ $(j \neq i)$ with $p_i - q_i > p_j - q_j$ and $p_i + q_i = p_j + q_j$, then $F(P,Q) > F(P',Q)$ with P' defined from P as in (T).

Axiom (Se1) implies that it is more desirable to reduce male-female gaps in the lesser populated occupations. This is desirable as these are generally higher paid occupation and hence should be given larger weights. (Se2) states that larger gaps should be given larger weights.

We define a composition of two vectors $P \in \Gamma_n^0$ and $Q \in \Gamma_m^0$ (m > 1) $P*Q$ as $P*Q = (p_1q_1, \ldots, p_1q_m, p_2q_1, \ldots, p_2q_m, \ldots, p_nq_1, \ldots, p_nq_m) \in \Gamma_{mn}^0$.

We now come to our final axiom.

Subgroup Consistency (**C**): $F(P*Q,R*S) = F(P,R) + F(Q,S)$ where $P,R \in \Gamma_n^0$, $Q,S \in \Gamma_m^0$ and $P*Q,R*S \in \Gamma_{mn}^0$.

This axiom states that subdividing each occupation into sub-occupations and then measuring segregation is same as adding the segregation indices under these two kinds of occupational grouping. (C) has been proposed by Aczél and Daróczy (1975).

3. A COMPARATIVE STUDY OF SOME POPULAR SEGREGATION MEASURES

The popular-most measure of segregation is the **Dissimilarity Index** proposed by Duncan-Duncan (1955). It is defined as

$$I(P,Q) = \frac{1}{2}\sum_{i=1}^{n} |q_i - p_i|, \tag{2}$$

which satisfies axioms (F), (D), (S), (A), (T) but do not satisfy the others (for further discussion on this, see Kakwani, 1994).

The Coefficient of Variation (C.V.) type index,

$$CV(P,Q) = \sum_{i=1}^{n} q_i \left[\frac{p_i}{q_i} - 1\right]^2, \tag{3}$$

proposed by Hutchens (1991) which obviously violates (S) and (C) but satisfies the other axioms.

The Gini type index proposed by Silber (1989),

$$G(P,Q) = \sum_{i=1}^{n} p_i(1 - q_i - 2\sum_{i+1}^{n} q_i), \tag{4}$$

which again violates (S) and (C).

Another very important measure of seggregation is Kakwani's (1994) S_β. It is defined as

$$S_\beta(P,Q,M,f) = \frac{a^\beta(1-a)^\beta}{a^\beta+(1-a)^\beta}\sum_{i=1}^{n} \frac{|p_i-q_i|^{\beta+1}}{\{aq_i+(1-a)p_i\}^\beta} \tag{5}$$

where $m(f)$ is the total number of males (females) in the labor force and a is the proportion of females in the labor force ($a = \frac{f}{m+f}$). When $\beta = 0$, S_β is equal to $I(P,Q)$, the Dissimilarity index proposed by Duncan-Duncan. S_β satisfies all the axioms considered in this paper with the qualification that it satisfies (S) in a restricted sense in that, (S) do not involve m or f (for further discussions on S_β, see Kakwani, 1994).

We now come to our most important class of divergence measures. In the terminology of Kullback (1957), the Directed Divergence of P and Q are the two sums in

$$(6) \qquad \beta \sum_{k=1}^{n} q_k Log \frac{q_k}{p_k} + \gamma \sum_{k=1}^{n} p_k Log \frac{p_k}{q_k}$$

or $\beta D_n(Q,P) + \gamma D_n(P,Q)$ for β, γ arbitrary non-negative real numbers, not both zero. The function $D_n(P,Q) = \sum_{k=1}^{n} p_k Log \frac{p_k}{q_k}$ is non-negative and attains minimum value 0 when $P = Q$. Thus it serves as a segregation measure between distributions P and Q. It is used in Statistics, Pattern Recognition, Coding Theory, Signal Processing and Information Theory. However, this Directed Divergence is neither symmetric nor does it satisfy the triangle inequality.[1] Thus, its application as a metric or measure of distance is limited. So Jeffreys (1946) introduced the notion of Symmetric Divergence as the sum (6) with $\beta = \gamma$ as

$$J_n(P, Q) = \beta(D_n(Q, P) + D_n(P, Q))$$

or in explicit form

$$(7) \qquad J_n(P, Q) = \beta \sum_{k=1}^{n} (p_k - q_k) Log \frac{p_k}{q_k}$$

($\beta > 0$ real). The measure in (7) is also called **J-Divergence** in honor of Jeffreys.

It is an easy exercise to check that the measure in (6) satisfies all the axioms except (S) which is satisfied only if $\beta = \gamma$, that is, by the **J-Divergence** measure of segregation.

4. THE CHARACTERIZATION RESULTS

We are concerned with properties that characterize $F(P,Q)$ as defined in (1) as a positive linear combination (6) of the Directed Divergences of P and Q. The results that follow are proved in Johnson (1979), Kannappan et al. (1993), Kannappan (1972, 1972a) and Aczél and Daróczy (1975). We state our first result as

Lemma 1: Suppose F satisfies axioms (F), (D) and (C), then f has the form

$$f(u, v) = g\left(\frac{u}{v}\right).u \qquad (8)$$

The proof of the lemma is in Johnson (1979) for a slightly different set-up, hence we omit the proof. It can be further proved that

Theorem 2: Suppose F satisfies axioms (F), (D) and (C), then F has the form (6) for some constants $\beta, \gamma \geq 0$, not both zero.

The proof of Theorem 2 is in Kannappan (1972, 1972a) and also Johnson (1979). It relies on Lemma 1. Theorem 2 leads us to our final result,

Corollary 3: Suppose F satisfies axioms (F), (D), (C) and (S), then F has the form $J_n(P,Q)$ as in (7).

Proof: Proof is immediate if we impose the symmetry restriction on (6) which implies $\beta = \gamma$ and we get the desired result. ■

To take care of the cases where p_k or q_k takes the value zero for some k, $1 \leq k \leq n$, (to define $F(P,Q)$ in Γ_n instead of Γ_n^0), we have to extend the range of $f(\cdot)$ to the value ∞. Here we adopt the convention, $uLog(u/0) = \infty$, $0Log(0/u) = 0$ and $0Log(u/0) = 0$. With this modification we can extend our results to F defined on Γ_n.

5. NUMERICAL ILLUSTRATION

For our illustrative purpose, we carry out two numerical exercises. the first one of which deals with gender segregation in employment

TABLE I

Occupational segregation in India over the years 1972–1988 (data from NSS quinquennial survey)

Round	Labor force (in millions)	Proportion of females	Kakwani S_β				J_n
			0.5	1	1.5	2	
27	236.3	0.329	0.0652	0.0327	0.0219	0.0105	0.2081
32	264.3	0.329	0.0748	0.0403	0.0232	0.014	0.2475
37	302.7	0.339	0.0746	0.0386	0.0211	0.0121	0.2218
43	322.2	0.331	0.0774	0.041	0.023	0.0136	0.2596

pattern in the Indian scenario as revealed by the NSS Quinquennial survey for the rounds 27 (year 1972–1973), 32 (year 1977–1978), 38 (year 1982–1983) and 43 (year 1987–1988). The data source is Sarvekshana (1990), the official journal of NSS. We consider the usually employed labor force in the major occupational groups listed in Sarvekshana. According to the NSS, those who worked for a relatively larger period of a reference period of 365 days preceding the date of survey constitute this labor force. The definition of status was different for round 27, but if we consider all usually employed, then the figures are comparable over all the four rounds considered here.

We present our findings in Table I. Column 1 of Table I shows the rounds considered, column 2 reports the size of total labor force according to the definition of NSS. Third column gives the proportion of females in total labor force. Columns 4–7 reports the value of Kakwani's S_β for $\beta = 0.5$, 1, 1.5 and 2 respectively and the last column gives the values of the *symmetric divergence measure* of Jeffreys (1946). All the measures show oscillatory trend and hence no concrete conclusion can be drawn on the basis of this data. This is disappointing, since the government has implemented many development programmes geared to step-up female participation in the labor force in recent years. On the basis of our findings, there seems to be a need of further improvements in this direction.

The variability in seggregation over time is uniformly much less according to Kakwani's S_β for any β (maximum is about 0.012) than according to the J-divergence, J_n (the difference between maximum and minimum is 0.05). This is due to the fact that J_n is unbounded (the range of values is $(0,\infty)$) whereas S_β is bounded between 0 and 1. For this reason, the variabilities in absolute term are not comparable. But this unboundedness makes the measure J_n more responsive to changes in the occupational distribution over the years than S_β. Also, J_n satisfies symmetry in the unrestricted sense, unlike S_β, which renders a symmetric viewpoint from both the male and female perspectives. This could be an advantage in a world where there is a need to look at gender related issues from the female perspective. Hence, the measure J_n might be more useful for evaluating policy impact than S_β. This also coroborates out theoretical discussion in Section 3.

The second exercise deals with employment data from 12 countries pertaining to the years 1987, 1990 and 1993. The data is again on employment segregation by major occupational groups according to ILO. The source of data is Year Book of Labor Statistics (1994) and Sarvekshana (1992: 94). To make our discussion on occupational seggregation stark, we have chosen 11 developed countries from different continents (to get a small geographically representative sample) and India. The developed countries are chosen to analyze validity of the popular belief that gender discrimination is lower in developed countries. The figures computed for those can be compared with that for India. The data is from labor force sample surveys and general household sample surveys. It covers all status groups, not only wage and salaried employees, but also employers, own account workers, unpaid family workers and members of producer's cooperatives.

We present the findings in Table II where the first column shows the names of the countries. Columns 2–4 reports the values of the Jeffreys' divergence measure for the years 1987, 1990 and 1993 respectively.

The results turn out to be quite interesting. It shows that among the 11 countries chosen other than India, 4 countries, namely USA, Canada, Holland and Sweden show a downward trend in the values of J_n for the 3 years under consideration which is quite encouraging. Holland shows the highest rate of decrease. But on the other hand, we have 5 countries including highly developed Japan

TABLE II

Jeffreys' occupational segregation measure J_n for 12 selected countries over the years 1987–1993 (data from ILO Year Book of Labor Statistics)

Country	1987	1990	1993
Canada	1.0367	1.0094	0.9632
U.S.A.	0.8799	0.8449	0.819
Hong Kong	0.2639	0.4627	0.6566
Israel	0.9032	0.9552	0.988
Japan	0.3168	0.3352	0.3496
Rep. of Korea	0.2149	0.2282	0.322
Austria	1.0704	1.0408	1.1056
Spain	0.7144	0.8024	0.852
Holland	1.0144	0.9784	0.92
Sweden	0.784	0.7658	0.7147
Australia	0.9856	1.0094	1.0003
India	0.2596	0.2013	0.2729

and the fast developing Hongkong and Republic of Korea showing unambiguously increasing trend in gender segregation over the period considered. This is very surprising since labor mobilization is expected to become more egalitarian as the country becomes more developed. Our analysis throws some doubt on this belief. Hongkong shows a rapid increase in the measure J_n suggesting very little effort on the part of the government towards equalizing occupational patterns among sexes. For the other 2 countries and India the trend is oscillatory and no conclusion can be drawn based on this analysis. (In fact, for India, the trend is always oscillating as revealed by Tables I and II.)

Another quite interesting and surprising conclusion is that, almost uniformly, the value of J_n for the chosen developed countries is higher than that for India. (Only Rep. of Korea has a lower value for the years 1987 and 1990.) This is at variance with the popular belief stated above.

Admittedly our analysis of occupational seggregation across gender is an incomplete one to the extent that we have not looked at a larger sample of countries taking into account different levels of development and income. Our purpose was just to illustrate the performance of our measure using real life data, so we did not attempt any elaborate exercise. For an excellent analysis on this issue and relating seggregation to development, industrial structure and inequality, the interested reader is referred to Semyonov and Jones (1999).

6. CONCLUSION

Gender-wise occupational segregation has been an important topic of study for social scientists for some time. Over time, many indices of segregation have been proposed in order to measure this phenomenon. But none of them are symmetric between the two sexes. This paper suggests and characterizes a symmetric measure of occupational segregation which was originally proposed in the literature by Jeffreys (1946) in a different context. We Also provide two numerical illustrations using NSS data in the Indian context and ILO data for 11 selected countries from different continents. The calculations based on the Indian data shows a discouraging picture in terms of decreasing gender-wise segregation over the years. The ILO data also gives a similar result in that the developed countries show an increasing trend in occupational segregation. But for some countries the trend is encouraging. Also the comparative performance of the developed countries and India, a developing country, is quite surprising.

NOTE

[1] A function $f(x, y)$ is said to satisfy the triangle inequality if $f(x, y) + f(y, z) \geq f(x, z)$.

REFERENCES

Aczél, J. and Z. Daróczy: 1975, On Measures of Information and their Characterization (Academic Press, New York).

Butler, R. J.: 1987, 'New indices of segregation', Economics Letters 24, pp. 359–363.

Chakravarty, S. R. and J. Silber: 1992, 'Employment segregation indices: An axiomatic characterization', in W. Eichhorn (ed.), Models and Measurement of Welfare and Inequality (Springer-Verlag, New York).

Duncan, O. D. and B. Duncan: 1955, 'A methodological analysis of segregation indices', American Sociological Review 20, pp. 210–217.

Hutchens, R. M.: 1991, 'Segregation curves, Lorenz curves and inequality in the distribution of people across occupations', Mathematical Social Sciences 21, pp. 31–51.

Jeffreys, H.: 1946, 'An invariant form for the prior probability in estimation problems', Proc. Roy. Soc. London, Ser. A 186, pp. 453–461.

Johnson, R. W.: 1979, 'Axiomatic characterization of the directed divergence and their linear combinations', IEEE Transactions on Information Theory IT-25, pp. 709–716.

Kakwani, N. C.: 1994, 'Segregation by sex: Measure and hypothesis testing, research on economic inequality', JAI Press Inc., pp. 1–26.

Kannappan, Pl.: 1972, 'On Shannon's entropy, directed divergence and inaccuracy', Z. Wahrscheinlichkeitstheorie Verw. Geb. 22, pp. 95–100.

Kannappan, Pl.: 1972a, 'On directed divergence and inaccuracy', Z. Wahrscheinlichkeitst heorie Verw. Geb. 25, pp. 49–55.

Kannappan, Pl., P. K. Sahoo and J. K. Chung: 1993, 'On a functional equation associated with the symmetric divergence measures', Utilitas Mathematica 44, pp. 75–83.

Karmel, T. and M. Maclachlan: 1988, 'Occupational sex segregation – increasing or decreasing?', Economic Record 64, pp. 187–195.

Kullback, S.: 1957, Information Theory and Statistics (Dover, New York).

Lewis, D.E.: 1982, 'The measurement of the occupational and industrial segregation of women', Journal of industrial Relations 24, pp. 406–423.

Moir, H. and J. Selby-Smith: 1979, 'Industrial segregation in the Australian labor market', Journal of industrial Relations 21, pp. 281–291.

Sarvekshana: 1990, Statistical Tables, Special Issue (September).

Sarvekshana: 1992, Statistical Tables.

Sarvekshana: 1995, Statistical Tables.

Semyonov M. and F. L. Jones: 1999, 'Dimensions of gender occupational differentiation in seggregation and inequality: A cross-national analysis', Social Indicators Research 46, pp. 225–247.

Silber, J.: 1989, 'On the measurement of employment segregation', Economics Letters 30, pp. 237–242.

Silber, J.: 1992, 'Occupational segregation indices in the multidimensional case: A note', The Economic Record (September) 68(202), pp. 276–277.

Year Book of Labor Statistics: 1994, International Labor Organization.

Indian Statistical Institute
Kolkata
India