Brazilian Review of Econometrics. 2022: Economics of Family and Gender, v. 42 n. 2: e0120221

DECOMPOSING GENDER SEGREGATION[1]

Diego Camargo Botassio
Ministry of the Economy, Brazil, and PSP Hub.

Rodolfo Hoffmann
Department of Economics, Administration and Sociology, University of São Paulo, Brazil.

December 14, 2022

Abstract

We provide insights into the class of additively decomposable segregation measures regarding its decomposition. We discuss the implications of choosing different parameters for the decomposition analysis. The within and between-groups terms are independent only when the parameter equals zero or one. In empirical works, these values should be preferred, especially if the between-groups component has a high value. Furthermore, the term "democratic" is appropriate for the decomposition when the measure's parameter equals one-half. We illustrate the use of this class of measures in analyzing the evolution of gender decomposition, considering various parameter values, through a household sample survey.

Keywords: Segregation; Gender segregation; Additively decomposition; Segregation measurement.

JEL: C63; J01; J16; J82.

1. Introduction

Gender segregation is a significant aspect of studies in labor economics. Indexes are utilized to measure the level of gender segregation in occupations predominantly held by women and men. An interesting approach to this topic is the decomposition of its measure. First, occupations are classified into groups according to a criterion. Then, the segregation decomposition consists in determining how the overall measure can be decomposed into two parts: one regarding the within-groups differences and the other regarding the segregation between groups of occupations.

The concept of segregation decomposition, initially introduced by Theil and Finizza (1971), has been analyzed by several authors, including Mora and Ruiz-Castillo (2003), Hutchens (2004) and Frankel and Volij (2011). Mora and Ruiz-

Diego Camargo Botassio: E-mail: diegocbotassio@gmail.com
Rodolfo Hoffmann: E-mail: hoffmannr@usp.br

[1] We thank the financial support from the Brazilian National Council for Technological and Scientific Development (CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico).

 https://doi.org/10.12660/bre.v42n22022.88798

Castillo (2003) contribute to the theoretical and empirical literature developing an additively decomposable segregation index building upon the ideas by Theil and Finizza (1971) and Fuchs (1975) and analyze the gender segregation in Spain. Frankel and Volij (2011) make important remarks on the theoretical properties of segregation decomposition.

In this paper, we analyze the decomposition of the class of additively decomposable segregation proposed by Hutchens (2004). This class of measures is a function of a real-valued parameter and its choice plays an important role in the decomposition properties. We highlight that the segregation decomposition characteristics using this class of measures have not been deeply discussed in the literature and have neither been explored in empirical works. This paper aims to address these gaps in segregation literature, significantly important for applied economists and sociologists who are interested in studying the evolution or decomposition of segregation by gender, skin color, etc.

To achieve our goals, we use a mathematical approach that is usual in this field (Hutchens 1991, 2004; Mora and Ruiz-Castillo 2003; Chakravarty and Silber 2007; Frankel and Volij 2011; Botassio and Hoffmann 2020). This approach enables us to contribute rigorously and comprehensively to the extensive existing literature. Furthermore, this formalism makes it possible to achieve nontrivial/novel conclusions, that are hard or impossible to be made with a non-mathematical approach.

In addition to this introduction, in section 1 we present the notation and the class of additively decomposable segregation measures. In section 2 we make the considerations regarding the decomposition coefficients (or weights) of this class, highlighting some important cases. Section 3 illustrates how the class of additively decomposable segregation measures can be used to analyze the segregation decomposition evolution in empirical works using data from a household sample survey. Section 4 summarizes the conclusions of the paper.

2. Definitions and the class of segregation measures

Consider an economy with female and male workers, represented by F and M, respectively, divided into J occupations. If F_j represents the number of women in the jth occupation, the women's share in the jth occupation is f_j, with $j = 1,2, \ldots, J$ ($f_j = F_j / F$). The values m_j and M_j are similarly defined for the men. Let $\mathbf{D}$ be the $2 \times J$ matrix distribution of these genders into the J occupations, and $D_J = \{\mathbf{D} \in \mathrm{R}_+^{2\times J}: F, M > 0\}$ be the set of all possible matrices with J occupations. Denote by $D = \cup_{J\geq 2} D_J$ the set of all feasible distributions with $J \geq 2$. Then, a *segregation measure* is a function $\Omega: D \rightarrow \mathrm{R}_+$ that maps a $\mathbf{D} \in D$ to a non-negative real number.

Proposed by Hutchens (2004), the *class of additively decomposable segregation measures*[2] is

$$I_\varepsilon(\mathbf{D}) = \frac{1}{\varepsilon(1-\varepsilon)}\left[1 - \sum_{j=1}^{J} m_j^\varepsilon f_j^{1-\varepsilon}\right] \tag{1}$$

where ε is a real-valued parameter. Some special cases of (1) are Theil's T and L indexes, which are obtained by L'Hospital's rule for $\varepsilon \to 0$ and $\varepsilon \to 1$, respectively. These expressions are $I_0(\mathbf{D}) = \sum_{j=1}^{J} f_j \ln(f_j/m_j)$ and $I_1(\mathbf{D}) = \sum_{j=1}^{J} m_j \ln(m_j/f_j)$. There are two increasing transformations often related to $I_\varepsilon(\mathbf{D})$. The first one is Atkinson's family of measures for segregation (James and Taeuber 1985), which is $A_\varepsilon(\mathbf{D}) = 1 - [1 - \varepsilon(1-\varepsilon)I_\varepsilon(\mathbf{D})]^{1/(1-\varepsilon)}$ for $0 < \varepsilon \neq 1$, and $A_1(\mathbf{D}) = 1 - \exp[-I_1(\mathbf{D})]$. The second is the coefficient of variation for segregation (Hutchens 1991, p. 48), whose expression is $CV(\mathbf{D}) = \sqrt{2I_{-1}(\mathbf{D})}$. The properties of $I_\varepsilon(\mathbf{D})$ are analyzed by Hutchens (2004), Frankel and Volij (2011), and Botassio and Hoffmann (2020).

The class $I_\varepsilon(\mathbf{D})$ or the square of $CV(\mathbf{D})$ are additively decomposable, that is, are measures that can be decomposed into two parts: one regarding the segregation between- groups of occupations and the other the segregation within-groups. The next section explores this feature and shows how the parameter's choice affects the segregation decomposition.

3. The segregation decomposition

The idea of decomposing segregation is very simple. Imagine occupations classified in groups according to a criterion (e.g., the International Standard Classification of Occupations - ISCO-08 - which divides occupations into 10 groups).[3] The subscript g denotes a group of occupations and takes values 1,2, ..., G. If n_g indicates the number of occupations in the gth group in an economy with J occupations divided into G groups, we must have $\sum_{g=1}^{G} n_g = J$. From (1) the overall segregation in this economy is

$$I_\varepsilon(\mathbf{D}) = \frac{1}{\varepsilon(1-\varepsilon)}\left[1 - \sum_{g=1}^{G}\sum_{j=1}^{n_g} m_{gj}^\varepsilon f_{gj}^{1-\varepsilon}\right] \tag{2}$$

with $m_{gj} = M_{gj}/M$ and $f_{gj} = F_{gj}/F$, where M_{gj} and F_{gj} are the totals of men and women in the jth occupation.

[2] In his article, Hutchens (2004) uses the transformation $H(\boldsymbol{D}) = \varepsilon(1-\varepsilon)I_\varepsilon(\boldsymbol{D})$ for $0 < \varepsilon < 1$, which is referred to as 'a generalized entropy measure of segregation'. Expression (1) is presented in Hutchens' (2004, p. 574) paper with a slightly different notation considering $c = 1 - \varepsilon$.

[3] The website of the International Labour Organization presents in details this classification. See: https://www.ilo.org/public/english/bureau/stat/isco/index.htm.

The segregation within the gth group, according to (1), is

$$I_{\varepsilon g}(\mathbf{D_g}) = \frac{1}{\varepsilon(1-\varepsilon)}\left[1 - \sum_{j=1}^{n_g}\left(\frac{M_{gj}}{M_g}\right)^{\varepsilon}\left(\frac{F_{gj}}{F_g}\right)^{1-\varepsilon}\right]. \quad (3)$$

M_g and F_g indicate the number of men and women in the gth group. $\mathbf{D}_g$ is a $2 \times n_g$ sub-matrix of $\mathbf{D}$ which includes only information regarding the n_g occupations in the gth group. Define by $f_g = F_g/F$ and $M_g = M_g/M$ the shares of each gender in the gth group. Adding and subtracting $\sum_{g=1}^{G} m_g^{\varepsilon} f_g^{1-\varepsilon}$ inside the brackets in (2), after some algebra, we get

$$I_{\varepsilon}(\mathbf{D}) = \sum_{g=1}^{G} w_g(\mathbf{D_g}) I_{\varepsilon g}(\mathbf{D_g}) + B_{\varepsilon}(\mathbf{D}), \quad (4)$$

where $w_g(\mathbf{D_g}) = m_g^{\varepsilon} f_g^{1-\varepsilon}$ is the weight attached to group g, $I_{\varepsilon g}(\mathbf{D}_g)$ is the segregation within the gth group given by (3), and $B_{\varepsilon}(\mathbf{D})$ is the between-groups term, which is

$$B_{\varepsilon}(\mathbf{D}) = \frac{1}{\varepsilon(1-\varepsilon)}\left[1 - \sum_{g=1}^{G} m_g^{\varepsilon} f_g^{1-\varepsilon}\right]. \quad (5)$$

Equation (4) is the additively decomposition of $I_{\varepsilon}(\mathbf{D})$. The left-hand side of (4) is the *overall segregation*, and the first and second terms on the right-hand side are, respectively, the *within* and *between-groups components*. Using (5) and the expression of the weights, it can be deduced that the sum of those weights is

$$\sum_{g=1}^{G} w_g(\mathbf{D_g}) = 1 - \varepsilon(1-\varepsilon) B_{\varepsilon}(\mathbf{D}). \quad (6)$$

Considering $\varepsilon = 0$ or 1 there are two considerations to be made from (6) that seem obvious, but their implications are important. First, $\varepsilon = 0$ or 1 is sufficient to ensure that the sum of the weights is one, which means that the within-groups component is a weighted average (and not only a weighted sum) of the within-groups segregation measures. Second, and most importantly, for $\varepsilon = 0$ or 1, the within-groups term is independent of the between-groups component. In Theil's (1967, p. 125) words, "one should prefer a measure for which the within-set components, including their weights, are independent of the between-set component". If one agrees with this statement, using $\varepsilon = 0$ or 1 is preferable to all other values. This shows us the generality and potential of expression (1). Note that if $\varepsilon = 0$ or 1, the within-groups component is affected only by the shares of one gender in the groups (weights f_g or m_g), but not by the between-groups component.

Another implication is on the sum of the weights for ε other than zero or one. Since $\varepsilon(1 - \varepsilon) > 0$ if $\varepsilon \in (0,1)$, we have $\sum_{g=1}^{G} w_g(\mathbf{D_g}) \leq 1$ whenever $0 < \varepsilon < 1$

according to (6). On the other hand, if $\varepsilon > 1$ or $\varepsilon < 0$, then $\sum_{g=1}^{G} w_g(\mathbf{D_g}) \geq 1$. Theil (1967, p. 125) noticed this last fact specifically for the square of the coefficient of variation for income. Considering $\varepsilon = -1$ in the segregation case, for instance, we have $\sum_{g=1}^{G} w_g(\mathbf{D_g}; \varepsilon = -1) = 1 + 2B_{-1}(\mathbf{D})$, where $2B_{-1}(\mathbf{D})$ is equal to the squared between-groups coefficient of variation for segregation. This implies that the weighting in the within-groups component is larger when the between-groups term is higher. So, again, considering $\varepsilon = 0$ or 1 is preferable [especially if $B_\varepsilon(\mathbf{D})$ has a high value] because they guarantee that both components in the decomposition are independent. For the cases where $\varepsilon \neq 0,1$, the sum of the weights is positive and it is equal to one only if $B_\varepsilon(\mathbf{D}) = 0$, which means that the overall segregation is a weighted average of the within-groups segregation measures.

In the context of measuring inequality, Theil's L index is a democratic measure, because the weights are equal to the participation of the groups in the population (Bourguignon 1979). On the other hand, the "democratic" denomination is not appropriate for Theil's L index for segregation. Let us explain why.

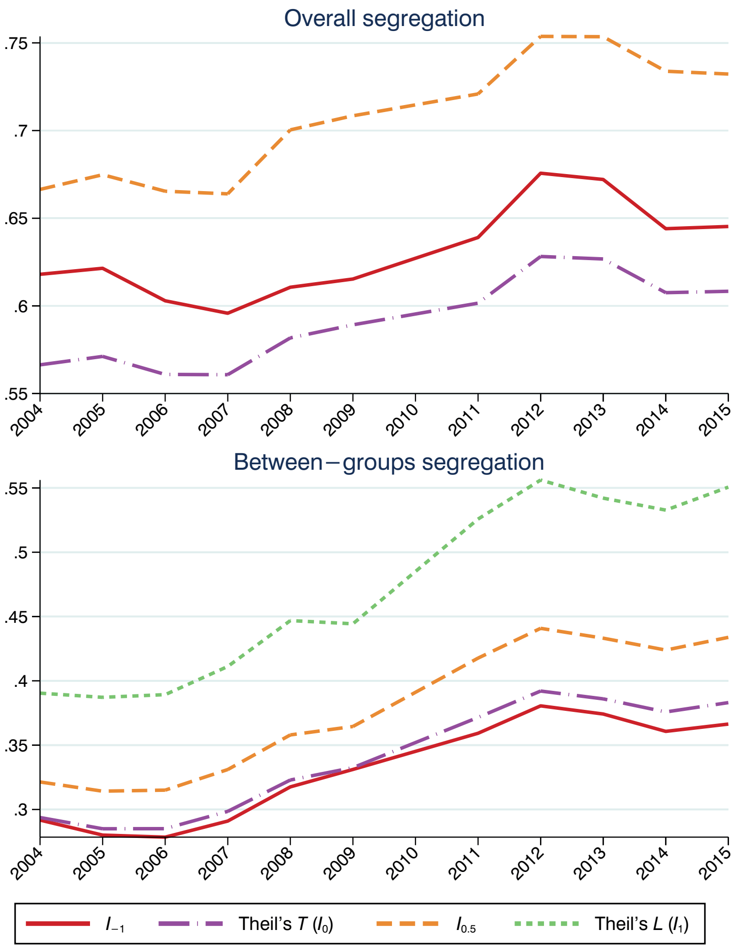
As the participation of the gth group is weighted by $w_g(\mathbf{D_g}) = m_g^{\varepsilon} f_g^{1-\varepsilon}$ in the within-groups component, if we consider the Theil's L index [$I_1(\mathbf{D})$], then the weighting is done by the men's share in the groups (m_g). On the other hand, using the Theil's T index [$I_0(\mathbf{D})$], the weighting is done by the women's share (f_g). In either case, the shares of one type are absolutely discriminated in the weighting process. We can say that it is fairer to use $\varepsilon = 0.5$, since the weights are $w_g(\mathbf{D_g}; \varepsilon = 0.5) = \sqrt{m_g f_g}$ (a geometric average of the gender's share in the group). This result is also valid for Hutchens' (2001, 2004) Square Root Index, since it is equal to $I_{0.5}(\mathbf{D})/4$. Note that the democratic denomination using $\varepsilon = 0.5$ concerns the decomposition analysis, especially the within-groups component. For the overall segregation, this concept should not be confused with the symmetry in types property, which is well analyzed in the literature (e.g., Frankel and Volij 2011; Hutchens 2001, 2004). See, for instance, the analogy with inequality. Theil's L inequality index is a democratic measure because of the weighting of the within-groups component, but there is no similar interpretation regarding the overall inequality.

4. Data and empirical results

For an application of the additively decomposable class of segregation measures, we use microdata from the Brazilian National Household Sample Survey (PNAD in the Portuguese acronym) from 2004 to 2015 (the survey was not conducted in 2010). The survey uses a five-digit occupation classification according to the Brazilian Household Occupations Classification (CBO-Domiciliar). This classification (variable V4010) is used to construct the 48 two-digits occupations classified into 10 one-digit groups of occupations.[4]

Figure 1 presents the evolution of the overall segregation [48 occupations - the left-hand side of (4)] and the between-groups component [10 groups - second term of the right-hand side of (4)]. We consider the index $I_{-1}(\mathbf{D})$ (half of the squared coefficient of variation), Theil's T [$I_0(\mathbf{D})$] and L [$I_1(\mathbf{D})$] indexes, and the measure $I_{0.5}(\mathbf{D})$ (four times the Square Root index).

Figure 1 – Occupational gender segregation decomposition. Brazil, 2004-2015



[4] Botassio and Hoffmann (2019) analyzed gender segregation between 7 groups of activity, while here we consider the classification by type of occupation. See the appendix for the classification of 48 types of occupation in 10 groups.

The overall segregation using Theil's L index is not defined for 2004, 2008, and 2011. That happens because, at the two-digits classification, in the sample there were no women occupied as operators of energy production and distribution facilities in 2004, and there were no women occupied as poly-scientific professionals in 2008 and 2011.[5]

A simple and efficient way to determine the segregation trend is to estimate a linear regression model. We estimate the equation $\ln(measure_t) = \alpha + \beta t + u_t$, where $measure_t$ is the value of a measure at year t and u_t is the error term. In this model, $[\exp(\beta) - 1] \cdot 100\%$ is the average annual segregation growth rate. Table 1 presents these results for the indexes from Figure 1.

TABLE 1
GENDER OVERALL AND BETWEEN-GROUPS SEGREGATION
EVOLUTION (REGRESSIONS). BRAZIL, 2004-2015

Overall segregation - 48 occupations				
	$I_{-1}(\mathbf{D})$	Theil's T	$I_{0.5}(\mathbf{D})$	Theil's L
Trend	0.008*	0.010**	0.012**	-
	(0.002)	(0.002)	(0.002)	-
Constant	–16.974*	–20.161**	–24.131**	-
	(4.703)	(3.626)	(3.766)	-
R-squared	0.552	0.765	0.815	-
Between-groups segregation - 10 groups				
	$I_{-1}(\mathbf{D})$	Theil's T	$I_{0.5}(\mathbf{D})$	Theil's L
Trend	0.030**	0.033**	0.036**	0.039**
	(0.004)	(0.004)	(0.004)	(0.004)
Constant	–61.332**	–66.751**	–72.410**	–78.321**
	(8.282)	(7.606)	(7.615)	(7.345)
R-squared	0.855	0.892	0.907	0.916

Notes: Standard errors in parentheses; * p < 0.01, ** p < 0.001.

Results show that, depending on the value of the parameter ε, the overall segregation increased from 0.8% [for $I_{-1}(\mathbf{D})$] to 1.2% [$I_{0.5}(\mathbf{D})$] per year. On the other hand, the between-groups segregation increased from 3% [$I_{-1}(\mathbf{D})$] to 3.9% (Theil's L index).

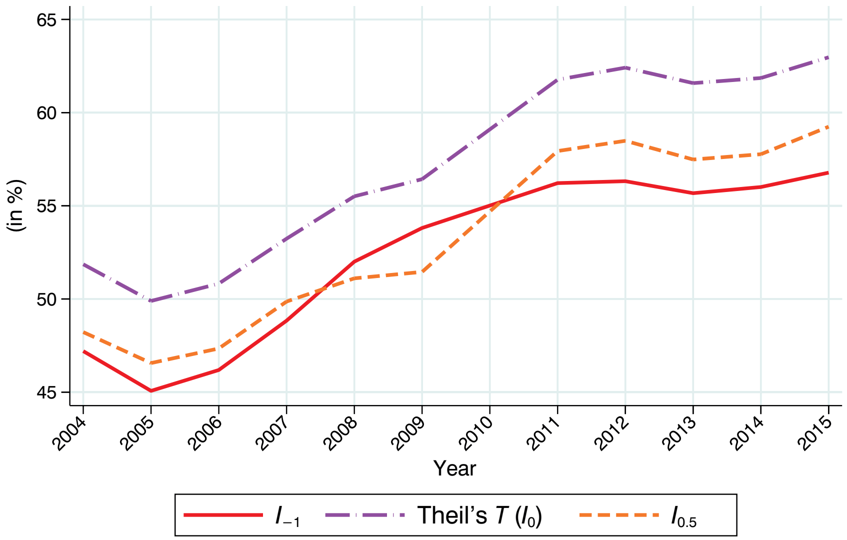
As the overall segregation annual growth rate is lower than the between-groups rate, regardless of the parameter ε, the increasing trend of the overall segregation is due to the increasing segregation between groups (the trend of the within segregation

[5] Theil's L index $[I_1(\mathbf{D}) = \sum_{j=1}^{J} m_j \ln(m_j / f_j)]$ is not defined if there is a $f_j = 0$. For Theil's T index $[I_0(\mathbf{D}) = \sum_{j=1}^{J} f_j \ln(f_j / m_j)]$ we used the $\lim_{f_j \downarrow 0} f_j \ln(f_j/m_j) = 0$.

is decreasing). In other words, segregation considering more homogeneous occupations is decreasing, but the between-groups segregation (more heterogeneous) is increasing at a higher rate. So, the overall segregation is increasing.

An analysis that allows us to verify it is to analyze the between-groups component share on the overall segregation. Figure 2 presents these results.

Figure 2 – Between-Groups Share in the Overall Segregation. Brazil, 2004-2015



In the period, the between-groups share of the overall segregation increased by 20.3% for $I_{-1}(\mathbf{D})$ (9.6 percentage points - p.p.), 21.4% for Theil's T index (11.1 p.p.), and 22.9% for $I_{0.5}(\mathbf{D})$ (11 p.p.).

5. Conclusions

In this paper, we pointed out some properties of the class of additively decomposable segregation measures, regarding the parameter's choice on the segregation decomposition. We conclude that if ε is equal to 0 or 1, then the within-groups component is an average of the within-groups segregations with weights equal to the share of one of the categories in the respective group. In this sense, these values are preferable, because they guarantee that the within-groups component is independent of the between-groups in the decomposition analysis. Further, the denomination of democratic measure is appropriate for $\varepsilon = 0.5$ concerning the decomposition [Hutchens' Square Root index or $I_{0.5}(\mathbf{D})$]. Further, the choice of ε requires a value judgment (Hutchens 2004) that can be associated with the parameter's aversion to segregation (Botassio and Hoffmann 2020).

Also, empirical results using Brazilian labor market data show that both the overall and the between-groups segregation have increased from 2004 to 2015, and the upward trend of the overall segregation is boosted by the between-groups segregation trend.

References

Botassio, D. C., & Hoffmann, R. (2019). Measuring gender segregation: *Evolution and decomposition analysis in the Brazilian labor market*. In: Anais do XXII Encontro de Economia da Região Sul.

Botassio, D. C., & Hoffmann, R. (2020). Measuring gender segregation. *Economic Theory Bulletin, 8(1)*, 25–47. doi:10.1007/s40505-019-00162-8.

Bourguignon, F. (1979). Decomposable income Inequality measures. *Econometrica, 47(4)*, 901–920. doi:10.2307/1914138.

Chakravarty, S. R., & Silber, J. (2007). A generalized index of employment segregation. *Mathematical Social Sciences, 53(2),* 185–195. doi:10.1016/j.mathsocsci.2006.11.003.

Frankel, D. M., & Volij, D. (2011). Measuring school segregation. *Journal of Economic Theory, 146(1)*, 1–38. doi:10.1016/j.jet.2010.10.008.

Fuchs, V. R. (1975). Explorations in Economic Research, Chap. A note on sex segregation in professional occupations, 105–111.

Hutchens, R. (1991). Segregation curves, Lorenz curves, and inequality in the distribution of people across occupations. *Mathematical Social Sciences, 21(1)*, 31–51. doi:10.1016/0165-4896(91)90038-S.

Hutchens, R. (2001). Numerical measures of segregation: desirable properties and their implications. *Mathematical Social Sciences, 42(1)*, 13–29. doi:10.1016/S0165-4896(00)00070-6.

Hutchens, R. (2004). One Measure of Segregation. *International Economic Review, 45(2)*, 555–578. doi:10.1111/j.1468-2354.2004.00136.x.

James, D. R., & Taeuber, K. E. (1985). Measures of Segregation. *Sociological Methodology, 15*, 1–32. doi:10.2307/270845.

Mora, R., & Ruiz-Castillo, J. (2003). Additively decomposable segregation indexes. The case of gender segregation by occupations and human capital levels in Spain. *The Journal of Economic Inequality, 1(2)*, 147–179. doi:10.1023/A:1026198429377.

Theil, H. (1967). Economics and Information Theory. Amsterdam: North-Holland.

Theil, H., & Finizza, A. J. (1971). A note on the measurement of racial integration of schools by means of informational concepts. *The Journal of Mathematical Sociology, 1(2)*, 187–193. doi:10.1080/0022250X.1971.9989795.

Appendix

Table A1
Classification of the 48 Two-Digit Occupations into 10 One-Digit Groups of Occupations

Group (1-digit)	Occupation (2-digits)	Classification
1		**Legislators, senior officials, and managers**
	11	Legislators and senior officials
	12	Corporate managers
	13	General managers
2		**Arts and sciences professionals**
	20	Poliscientific professionals
	21	Physical, mathematical, and engineering science professionals
	22	Life science and health professionals
	23	Teaching professionals
	24	Legal professionals (Lawyers and Judges)
	25	Social science and related professionals
	26	Information and religious professionals, writers, and creative or performing artists
3		**Technicians and associate professionals**
	30	Multipurpose technicians
	31	Physical, engineering, and related science technicians
	32	Life science and health associate professionals
	33	Teaching associate professionals
	34	Ship, road, and aircraft controllers and technicians
	35	Administrative associate professionals
	37	Artistic, entertainment, and sports associate professionals
	39	Other associate professionals
4		**Clerks**
	41	Office clerks
	42	Customer services clerks
5		**Service workers and shop and market sales workers**
	51	Services workers
	52	Salespersons and demonstrators
6		**Skilled agricultural and fishery workers**
	61	Agricultural producers
	62	Agricultural workers
	63	Fishery workers, hunters, and forestry workers
	64	Agricultural and forestry mechanization workers

7		**Craft and related trades workers**
	71	Extraction and building trades workers
	72	Metal, machinery, and related trades workers
	73	Electrical and electronic manufacturing and installation workers
	74	Precision and musical instrument and instrument assemblers
	75	Jewelers, potters, glass-makers, and related trades workers
	76	Handicraft workers in wood, textile, leather, and related materials
	77	Wood treaters, cabinet-makers, and related trades workers
	78	Cross-functional workers
8		**Plant and machine operators and assemblers**
	81	Workers in the continuous process industries and other industries
	82	Steel installations and construction materials workers
	83	Cellulose, paper, cardboard, and artifact manufacturing plant and machinery workers
	84	Food processing and related trades workers
	86	Operators of energy production and distribution facilities, utilities, water collection, treatment, and distribution
	87	Other elementary industrial workers
9		**Maintenance and repair workers**
	91	Mechanical maintenance and repair workers
	95	Multiple maintainers
	99	Other conservation, maintenance, and repair workers
0		**Armed forces, policemen, and firemen**
	01	Aeronautics member
	02	Army member
	03	Navy member
	04	Policemen
	05	Fireman