The Analysis of Sex Segregation; or When is Index Measurement not Index Measurement?

MARTIN WATTS
Department of Economics,

University of Newcastle,

Callaghan,

New South Wales,

Australia 2308.

Tel.: 61 2 4921-5069. Fax: 61 2 4921-6919.

Email: ecmjw@cc.newcastle.edu.au

Abstract

In their paper in this issue, Grusky and Charles (1998) make a number of dubious claims about the measurement and interpretation of sex segregation. First, they are incorrect in claiming that only log odds measures yield margin free measures of segregation. Second, the estimation and testing of a limited class of log-linear models does not provide an independent test of the appropriateness of a log odds ratio index to measure segregation. Grusky and Charles'estimation informs them of the statistically justifiable degree of occupational disaggregation, and not whether a log odds ratio is superior to say a linear index in the measurement of segregation. Finally, their index A is beset with problems of interpretation, not withstanding their arguments, and their additional measures, A_W, A_B suffer similar problems.

The authors are correct in arguing that measurement procedures should be margin free. Further, I concur with the view that the adoption of a single annual summary measure of segregation cannot be justified, because it is premised on the assumption that individual occupations, or groups of occupations, exhibit similar trends in sex segregation.

The Analysis of Sex Segregation; or When is Index Measurement not Index Measurement?

Introduction

In their paper in this issue, Grusky and Charles (1998), hereafter GC (1998), reject the computation of the Ip index advocated by Watts (1998). They recommend the use of a log odds ratio to measure changes in the pattern of segregation over time or across countries, based on adoption of log-linear modelling. They imply that their estimation of log linear models constitutes an independent test of the appropriateness of their different measures of segregation.

In this reply, I argue that the decomposition of the Ip index is margin free and possesses some desirable properties for measuring changes in segregation, both in aggregate and by occupational group. Second, the estimation of log-linear models does not provide an independent test of the appropriate measure of segregation, but, at best, the appropriate degree of occupational disaggregation. Third, the indices derived from the parameter estimates of the log-linear models by GC have undesirable properties.

There are two fundamental areas of agreement:

First, it is essential that margin free measures of segregation are employed, otherwise the interpretation of differences across countries or changes over time is impossible. In a number of papers relating to the measurement and empirical analysis of trends in segregation, some of which are cited by GC, I have relentlessly argued this point. Thus the repeated claim by GC (eg. p.1,

p.2, p.3 and p.12) about the importance of measures being margin free is unnecessary.

Second, the use of simple aggregate measures of occupational segregation is based on the premise that 'universal segregative and integrative forces dwarf occupation specific forces' (Weeden 1998, p.4), so that changes in the summary measure are assumed to adequately capture the complexity of changes across groups of occupations (see also GC, 1998, p.11).

I have always expressed dissatisfaction with the use of a summary annual index to measure complex patterns of change over time (see, for example, Watts 1992, Watts 1998). This view underpinned my decomposition of aggregate movements of the Ip index into the contributions by occupational group, both to the level and rate of change of segregation in empirical research that commenced in 1992 (eg Watts and Rich, 1992, 1993 and Watts, 1995). Further we have argued for the distinction to be made between part-time and full-time employment in studies of segregation (Watts and Rich, 1991, 1993).

After claiming that D inspired indices, such as Ip, cannot replace complex qualitative differences in segregation, GC (1998, p.12) state in their conclusion that 'We cannot imagine that Watts truly wishes to suppress such results and rely exclusively on summary measures'. This statement, among many others in the paper, is at best misleading and at worse quite mischevious by completely misrepresenting my position.

Methodology

In their paper, GC are not clear about their objectives for conducting research on segregation. By using an index, I am attempting to measure changes in the pattern of segregation over time across Occupational Groups, differentiated by skill, in a systematic manner, without sacrificing the integrity of the detailed occupation data. Of critical importance in measuring change is devising an index that has desirable properties. From an economic/social policy perspective, the rates of integration across these different groups may provide an indication of barriers to entry which men or women face trying to enter atypical occupations. This empirical analysis can be complemented by case studies to assist in policy design.

CG prefer 'the high ground of elaborating a generic modelling approach' to participating in 'index wars'' (p.4). This statement is merely rhetoric in that they are not prepared to concede that their interpretation of changes in the data does ultimately rely on index measurement, which should be subjected to careful scrutiny.[i] They claim that this approach allows researchers to test and reject scalar measures of sex segregation (1998, p.1). Which scalar measures of segregation are tested within the general modelling framework?

The advocacy of the log odds ratio cannot be justified, by reference to log-linear modelling. In their empirical work the authors investigate a narrow class of log-linear models differing in the functional form of the exponent, which reflects different groupings of occupations. The indices used depict the levels of segregation across occupations in the different specifications are functions of the corresponding estimated scale effects and take the same general log form, but necessarily reflect the grouping of occupations.[ii]

The indices are not pre-determined by the 'correct' statistical specification of the segregation process. The log-linear specification reflects the research interests of the investigator. The measure of segregation which is adopted must be independently justified.

There are a large number of specifications, linear and logarithmic that could be adopted to capture the pattern of segregation. GC make no effort to estimate and explore the statistical properties of different specifications, but seem to be wedded to a log-linear approach and hence a particular form of measurement of segregation. This position is only sustainable if the associated measures of segregation have desirable properties. Their log linear modelling is further explored below.

Properties of the IP Index

GC (1998) make a number of assertions about the properties of the IP index, which they fail to substantiate. They claim that the KM decomposition does not yield a conventional margin free measure of change (p.2). The numerical transformation of the period one distribution of employment by sex and occupation yields a distribution of employment by sex and occupation that has the same occupational shares and overall sex composition. The difference between the indexes, which is the numerator of the expression for the Composition Effect is margin free.[iii] Similarly the computation of Composition Effects for the Occupational Groups are based on margin free comparisons.

I plead guilty to the charge of interpreting the IP values for Occupational Groups as measures of intrinsic segregation (GC, 1998, p.3). It should be noted, however, that, strictly speaking, there is no meaningful way to compare rigorously the levels of segregation across different groups of occupations. The numbers of occupations within these groups may be unequal. What benchmark should the sex composition of the occupations be measured against? Certainly no margin free comparison can be made with 2 occupational groups.

GC (1998, p.10) generate segregation profiles by major occupation for the different countries, based on scale values ϕ_{ck}. Are these scale values strictly comparable, given the unequal numbers of detailed occupations in the major occupations?

My main focus is to use the procedure to enable the margin–free decomposition of trends in segregation for all occupations together and occupational groups, as noted by GC (1998, p.3). This overcomes the problem of using single summary measures. I concur with GC's view that "It is high time that advocates of particular indices are held accountable for the data reduction that their indices imply" (1998, p.5).

Log-Linear Modelling

Having rejected any form of decomposition procedure, the authors claim that log-linear econometric estimation is required and that the only index with the desirable properties is one based on the log of the odds ratio (p.2).[iv]

A summary index, A can be derived from the saturated log-linear model (GC 1998 equation 2, p.5) where

$$A = \exp\left\{(1/J)\sum_{j=1}^{J}[\ln(F_{jk}/M_{jk}) - \{(1/J)\sum_{j=1}^{J}\ln(F_{jk}/M_{jk})\}]^2\right\}^{1/2} \quad (1)$$

and F_{jk}, M_{jk} denote the respective number of females and males in the jth occupation in time period k and J denotes the total number of occupations.

This measure allows for qualitative variability in the underlying structure of segregation (CG, 1998, p.5). Certainly its individual (occupation) components differ, as is the case in the computation of all indices. Can this measure inform the researcher about the pattern of segregation?

The index is characterized by Occupations Invariance because the occupations are treated as being of equal size. This is justified by the assertion that occupations, not individuals, are the unit of analysis, so that the relative size of occupations is unimportant (GC, 1998, p.6; see also Weeden, 1998, p.24).

Weeden does not maintain a consistent position on this issue. He computes measures based on A and then notes that, due to the extreme values of the log of sex ratios across some occupations, the measure exhibits volatility (see also Watts, 1998, pp.17-19). He then undermines his advocacy of occupations as the basic unit of analysis by aggregating the smaller occupations by employment together and notes the measure exhibits less volatility (p.25). Elsewhere he qualifies the results by noting that the volatility of the index is a consequence of changes in occupations that are highly segregated (Weeden, 1998, p.37).[v]

It is acknowledged by all researchers that an integrated distribution of employment by sex is represented by a uniform odds ratio F_{jk}/M_{jk} across all occupations for a given context, k. The index A takes the value zero if the sex distribution of employment is integrated. The logical benchmark by which to measure the extent of sex segregation is then the uniform odds ratio F/M, but the index A uses the mean of the logs of the odds ratios across occupations as the benchmark for the computation of the deviation from the benchmark. This is somewhat odd since it is determined by the actual sex distribution of employment across occupations.[vi] For this reason, my offer of a compromise division of labour (GC 1998, p.8) is withdrawn!

The problem of cells with zero entries is not overcome by ad hoc estimation procedures to fill the cells. An occupation which has a history of (fe)male dominance and in a particular year has a zero female (male) entry will end up with an estimated female (male) figure which continues to signify (fe)male dominance. This occupation will disproportionately influence the magnitude of the index. Also, the replacement of a zero entry by estimation procedures may introduce a distortion if the zero entry is correct, rather than indicating the vagaries of sampling.

GC (1998, p.4) emphasise the superiority of their modelling approach and the derivative measures, asserting that the critique which is directed at their index A cannot be convincingly generalized to their larger approach within the log-linear framework.

The definitions of the two important summary indices, A_W and A_B are relegated to a footnote (p.6), presumably to avoid scrutiny, so there is no discussion of their properties. Given the functional form of these measures, it is disingenuous to claim that these measures are somehow immune from similar criticisms to those levelled at A.

These measures are 'based' on a simple multi-level model of the form

$$m_{ijk} = \alpha_k \beta_{ik} \gamma_{jk} e^{z_i v_{jk} + z_i \phi_{ck}}$$

It can be shown that the summary indices take the form:

$$A_w = \exp\left\{ (J_C / J) \sum_C (1 / J_C) \sum_{j \in C} [\ln(F_{jk} / M_{jk}) - \{(1 / J_C) \sum_{j \in C} \ln(F_{jk} / M_{jk})\}]^2 \right\}^{1/2}$$

and

$$A_B = \exp\left\{ (1 / C^*) \sum_C [(\sum_{j \in C} \ln(F_{jk} / M_{jk}) / J_C) - \{(1 / J) \sum_{j=1}^{J} \ln(F_{jk} / M_{jk})\}]^2 \right\}^{1/2}$$

where C denotes a major occupational category with J_C occupations and there are C* major occupational categories.[vii]

The interpretation of these measures is beset with difficulty. A_w is measuring the deviation of the log odds ratios across occupations from their respective means within the corresponding major occupational categories. The mean ratio employed in the equation is flawed. As noted above, this benchmark is determined by the female/male ratios across individual occupations within the major occupational categories, rather than by the overall female/male ratio.[viii]

A reduction in the measure, associated with a major occupational category, could be accompanied by an increase in the absolute difference between the overall female share of employment and the female share of this group of occupations and a rise or fall in the overall index. Thus, a reduction in

occupational segregation within a major occupational category may not be synonymous with lower economy-wide segregation.

For example, the observation that the gross level of segregation within Craft occupations had fallen, in the light of say a decline in the female share of employment in these occupations to less than 10%, would be of little significance, if the overall (economy wide) female share of employment was increasing. It would merely indicate that female shares of employment across the occupations within this major occupational category had become less dispersed.[ix]

I accept that if it can be established statistically that a certain level of occupational disaggregation is appropriate, then the argument about Organizational Equivalence loses some force (GC, 1998, p.6), but, as noted above, I reject the testing procedure used by CG, because the benchmarks used, namely major occupation 'means' are inappropriate.

Certainly, if occupations are grouped according to whether they are female dominated or male dominated, the overall Ip index will be unchanged. In general, however, such a grouping will make little sense. Again this points to the adoption of occupational groupings based on skills, rather than statistical significance, and, of course, the use of a measurement procedure that does not sacrifice the integrity of the detailed occupational data.

GC (1998, p.10) argue that 'all segregation indices, conventional or otherwise, are highly flawed for the present data' because micro-level estimates are widely scattered. Taken at face value, this suggests that it is unlikely that a commonality will exist at the level of major occupational groups in other

empirical studies. Consequently, the problem of Organizational Equivalence will continue to haunt their work.

In addition, their statement highlights the underlying methodological differences between our respective approaches. My project is not to find statistically similar patterns of segregation within groups of detailed occupations, which would appear to be a thankless task, but rather to document rigorously patterns of change.

Conclusion

The authors also accuse me of forgoing all pretense of independent conceptualization in the adoption of the Ip index and its decomposition for the study of trends in sex segregation. Nowhere do the authors explain their conception of what segregation is and what properties should characterize its measurement. Their statistical approach claims to differentiate statistically between different degrees of occupational aggregation within a narrow class of log-linear models, but the general form of their index measurement is influenced by their choice of a log odds ratio, despite its undesirable properties. What is the traditional measure of segregation that GC claim to operationalize faithfully (p.12)? Further, it is hard to justify the supplementary use of D (GC (1998, p.12), even for continuity with past work, when it is flawed as a measure and misrepresented by most researchers.

The use of the Ip index with occupations differentiated into Occupational groups represents a coherent and integrated approach to the measurement of occupational segregation. On the other hand, any measure founded on the log

of the odds ratio is subject to the vagaries of extreme values and the assumption that all occupations should be weighted equally.

Grusky and Charles may be seeking the high ground, but (index) measurement remains central to their analysis of segregation and these indices are fundamentally flawed.

References

Cortese C.F., Frank, R.F. & Cohen, J. (1976). "Further Considerations on the Methodological Analysis of Segregation Indices". *American Sociological Review*, 41, 630-637.

Grusky, D.B. and Charles, M. (1998) "The Past, Present and Future of Sex Segregation Methodology", *Demography* ****

Watts, M.J. and J. Rich (1991). "Equal Employment Opportunity in Australia? The Role of Part-time Employment in Occupational Sex Segregation", *Australian Bulletin of Labour*, June, 160-79.

Watts, M.J. and J. Rich (1992). "Labour Market Segmentation and the Persistence of Occupational Sex Segregation in Australia", *Australian Economic Papers*, 31(58), June, 58-76.

Watts, M.J. and J. Rich (1992) "Occupational Segregation in Britain, 1979-89: The Role of Part-Time Employment" *International Review of Applied Economics*, 6(3), September, 286-308.

Watts, M.J. and J. Rich (1993). "Occupational Sex Segregation in Britain, 1979-89: The Persistence of Sexual Stereotyping", *Cambridge Journal of Economics*, 17(2), June, 159-177.

Watts, M.J. (1994) "A Critique of Marginal Matching", *Work, Employment and Society* 8(3), 421-31.

Watts, M.J. (1995) “Divergent Trends in Gender Segregation by Occupation in the USA: 1970-92”, *Journal of Post Keynesian Economics* 17(3), 357-379.

Watts, M.J. (1997) “Multi-Dimensional Indexes of Occupational Segregation: A Critical Assessment”, *Evaluation Review* 21(4), 461-82.

Watts, M.J. (1998) “Occupational Gender Segregation: Index Measurement and Econometric Modelling”, *Demography* ****

Weeden, K (1998) “Revisiting Occupational Sex Segregation in the United States, 1910-1990: Results from a Log-Linear Approach’, *Demography* XX

Endnotes

[i] The claim that participants in the index war are looking for a comprehensive analytic solution to all answers that might possibly be posed (pp.3-4) is exaggerated.

[ii] Weeden (1998, p. 44) is even prepared to group occupations together that have similar patterns of segregation for the purpose of statistical analysis. I would claim that homogeneity with respect to skill is of much greater relevance to modern economic and social policy research.

[iii] GC (1998, p.1) again misrepresent the properties of the Index of Dissimilarity by asserting that it measures the proportion of the male, female, or total labor force that requires reallocation to 'produce an even distribution' (White , 1985, p.202). Watts (1998, p.9) notes that this error is common amongst researchers of gender segregation and outlines the correct interpretation of the D index by reference to Cortese, Frank and Cohen (1976, p.634-35). Further, it is not clear how the D index and 'its many cousins, such as IP' can measure the same magnitude.

[iv] Their multiplicative shift model (equation 1) includes the term ϕ_k as one part of the exponent, yet the indicator variable for gender, Z_1 takes the value 0, although it is arbitrary. For this value the exponential term is unity, so it is unclear how ϕ_k can represent a multiplicative shift, which is not uniform by gender.

[v] In a footnote, Weeden (1998, p.38) argues that the sensitivity of the index is only a problem, if individuals are defined as the unit of analysis or changes arise from sampling variability.

[vi] Weeden (1998, p.22) argues incorrectly that it measures 'the typical amount that the occupation sex ratios deviate from perfect integration, where perfect integration is defined by an identical (logged) sex ratio in each occupation'.

[vii] These derivations require that the sum of the ϕ_{ck}'s each weighted by the corresponding number of occupations in the major occupational category is zero, not the simple sum of these macrolevel scale variables (cf. GC, 1998, footnote 4, p.13).

[viii] In a similar vein, Silber (1989a, pp.111-12) shows that the overall Gini coefficient can be decomposed into three distinct components based on population sub-groups, namely the intra-group Gini, the inter-group Gini and an interaction term. Again the computation of the intra-group Gini coefficient is not measured with respect to an external yardstick. Thus the decompositions of the Gini coefficient are not suitable for the analysis of patterns of change in the extent of gender segregation across major occupational categories (see Watts, 1997).

[ix] Rubery (1988) and Figart and Mutari (1993), amongst others, examine movements in gender segregation within particular groups of industries or occupations separately, using D, but see the critique by Watts (1994, p.426-27).