UNIVERSIDAD DE LA SALLE

Computing occupational segregation indices with standard errors

An ado-file application with an illustration for Colombia

Jairo G. Isaza-Castro
jisaza@lasalle.edu.co
Karen Guerrero; Karen Hernandez; Jessy Hemer

Stata Conference at Baltimore (MN), July 29th 2017

Motivation

- Analyzing changes in segregation indices over time or across population groups requires some reference to their variability. Having a representative sample allows to calculate an estimator for the population value of any segregation index –but this yields no information about its dispersion (Deutsch et al. 2002)
- Bootstrap provides a solution for situations like this (*cfr.* Deutsch et al. 2002; Jenkins et al. 2002)
- We developed an ado file called "**`segregation`**" which allows the user to compute three segregation indices with standard errors and confidence intervals:
  - Duncan and Duncan (1955) dissimilarity index
  - Gini Coefficient based on the distribution of jobs by gender (see Deutsch et al. 1994) and
  - Karmel and MacLachlan (1988) index of labor market segregation

Outline

- What we mean by “occupational segregation”
- Selected occupational segregation indices
- The algorithm
- Results and discussion
- Pending issues for further research

What we mean by "occupational segregation"

Three overlapping concepts (Blackburn and Jarman, 2005):

- **Segregation** which refers to the existence of a differentiated pattern of jobs predominantly performed by either women or men.
- **Exposure**, which is related to the degree of social interaction that one minority group has with the rest of the population in the labour market.
- **Concentration**, that relates to the composition of the labour force in terms of minority/majority groups of the population and is measured in one or more occupations.

UNIVERSIDAD DE LA SALLE

Occupational segregation indices

Index	Statistical formulas	Definition
Dissimilarity index (Duncan & Duncan, 1955)	$DI = \frac{1}{2}\sum_{i=1}^{n}\left\|\frac{F_i}{F} - \frac{M_i}{M}\right\|, i = 1,2,\dots,n$	✓where *n* is the number of occupations, F_i and M_i are the number of female and male workers in occupation *I,* respectively, and *F* and *M* refer to the total number of female and male workers.
Gini coefficient of the distribution of jobs (Silber, 1986)	$GI = \frac{1}{2}\sum_{i=1}^{n}\sum_{i=j}^{n}\frac{M_i}{M}\frac{M_j}{M}\left\|\frac{F_i/M_i - F_j/M_j}{F/M}\right\|$	✓ where Mi and Fi are defined as explained above. ✓ **it represents a weighted relative mean of deviations of the male/female ratios from an average gender distribution of jobs within occupations**.
Karmel and MacLachlan (1988) index	$KM = \sum_{i=1}^{n}\left\|a\frac{M_i}{T} - (1-a)\frac{F_i}{T}\right\|$	✓ where a (=F/(M+F)) represents the female participation in the labour force and T = M + F.

Command structure

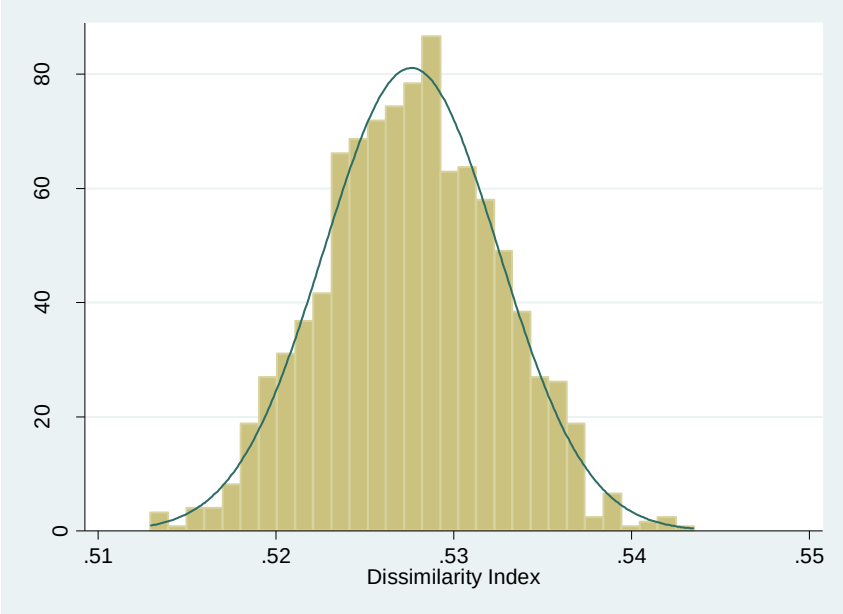
segregation depvar groupvar [weight] [if exp], n(#) [by(varname)]

- where ***depvar*** is a categorical variable deemed to be relevant for the analysis, ***groupvar*** features the dichotomous variable defining the analysis groups (i.e., gender or ethnic group), ***[weight]*** specifies the weight variable (in terms either of frequencies or sampling weights), ***n(#)*** indicates the number of resamples from the original sample to be taken and, ***by(varname)*** declares a categorical variable across which the command can be repeated.

The algorithm

- Steps

1. It takes a view of the original data into Mata for the relevant variables (occupation variable and dichotomous grouping variable –plus conditional variables if necessary)
2. Then it draws a number of random samples (i.e., 1200) with replacement from the original Mata view in order to obtain a distribution for each one of the three segregation measures described above.
3. Finally it estimates the means for the segregation measures to draw the results table with their corresponding standard errors and confidence intervals (at the 95%).



Results

UNIVERSIDAD DE LA SALLE

The dataset…

```
Contains data from C:\Users\JairoG\Dropbox\jairo\2017\Stata Conference 2017\GEIH_rural_2011.dta
  obs:        37,192
 vars:             4                          22 Jul 2017 15:22
 size:       260,344
-------------------------------------------------------------------------------------------------
storage   display    value
variable name   type    format     label      variable label
-------------------------------------------------------------------------------------------------
estrato1        byte    %8.0g      estrato1   sextile by quality of life score
p6020           byte    %8.0g      p6020      sex
fex_c           float   %9.0g                 frequency weights
isco            byte    %10.0g                int. standard classification of occupations 1968
-------------------------------------------------------------------------------------------------
```

Results

UNIVERSIDAD DE LA SALLE

Do-file Editor - rutina segregation*

File Edit View Project Tools

campana* rutina segregation* Untitled1.do

```
1   clear all
2   set more off
3   cd "E:\Dropbox\Archivos_Memo\Unisalle\jairo\2017\Stata Conference 2017"
4
5   use GEIH_rural_2011, clear
6
7   /* To obtain the three segregation measures from 1200 resamples */
8   segregation isco p6020 , n(1200)
9
10  /* To obtain the three segregation measures with the "if" conditional */
11  segregation isco p6020 if estrato1==1, n(1200)
12
13  /* Segregation measures with weighted data */
14  segregation isco p6020 [fw=fex_c], n(1200)
15
16  /* Segregation measures by strata */
17  segregation isco p6020 , n(1200) by(estrato1)
18
19  /* Several options combined */
20  segregation isco p6020 [fw=fex_c] if estrato1<4, n(1200) by(estrato1)
21
22
```

```
. /* To obtain the three segregation measures from 1200 resamples */
. segregation isco p6020 , n(1200)

Mean estimation                     Number of obs    =    1200

------------------------------------------------------------------
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------------
        Gini |   .7822177   .0020915      .7781142    .7863211
      Duncan |   .6163188   .0018427      .6127036    .6199341
         Kmi |   .2325772   .0004544      .2316857    .2334687
------------------------------------------------------------------

.
. /* To obtain the three segregation measures with the "if" conditional */
. segregation isco p6020 if estrato1==1, n(1200)
(19310 real changes made)

Mean estimation                     Number of obs    =    1200

------------------------------------------------------------------
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------------
        Gini |   .9066739   .0031356      .9005221    .9128257
      Duncan |   .8399941   .0045741      .8310199    .8489683
         Kmi |   .3259277   .0017544      .3224857    .3293698
------------------------------------------------------------------
```

- Conventional results based from 1200 resamples
- Conditional results for Strata 1

```
. /* To obtain the three segregation measures from 1200 resamples */
. segregation isco p6020 , n(1200)

Mean estimation                     Number of obs    =    1200

------------------------------------------------------------------
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------------
       Gini |   .7822177   .0020915      .7781142    .7863211
     Duncan |   .6163188   .0018427      .6127036    .6199341
        Kmi |   .2325772   .0004544      .2316857    .2334687
------------------------------------------------------------------

.
. /* To obtain the three segregation measures with the "if" conditional */
. segregation isco p6020 if estrato1==1, n(1200)
(19310 real changes made)

Mean estimation                     Number of obs    =    1200

------------------------------------------------------------------
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------------
       Gini |   .9066739   .0031356      .9005221    .9128257
     Duncan |   .8399941   .0045741      .8310199    .8489683
        Kmi |   .3259277   .0017544      .3224857    .3293698
------------------------------------------------------------------
```

- Conventional results based from 1200 resamples
- Conditional results for Strata 1

- The command can also compute results with weighted data

```
. /* Segregation measures with weighted data */
. segregation isco p6020 [fw=fex_c], n(1200)

Mean estimation                         Number of obs    =    1200

------------------------------------------------------------------------------
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        Gini |   .8210078   .0028506      .8154151    .8266005
      Duncan |   .6569272   .0026704      .6516881    .6621664
         Kmi |   .2859661    .001086      .2838354    .2880967
------------------------------------------------------------------------------
```

- The command can also compute results with weighted data

In this case, using weights moves all indices upwards but this does not have always to be the case

```
. /* Segregation measures with weighted data */
. segregation isco p6020 [fw=fex_c], n(1200)

Mean estimation                         Number of obs    =      1200

------------------------------------------------------------------------------
             |       Mean   Std. Err.      [95% Conf. Interval]
-------------+----------------------------------------------------------------
        Gini |   .8210078   .0028506       .8154151    .8266005
      Duncan |   .6569272   .0026704       .6516881    .6621664
         Kmi |   .2859661    .001086       .2838354    .2880967
------------------------------------------------------------------------------
```

```
. /* Segregation measures by strata */
. segregation isco p6020 , n(1200) by(estrato1)
(1 vector posted)
estrato 1

Mean estimation                     Number of obs    =    1200

------------------------------------------------------------------
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------------
        Gini |   .8523668   .0027112      .8470475    .8576861
      Duncan |   .8032839   .0032977      .7968141    .8097538
         Kmi |   .1998572   .0026832      .1945929    .2051216
------------------------------------------------------------------
estrato 2

Mean estimation                     Number of obs    =    1200

------------------------------------------------------------------
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------------
        Gini |    .595272   .0030826      .5892242    .6013198
      Duncan |   .5336536   .0039289      .5259453     .541362
         Kmi |    .184353   .0020711      .1802897    .1884163
------------------------------------------------------------------
estrato 3

Mean estimation                     Number of obs    =    1200

------------------------------------------------------------------
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------------
        Gini |   .8804594   .0035381      .8735179    .8874009
      Duncan |   .8260801   .0050609      .8161509    .8360093
         Kmi |    .324485   .0024263      .3197247    .3292453
------------------------------------------------------------------
```

- The "[by(*varname*)]" option

- Several options can also be applied simultaneously
  - ✓ weights
  - ✓ if
  - ✓ by

```
. /* Several options combined */
. segregation isco p6020 [fw=fex_c] if estrato1<4, n(1200) by(estrato1)
(50278 real changes made)
(1 vector posted)
estrato 1

Mean estimation                     Number of obs    =    1200

------------------------------------------------------------------
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------------
        Gini |   .9482807   .0027633      .9428592    .9537022
      Duncan |   .9265347   .0039048      .9188737    .9341957
         Kmi |   .3988634   .0018887      .3951578     .402569
------------------------------------------------------------------
estrato 2

Mean estimation                     Number of obs    =    1200

------------------------------------------------------------------
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------------
        Gini |   .5350322   .0056822      .5238841    .5461804
      Duncan |   .4715174   .0061122      .4595255    .4835093
         Kmi |    .219724   .0028514      .2141297    .2253183
------------------------------------------------------------------
estrato 3

Mean estimation                     Number of obs    =    1200

------------------------------------------------------------------
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------------
        Gini |   .5259643   .0044516      .5172304    .5346981
      Duncan |   .4404268   .0042713      .4320468    .4488068
         Kmi |   .2098955    .002093      .2057891    .2140018
------------------------------------------------------------------
```

- Several options can also be applied simultaneously
  - ✓ weights
  - ✓ if
  - ✓ by

```
. /* Several options combined */
. segregation isco p6020 [fw=fex_c] if estrato1<4, n(1200) by(estrato1)
(50278 real changes made)
(1 vector posted)
estrato 1

Mean estimation                     Number of obs     =      1200

------------------------------------------------------------------------------
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        Gini |   .9482807   .0027633      .9428592    .9537022
      Duncan |   .9265347   .0039048      .9188737    .9341957
         Kmi |   .3988634   .0018887      .3951578     .402569
------------------------------------------------------------------------------
estrato 2

Mean estimation                     Number of obs     =      1200

------------------------------------------------------------------------------
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        Gini |   .5350322   .0056822      .5238841    .5461804
      Duncan |   .4715174   .0061122      .4595255    .4835093
         Kmi |    .219724   .0028514      .2141297    .2253183
------------------------------------------------------------------------------
estrato 3

Mean estimation                     Number of obs     =      1200

------------------------------------------------------------------------------
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        Gini |   .5259643   .0044516      .5172304    .5346981
      Duncan |   .4404268   .0042713      .4320468    .4488068
         Kmi |   .2098955    .002093      .2057891    .2140018
------------------------------------------------------------------------------
```

Some pending issues

- We are working to give the user more choice to customize the output: different levels for confidence intervals, picking or dropping indices, reporting additional statistics
- More flexibility in order to account for complex sampling designs
- Other segregation measures proposed in the literature could also be incorporated
- Extensions to multi-group segregation indices
  - Hutchens `square root' segregation index with optional decompositions by subgroups (see Jenkings *et al.* 2006)
  - "seg" command calculates several indices to which standard errors could also be applied: Gini index, Theil Information Theory index, Squared Coefficient of Variation index and Simpson Diversity indexes (see Reardon & Firebaugh 2002)

References

BLACKBURN, R. M., BROOKS, B. & JARMAN, J. 2001. Occupational Stratification: The Vertical Dimension of Occupational Segregation. *Work, Employment & Society,* 15**,** 511-538.

DEUTSCH, J., FLUCKIGER, Y. & SILBER, J. (1994) Measuring occupational segregation: Summary statistics and the impact of classification errors and aggregation. *Journal of Econometrics*, 61, 133-146.

DUNCAN, O.D., DUNCAN, B., 1955: A Methodological Analysis of Segregation Indexes. *American Sociological Review* 20: 210-217.

HUTCHENS, R. 2004. One measure of segregation. *International Economic Review* 45(2): 555-578.

ISAZA-CASTRO, J.G. & REILLY, B.M. (2010) *Occupational Segregation by Gender: An Empirical Analysis for Urban Colombia (1986-2004)*. Paper presented at the Guanajuato Workshop for Young Economists, Guanajuato (Mexico).

JENKINS, S.P., MICKLEWRIGHT, J. and SCHNEPF, S.V. 2006. Social segregation in secondary schools: how does England compare with other countries? *Working Paper 2006-02*, Institute for Social and Economic Research, University of Essex. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.421.4691&rep=rep1&type=pdf -access: 27 June 2017

KARMEL, T. & MACLACHLAN, M. (1988) Occupational Sex Segregation--Increasing or Decreasing? *Economic Record*, 64, 187.

REARDON, S. F., & FIREBAUGH, G. 2002. "Measures of multigroup segregation." Sociological Methodology 32: 33-67.

SEMYONOV, M. & JONES, F. (1999) Dimensions of Gender Occupational Differentiation in Segregation and Inequality: A Cross-National Analysis. *Social Indicators Research*, 46, 225-247.

SHAO, J., and D. TU. 1995. *The Jackknife and Bootstrap*. New York: Springer.

SILBER, J. G. (1989) On the measurement of employment segregation. *Economics Letters*, 30, 237-243.

Gracias!