Implementing the Mutual Information Index in Stata

Ricardo Mora

Universidad Carlos III de Madrid

Madrid, Oct 2013

Outline

Introduction

The Mutual Information Index: M

- M is a multigroup index of segregation first proposed by Theil and Finizza (1971) in the context of racial school segregation.

The Mutual Information Index: M

- M is a multigroup index of segregation first proposed by Theil and Finizza (1971) in the context of racial school segregation.
- Frankel and Volij (2011) prove that M represents the unique nontrivial ordering that satisfies Scale Invariance, Independence, Symmetry, the Group and Unit Division Properties, and a technical continuity property.

The Mutual Information Index: M

- M is a multigroup index of segregation first proposed by Theil and Finizza (1971) in the context of racial school segregation.
- Frankel and Volij (2011) prove that M represents the unique nontrivial ordering that satisfies Scale Invariance, Independence, Symmetry, the Group and Unit Division Properties, and a technical continuity property.
- They also show that M is Strong Decomposable: In a between-within decomposition of the index, the within term is the weighted average of segregation in each cluster with weights equal to their demographic shares.

The Mutual Information Index: M

- M is a multigroup index of segregation first proposed by Theil and Finizza (1971) in the context of racial school segregation.
- Frankel and Volij (2011) prove that M represents the unique nontrivial ordering that satisfies Scale Invariance, Independence, Symmetry, the Group and Unit Division Properties, and a technical continuity property.
- They also show that M is Strong Decomposable: In a between-within decomposition of the index, the within term is the weighted average of segregation in each cluster with weights equal to their demographic shares.
- To our knowledge, no other multigroup index of segregation is strongly decomposable.

Some Properties of the M index

- Although M is not Composition Invariant, Mora and Ruiz-Castillo (2010) show that in pair wise comparisons this index admits two decompositions where invariant terms are identified.

Some Properties of the M index

- Although M is not Composition Invariant, Mora and Ruiz-Castillo (2010) show that in pair wise comparisons this index admits two decompositions where invariant terms are identified.
- M is closely related to the so called Information or Entropy index, H:
  - H is the M index normalized by the groups' entropy.
  - While M admits Strong Decomposability both by units and groups, H admits a weaker decomposability property that, for groups, is open to ambiguities in its interpretation (Mora and Ruiz-Castillo (2010).

Some Properties of the M index

- Although M is not Composition Invariant, Mora and Ruiz-Castillo (2010) show that in pair wise comparisons this index admits two decompositions where invariant terms are identified.
- M is closely related to the so called Information or Entropy index, H:
  - H is the M index normalized by the groups' entropy.
  - While M admits Strong Decomposability both by units and groups, H admits a weaker decomposability property that, for groups, is open to ambiguities in its interpretation (Mora and Ruiz-Castillo (2010).
- In this presentation, I will focus on the implementation of the Strong Decomposability Property with the M index in Stata.

Strong Decomposability

Motivation 1: Occupational Segregation by Gender

- Assume that individuals can be either employed along J occupations or doing housekeeping.

Motivation 1: Occupational Segregation by Gender

- Assume that individuals can be either employed along J occupations or doing housekeeping.
- Code *housekeeping* as the $J+1$ activity and let p_j be the proportion of workers in activity $j = 1, ..., J, J+1$.

Motivation 1: Occupational Segregation by Gender

- Assume that individuals can be either employed along J occupations or doing housekeeping.
- Code *housekeeping* as the $J+1$ activity and let p_j be the proportion of workers in activity $j = 1, ..., J, J+1$.
- If an individual is drawn randomly from the pool of individuals, the expected information of learning the worker's occupation is measured by her or his entropy $E_J = \sum_{j=1}^{J+1} p_j log\left(\frac{1}{p_j}\right)$.

Motivation 1: Occupational Segregation by Gender

- Assume that individuals can be either employed along J occupations or doing housekeeping.
- Code *housekeeping* as the $J+1$ activity and let p_j be the proportion of workers in activity $j = 1, ..., J, J+1$.
- If an individual is drawn randomly from the pool of individuals, the expected information of learning the worker's occupation is measured by her or his entropy $E_J = \sum_{j=1}^{J+1} p_j log\left(\frac{1}{p_j}\right)$.
- After learning that the worker is a woman (man), her (his) entropy decreases to E_J^f (E_J^m). M_J results from these reductions in expected information averaged over female and male workers,

$$M_J = p^f\left(E_J - E_J^f\right) + p^m\left(E_J - E_J^m\right) \tag{1}$$

The Gender Division of Labor

- Assume that jobs can be classified into part-time and full-time jobs.

The Gender Division of Labor

- Assume that jobs can be classified into part-time and full-time jobs.
- Extend the set of organizational units from the original $J + 1$ occupational categories to a new set which includes all interactions between the original occupations and the full-time vs. part-time status of the job. We now have $2J + 1$ categories.

The Gender Division of Labor

- Assume that jobs can be classified into part-time and full-time jobs.
- Extend the set of organizational units from the original $J+1$ occupational categories to a new set which includes all interactions between the original occupations and the full-time vs. part-time status of the job. We now have $2J+1$ categories.
- Accordingly, define the index for the gender division of labor as the M index over the expanded set of $2J+1$ activities:

$$M_{GDL} = p^f \left(E_{2J} - E^f_{2J} \right) + p^m \left(E_{2J} - E^m_{2J} \right) \tag{2}$$

Strong Unit Decomposability

- Let M_{FPLF} be the M index of gender segregation where the only organizational units are *working part-time*, *working full-time*, and *housekeeping*.

Strong Unit Decomposability

- Let M_{FPLF} be the M index of gender segregation where the only organizational units are *working part-time*, *working full-time*, and *housekeeping*.
- By *Strong Decomposability*:

$$M_{GDL} = M_{FPLF} + M_{2J}^{W(FPLF)} \tag{3}$$

Strong Unit Decomposability

- Let M_{FPLF} be the M index of gender segregation where the only organizational units are *working part-time*, *working full-time*, and *housekeeping*.
- By *Strong Decomposability*:

$$M_{GDL} = M_{FPLF} + M_{2J}^{W(FPLF)} \tag{3}$$

- Hence, the ratio $\frac{M_{2J}^{W(FPLF)}}{M_{GDL}}$ captures occupational segregation after controlling for gender differences in labor force participation and the incidence of part-time jobs.

Traditional notions of Occupational Segregation

- In traditional studies on occupational segregation, there is no distinction between part- and full-time jobs and only the working population is considered in the analysis.

Traditional notions of Occupational Segregation

- In traditional studies on occupational segregation, there is no distinction between part- and full-time jobs and only the working population is considered in the analysis.
- Let M_0 denote the traditional index of occupational segregation using as organizational units the original J occupations over the working population.

Traditional notions of Occupational Segregation

- In traditional studies on occupational segregation, there is no distinction between part- and full-time jobs and only the working population is considered in the analysis.
- Let M_0 denote the traditional index of occupational segregation using as organizational units the original J occupations over the working population.
- It can be shown that:

$$M_{GDL} = M_{PR} + p_{work} M_0 + M_{2J}^{W(PR,occ)} \tag{4}$$

Traditional notions of Occupational Segregation

- In traditional studies on occupational segregation, there is no distinction between part- and full-time jobs and only the working population is considered in the analysis.
- Let M_0 denote the traditional index of occupational segregation using as organizational units the original J occupations over the working population.
- It can be shown that:

$$M_{GDL} = M_{PR} + p_{work} M_0 + M_{2J}^{W(PR,occ)} \tag{4}$$

- Hence, $\frac{p_{work} M_0}{M_{GDL}}$ captures the contribution of occupational segregation to GDL after controlling for gender differences in labor force participation.

Motivation 2: Occupational Segregation by Race and Ethnicity

- Assume that workers in J occupations are distinguished by their ethnic (e) and gender (g) status.

Motivation 2: Occupational Segregation by Race and Ethnicity

- Assume that workers in J occupations are distinguished by their ethnic (e) and gender (g) status.
- The M index of occupational segregation jointly by ethnic and gender, M^*, is the average increase in the information we have about the worker's occupation that comes from learning her/his ethnicity and gender:

$$M^* = \sum_{e,g} p^{e,g} \left[E_J - E_{J|e,g} \right] \tag{5}$$

Strong Group Decomposability

- As the M index fulfills the Strong Group Decomposability, M^* satisfies the following two decompositions:

$$M^* = M^g + \sum_g p^g M^e(g) = M^e + \sum_e p^e M^g(e). \tag{6}$$

Strong Group Decomposability

- As the M index fulfills the Strong Group Decomposability, M^* satisfies the following two decompositions:

$$M^* = M^g + \sum_g p^g M^e(g) = M^e + \sum_e p^e M^g(e). \tag{6}$$

- This equation is important because it quantifies how much of ethnic and gender segregation is *exclusively* due to either ethnicity or gender.

Three Scenarios

$$I = M^* - \left(\sum_g p_g M^e(g) + \sum_e p_e M^g(e) \right). \tag{7}$$

- I is segregation jointly induced by ethnicity and gender that cannot be attributed *uniquely* to either of these two factors

Three Scenarios

$$I = M^* - \left(\sum_g p_g M^e(g) + \sum_e p_e M^g(e) \right). \tag{7}$$

- I is segregation jointly induced by ethnicity and gender that cannot be attributed *uniquely* to either of these two factors
- When $I = 0$, the exclusive contributions of ethnicity and gender add up to their joint effect, M^*.

Three Scenarios

$$I = M^* - \left(\sum_g p_g M^e(g) + \sum_e p_e M^g(e) \right). \tag{7}$$

- I is segregation jointly induced by ethnicity and gender that cannot be attributed *uniquely* to either of these two factors
- When $I = 0$, the exclusive contributions of ethnicity and gender add up to their joint effect, M^*.
- When $I > 0$, a part of M^* cannot be attributed to either factor in isolation. In this case, traditional measures overestimate the amount of segregation induced by each status.

Three Scenarios

$$I = M^* - \left(\sum_g p_g M^e(g) + \sum_e p_e M^g(e) \right). \tag{7}$$

- I is segregation jointly induced by ethnicity and gender that cannot be attributed *uniquely* to either of these two factors
- When $I = 0$, the exclusive contributions of ethnicity and gender add up to their joint effect, M^*.
- When $I > 0$, a part of M^* cannot be attributed to either factor in isolation. In this case, traditional measures overestimate the amount of segregation induced by each status.
- When $I < 0$, the combination of ethnicity and gender produces less segregation than we would observe if we simply add the net segregative effects of each status. In this case, traditional measures underestimate the portion of segregation that each status begets.

Computing M with Stata

The mutual ado

mutual *varname varlist* [*if*] [*in*] [*weight*] [using *filename*], options

- *varname*: discrete variable that represents group status (organizational unit)
- *varlist*: discrete variables whose combinations identify the organizational units (the group status)
- Options:
  - **gen**erate(*name*): variable name for the index.
  - **within**(*varlist1*): computes the within term of the index.
  - **by**(*varlist2*)]: computes the index for each cell defined by varlist2.
- frequency and importance weights are allowed
- each observation represents an individual or a group of individuals

Examples (1)

Occupational Segregation by Gender

```
mutual gender occup if occup!=., by(year) gen(M)
```

- It lists the M index of occupational segregation by gender for each year.

Examples (1)

Occupational Segregation by Gender

```
mutual gender occup if occup!=., by(year) gen(M)
```

- It lists the M index of occupational segregation by gender for each year.

Occupational segregation by gender within race

```
mutual gender occup [iw=count], by(year country)
gen(Mg_e) within(race)
```

- It lists the within term of the joint M index. Each observation represents several individuals.

Examples (1)

Occupational Segregation by Gender

```
mutual gender occup if occup!=., by(year) gen(M)
```

- It lists the M index of occupational segregation by gender for each year.

Occupational segregation by gender within race

```
mutual gender occup [iw=count], by(year country)
gen(Mg_e) within(race)
```

- It lists the within term of the joint M index. Each observation represents several individuals.

Global division of labor

```
mutual gender ea occup lmi [iw=count], by(year country)
gen(GDL)
```

- Several variables define the organizational units.

Examples (2)

Occupational segregation by gender and race within educational levels

```
mutual occ_3d gender race [iw=count], by(year country)
gen(M_star) within(educ)
```

- Several variables define the group status. It lists the joint M index for each year and country within educational levels.

Examples (2)

Occupational segregation by gender and race within educational levels

```
mutual occ_3d gender race [iw=count], by(year country)
gen(M_star) within(educ)
```

- Several variables define the group status. It lists the joint M index for each year and country within educational levels.

Global division of labor within educational levels

```
mutual gender ea06 occ_3d lmi [iw=count] using
"global.dta", by(year country) gen(GDL_edc) within(educ)
```

- It saves the results in “global.dta”

Empirical Illustrations

Segregated Integration (with M. Kreimer)

- “Segregated Integration: Recent Trends in the Austrian Gender Division of Labor”, (with Margareta Kreimer) WP 13-17. UC3M, Spain.

Segregated Integration (with M. Kreimer)

- “Segregated Integration: Recent Trends in the Austrian Gender Division of Labor”, (with Margareta Kreimer) WP 13-17. UC3M, Spain.
- Using micro data from the Austrian Labor Force Survey from 1995-2010, we explore how decreases in the gender differential in participation rates together with increasing differentials in the incidence of part-time jobs and stable or rising levels of occupational segregation by gender affect the gender division of labor.

Segregated Integration (with M. Kreimer)

- “Segregated Integration: Recent Trends in the Austrian Gender Division of Labor”, (with Margareta Kreimer) WP 13-17. UC3M, Spain.
- Using micro data from the Austrian Labor Force Survey from 1995-2010, we explore how decreases in the gender differential in participation rates together with increasing differentials in the incidence of part-time jobs and stable or rising levels of occupational segregation by gender affect the gender division of labor.
- To so so, we propose an index for the gender division of labor based on the Mutual Information index.

The Evolution of GDL

Recent Trends in the Austrian Gender Division of Labor

	1995	1997	1998	2003	2004	2005	2010
Gender Division of Labor	20.84	24.05	23.90	23.29	20.29	20.66	20.22
Full- vs. Part time and LF	5.47	5.51	5.54	6.08	6.22	5.91	5.66
Activity and LMI	11.99	12.05	12.03	11.60	10.75	10.23	9.48
Contribution of Occupational Segregation to GDL							
Within Activity and LMI	8.85	12.00	11.87	11.69	9.54	10.43	10.74
$p_{work}M_0$	10.77	14.03	13.88	14.19	12.15	13.30	13.76
M_0	16.90	22.27	21.84	22.00	19.27	20.53	20.64

Brief Summary of Results

- Our main results show that the gender division of labor is very stable along the 16-year period.

Brief Summary of Results

- Our main results show that the gender division of labor is very stable along the 16-year period.
- This is so because although the rising female labor force participation reduces the gender division of labor, increases in gender differences in the incidence of part-time jobs and increases in occupational segregation result in greater division of labor across genders.

Brief Summary of Results

- Our main results show that the gender division of labor is very stable along the 16-year period.
- This is so because although the rising female labor force participation reduces the gender division of labor, increases in gender differences in the incidence of part-time jobs and increases in occupational segregation result in greater division of labor across genders.
- These results are robust to alternative definitions of economic activity and labor market involvement and can also be found after controlling for educational levels and fields.

The joint effect of ethnicity and gender on occupational segregation

- "The intersection of gender and ethnicity : a new approach to the study of occupational segregation" (with Guinea-Martín and Ruiz-Castillo). WP 11-40. UC3M. Spain.

The joint effect of ethnicity and gender on occupational segregation

- "The intersection of gender and ethnicity : a new approach to the study of occupational segregation" (with Guinea-Martín and Ruiz-Castillo). WP 11-40. UC3M. Spain.
- We study census data for England and Wales in 2001.

The joint effect of ethnicity and gender on occupational segregation

- "The intersection of gender and ethnicity : a new approach to the study of occupational segregation" (with Guinea-Martín and Ruiz-Castillo). WP 11-40. UC3M. Spain.
- We study census data for England and Wales in 2001.
- We measure the joint effect of gender and ethnicity on occupational segregation by applying the M index to the product of the two genders and seven ethnic groups.

The joint effect of ethnicity and gender on occupational segregation

- "The intersection of gender and ethnicity : a new approach to the study of occupational segregation" (with Guinea-Martín and Ruiz-Castillo). WP 11-40. UC3M. Spain.
- We study census data for England and Wales in 2001.
- We measure the joint effect of gender and ethnicity on occupational segregation by applying the M index to the product of the two genders and seven ethnic groups.
- Specifically, we study
  - how much each source contributes to occupational segregation, controlling for the effect of the other
  - whether the combined impact of gender and ethnicity is greater than, equal to, or smaller than the sum of their individual effects

The joint effect of ethnicity and gender

	ethnic	*gender*	*joint*	*ethnic* w. *gender*	*gender* w. *ethnic*	*interaction*
all	1.4	20.1	21.7	1.6	20.3	-0.17
non-mixed areas	0.6	21.7	22.4	0.7	21.7	-0.05
mixed areas	2.5	18.0	20.8	2.8	18.3	-0.28
Controlling for Human Capital						
non-mixed areas	0.7	22.8	23.5	0.7	22.8	-0.07
mixed areas	3.2	19.1	22.6	3.5	19.3	-0.28

Brief Summary

- We confirm previous results showing the greater importance of gender over ethnicity as a source of occupational segregation.

Brief Summary

- We confirm previous results showing the greater importance of gender over ethnicity as a source of occupational segregation.
- However, we find that ethnicity contributes 13.5 percent of overall segregation in geographical areas where minorities concentrate.

Brief Summary

- We confirm previous results showing the greater importance of gender over ethnicity as a source of occupational segregation.
- However, we find that ethnicity contributes 13.5 percent of overall segregation in geographical areas where minorities concentrate.
- Contrary to intersectionality theories, we find that there is a small, “dwindling” interaction effect between the two sources of segregation: ethnicity slightly weakens the segregative power of gender, and vice versa.

Conclusions

- The Mutual Information Index is the only multigroup segregation index that, in addition to possessing other desirable properties, satisfies the Strong Group & Unit Decomposability Property.

Conclusions

- The Mutual Information Index is the only multigroup segregation index that, in addition to possessing other desirable properties, satisfies the Strong Group & Unit Decomposability Property.
- Consider, for example, the study of occupational segregation jointly by ethnicity and gender. An index that satisfies Strong GRoup Decomposability allows us to identify the proportion of occupational segregation by ethnicity and gender that can be attributed exclusively to either ethnicity or gender.

Conclusions

- The Mutual Information Index is the only multigroup segregation index that, in addition to possessing other desirable properties, satisfies the Strong Group & Unit Decomposability Property.
- Consider, for example, the study of occupational segregation jointly by ethnicity and gender. An index that satisfies Strong GRoup Decomposability allows us to identify the proportion of occupational segregation by ethnicity and gender that can be attributed exclusively to either ethnicity or gender.
- The computation of the index is easy and can be implemented with the help of an ado file.

Conclusions

- The Mutual Information Index is the only multigroup segregation index that, in addition to possessing other desirable properties, satisfies the Strong Group & Unit Decomposability Property.
- Consider, for example, the study of occupational segregation jointly by ethnicity and gender. An index that satisfies Strong GRoup Decomposability allows us to identify the proportion of occupational segregation by ethnicity and gender that can be attributed exclusively to either ethnicity or gender.
- The computation of the index is easy and can be implemented with the help of an ado file.
- The decomposability properties are illustrated by means of two examples: the joint study of gender and ethnicity in occupational segregation as well as the measurement of the gender division of labor.

Thank you

Frankel, D. M., Volij, O., 2011. Measuring school segregation. Journal of Economic Theory 146 (1), 1–38.
URL http://www.sciencedirect.com/science/article/pii/S0022053110001353

Mora, R., Ruiz-Castillo, J., 2010. Entropy-based segregation indices. Sociological Methodology 41, 159–194.

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.
URL http://www.tandfonline.com/doi/abs/10.1080/0022250X.1971.9989795