AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

Check for updates

The Stata Journal (2016)
16, Number 2, pp. 386–402

igmobil: A command for intergenerational mobility analysis in Stata

Marco Savegnago
Bank of Italy
Rome, Italy
marco.savegnago@bancaditalia.it

Abstract. In this article, I describe a new command, `igmobil`, that computes up to 20 intergenerational mobility (IGM) indices for continuous (that is, income or years of education) or discrete (that is, educational or occupational level) variables. I consider three classes of IGM indices: 1) single-stage indices, 2) indices derived from a transition matrix between parents' and children's socioeconomic status, and 3) indices based on inequality measures. Users may add a fourth class to specify any possible IGM index not included in `igmobil`. Standard errors and confidence intervals are calculated using a bootstrap procedure. Users can customize many aspects of the program output, including the type and dimension of the transition matrix, the parameters for some IGM indices (like the ones involving generalized entropy measures and the Atkinson index), and how standard errors and confidence intervals are calculated.

Keywords: st0437, igmobil, intergenerational mobility, IGM, bootstrap

1 Introduction

Intergenerational mobility (IGM) refers to the extent to which the advantages and disadvantages of individuals (according to dimensions such as income, wealth, or education) are transmitted across generations (Black and Devereux 2011). A key aspect is that IGM is a complex concept and may mean different things to different people. Fields (2008) identifies six different concepts of income mobility—namely, time independence, positional movement, share movement, nondirectional income movement, directional income movement, and equalizer of long-term incomes—and each concept can be quantified by a specific set of indices. For example, share movement mobility occurs when an individual's income increases with respect to the population mean.[1] A possible index for this example would be $\mathcal{M}(X_i, Y_i) = N^{-1}\sum_{i=1}^{n} |Y_i/\mu_Y - X_i/\mu_X|$, where Y_i and X_i denote, respectively, a child's and his or her parent's income for family i (see section 2). A natural consequence in this case is that there is no consensus on how IGM should be measured, so many indices are available to an applied researcher.[2] Therefore, because the complexity of IGM cannot be captured by a unique index, Fields and Ok (1999a) suggest using different measures.

1. Fields (2008) notes that a child can experience share movement even if he or she has the same income as his or her parent.
2. See Fields and Ok (1999a), Checchi and Dardanoni (2002), Black and Devereux (2011), and Jäntti and Jenkins (2013).

© 2016 StataCorp LP st0437

igmobil computes point estimates and inferential measures for 19 IGM indices commonly used in the applied research. These 19 indices are appropriate for measuring income mobility, although only some of them are suitable for analyzing mobility according to other dimensions, such as years of education or occupational level. Users can add a 20th index to this list by specifying their own user-written program, an example of which is provided in section 4.6.

igmobil also computes standard errors and confidence intervals using an embedded bootstrap procedure.

2 Intergenerational mobility indices

Let the vector (Y_i, X_i) describe the socioeconomic status (SES) of a child and his or her parent for family i. We are interested in the extent to which the child's SES, Y_i, depends on the parent's SES, X_i.[3] From a practical point of view, the abstract and multifaceted notion of SES must be proxied by an observable variable, which is typically chosen among income, wealth, health, education, occupational prestige, and the like. For simplicity, I will assume that the vectors $(\mathbf{Y}, \mathbf{X})$ contain information on permanent income. In the final paragraph of this section, I will briefly discuss how analysis can be carried out on other dimensions of SES other than income.

An intergenerational mobility index, $\mathcal{M}(Y, X)$, is any function $\mathcal{M} : \mathbb{R}^{2n} \rightarrow \mathbb{R}$, which maps the vectors of incomes $(\mathbf{Y}, \mathbf{X})$ into a scalar. Following the distinctions made in Cowell and Schluter (1998) and Checchi and Dardanoni (2002), I divide the IGM indices computed by igmobil into the following three broad classes: 1) single-stage indices, which are computed directly on microdata; 2) indices based on a transition matrix, which require discrete or discretized data; and 3) inequality reduction indices, which are based on inequality measures (such as the Gini coefficient) computed on the cross-sectional distribution of income for both generations. Users may add a fourth class to specify any possible IGM index not included in igmobil.

Table 1 describes the IGM indices estimated by the program. For each index, I report the bibliographic reference where that index was either first proposed or described in an applied work.

3. Alternatively, one can study the evolution of SES for the same individuals at different points in time. What follows applies to both intergenerational and intragenerational mobility, even if I refer to the former only.

Table 1. Intergenerational mobility indices estimated by `igmobil`

Index	Alias	Formula	Reference
		Single-stage indices	
$\mathcal{M}_1$	Abs. difference	$n^{-1}\sum \lvert X_i - Y_i \rvert$	Fields and Ok (1996)
$\mathcal{M}_2$	Sq. difference	$n^{-1}\sum (X_i - Y_i)^2$	Checchi and Dardanoni (2002)
$\mathcal{M}_3$	Fields and Ok	$n^{-1}\sum \lvert \ln X_i - \ln Y_i \rvert$	Fields and Ok (1999b)
$\mathcal{M}_4$	Share	$n^{-1}\sum (X_i/\mu_X - Y_i/\mu_Y)^2$	Fields (2008)
$\mathcal{M}_5$	Hart	$1 - \mathrm{Corr}(\ln Y_i, \ln X_i)$	Hart (1981)
$\mathcal{M}_6$	Spearman	$1 - \mathrm{Spearman}(\ln Y_i, \ln X_i)$	Black and Devereux (2011)
$\mathcal{M}_7$	Abs. CDF	$n^{-1}\sum \lvert F_Y(Y_i) - F_X(X_i) \rvert$	Checchi and Dardanoni (2002)
$\mathcal{M}_8$	Sq. CDF	$n^{-1}\sum \{F_Y(Y_i) - F_X(X_i)\}^2$	Checchi and Dardanoni (2002)
$\mathcal{M}_9$	1-OLS(levels)	$1 - \mathrm{OLS}(Y_i, X_i)$	Black and Devereux (2011)
$\mathcal{M}_{10}$	1-OLS(logs)	$1 - \mathrm{OLS}(\ln Y_i, \ln X_i)$	Black and Devereux (2011)
		Indices based on $\mathbf{P}_{K\times K}$ *transition matrix*	
$\mathcal{M}_{11}$	Prais (trace)	$(K-1)^{-1}\{K - \mathrm{trace}(\mathbf{P})\}$	Shorrocks (1978b)
$\mathcal{M}_{12}$	Bartholomew	$\{K(K-1)\}^{-1}\sum_i \sum_j p_{ij} \lvert i-j \rvert$	Bartholomew (1973)
$\mathcal{M}_{13}$	Eigenvalue2	$1 - \lvert$2nd largest eigenvalue$\rvert$	Sommers and Conlisk (1979)
$\mathcal{M}_{14}$	Determinant	$1 - \lvert \det(\mathbf{P}) \rvert$	Shorrocks (1978b)
		Indices based on inequality reduction	
$\mathcal{M}_{15}$	$\mathcal{S}$ or $\mathcal{F}$ – Gini	See (1), (2)	Fields (2010); Shorrocks (1978a)
$\mathcal{M}_{16}$	$\mathcal{S}$ or $\mathcal{F}$ – GE(a_1)	See (1), (2), and *Appendix*	Fields (2010); Shorrocks (1978a)
$\mathcal{M}_{17}$	$\mathcal{S}$ or $\mathcal{F}$ – GE(a_2)	See (1), (2), and *Appendix*	Fields (2010); Shorrocks (1978a)
$\mathcal{M}_{18}$	$\mathcal{S}$ or $\mathcal{F}$ – Atk(ϵ_1)	See (1), (2), and *Appendix*	Fields (2010); Shorrocks (1978a)
$\mathcal{M}_{19}$	$\mathcal{S}$ or $\mathcal{F}$ – Atk(ϵ_2)	See (1), (2), and *Appendix*	Fields (2010); Shorrocks (1978a)
		Index based on user-written program	
$\mathcal{M}_{20}$	user written (ex.)	See sections 2 and 4.6	

`igmobil` estimates the following classes of indices:

Class 1: Single-stage indices. The defining characteristic of indices belonging to this class is that the final IGM index is an aggregation of mobility that occurs between the families in the population (see Checchi and Dardanoni [2002]). Many well-known indices belong to this class, including the Fields and Ok index ($\mathcal{M}_3$; Fields and Ok [1999b]) and the indices based on the Pearson correlation ($\mathcal{M}_5$), the Spearman correlation ($\mathcal{M}_6$), and the intergenerational elasticity ($\mathcal{M}_{10}$, which is one minus the ordinary least-squares coefficient in a regression of child log-income on parent log-income). The other indices in this class are the average absolute difference ($\mathcal{M}_1$), the average squared difference ($\mathcal{M}_2$), the share index ($\mathcal{M}_4$, introduced in section 1), the average absolute ($\mathcal{M}_7$) and squared ($\mathcal{M}_8$) difference of the individuals' empirical cumulative density functions (CDFs), and the ordinary least-squares coefficient in a regression of child income on parent income ($\mathcal{M}_9$, which uses income levels instead of logs as in the intergenerational elasticity, $\mathcal{M}_{10}$).

Class 2: Indices based on a transition matrix. These indices are functional of the transition matrix $\mathbf{P}_{K\times K}$ between the parent's income level and the child's income level. The generic element p_{jk} represents the probability that the child's income falls in the kth class given that the parent's income falls in the jth class. Income levels can be either absolute (a size transition matrix) or set on the basis of quantiles of the marginal distributions of Y_i and X_i (quantile or fractile transition matrix). Indices derived from a transition matrix combine the elements on the main diagonal (as in the Shorrocks and Prais index, $\mathcal{M}_{11}$; see Shorrocks [1978b]); they consider the average "jump" of income classes (as in the Bartholomew index, $\mathcal{M}_{12}$; see Bartholomew [1973]); and they account for the second-largest eigenvalues ($\mathcal{M}_{13}$; see Sommers and Conlisk [1979]) or the determinant of the matrix itself ($\mathcal{M}_{14}$; see Shorrocks [1978b]). By default, `igmobil` computes 5×5 quantile matrices, although one can specify the desired number of quantiles or can switch to a size transition matrix (in this case, however, data must be discrete).

Class 3: Inequality reduction indices. This class captures the notion of mobility as a long-term income equalizer. The intuition is that, in the case of upward mobility, inequality in the average father–son income (a measure of "long-term" or "dynasty" income) will be smaller than inequality in only the father's income (a measure of "short-term" income). Let $Z=(Y+X)/2$ be the "long-term income" and $\mathcal{I}(\cdot)$ a cross-sectional measure of inequality (such as the Gini index). Then, consider the Shorrocks (1978a) $\mathcal{S}(\mathcal{I},Y,X)$ and the Fields (2010) $\mathcal{F}(\mathcal{I},Y,X)$ families as follows:

$$\text{Shorrocks: } \mathcal{S}(\mathcal{I},Y,X)=1-\frac{\mu_Z\,\mathcal{I}(Z)}{\mu_X\,\mathcal{I}(X)+\mu_Y\,\mathcal{I}(Y)} \tag{1}$$

$$\text{Fields: } \mathcal{F}(\mathcal{I},Y,X)=1-\frac{\mathcal{I}(Z)}{\mathcal{I}(X)} \tag{2}$$

The families $\mathcal{S}(\cdot)$ and $\mathcal{F}(\cdot)$ are conceptually similar: the only difference is that the Shorrocks family's indices do not distinguish whether the income dynamics are equalizing or disequalizing, while the Fields family's indices do. Moreover, the Fields family is very close to the Chakravarty, Dutta, and Weymark (1985) index, although these indices differ in their normative implications (see Fields [2010] for a detailed discussion on this point).

As cross-sectional inequality measures, I will consider the Gini index; the generalized entropy measure, GE(a); and the Atkinson index, Atk(ϵ). I provide a brief description of these last two inequality measures in the *Appendix*. Users can specify that the `igmobil` command uses either the $\mathcal{S}(\cdot)$ or the $\mathcal{F}(\cdot)$ family (the default is $\mathcal{F}(\cdot)$) and can specify up to two parameters each for the generalized entropy (by default, $a = 0$ and $a = 1$, for which the generalized entropy becomes the mean log deviation and the Theil index) and for the Atkinson index (by default, $\epsilon = 0.5$ and $\epsilon = 2$).

Class 4: Index from a user-written program. This option is useful if one wants to enrich the above list with another IGM index. For example, consider the following upward-mobility index suggested by Bhattacharya and Mazumder (2011):

$$\text{UP}_{\tau,s} = \Pr(r_{Y_i} - r_{X_i} > \tau \mid r_{X_i} \le s)$$

Here r_{Y_i} and r_{X_i} are the child's and the parent's relative positions in their marginal income distribution; that is, $r_{Y_i} = F_Y(Y_i)$ and $r_{X_i} = F_X(X_i)$. Thus $\text{UP}_{\tau,s}$ is the probability that the child's rank exceeds the parent's rank by τ given that the parent's rank is below s. I provide code for estimating such an index and for passing its value to the `igmobil` command.

Final remarks and inference. In the given examples, SES of each generation is proxied by income: In this case, all IGM indices computed by `igmobil` make sense, although the specific concept of mobility that one has in mind can lead to a preference for certain indices. If we use other continuous variables (such as years of education or a continuous measure of health status) as SES indicators, then we should probably estimate only some of the single-stage indices [like the indices based on absolute and squared differences ($\mathcal{M}_1$ and $\mathcal{M}_2$), the ones based on empirical CDFs ($\mathcal{M}_7$ and $\mathcal{M}_8$), and the ordinary least-squares coefficient (on levels, as in $\mathcal{M}_9$)]. If we use purely categorical variables (such as occupational status or educational attainment), we should instead estimate only indices based on a transition matrix ($\mathcal{M}_{11}$–$\mathcal{M}_{14}$), and so on. One advantage of `igmobil` is that it can accommodate all of these needs within the same framework.

An important—and probably undervalued—aspect of empirical research on IGM regards statistical inference. `igmobil` computes some complex IGM indices (like those derived from the quantile transition matrix, where we have extra variability resulting from the estimation of quantiles) for which we might expect poor inference if based on asymptotic arguments. For this reason, I implement a bootstrap procedure in the `igmobil` command; options for this procedure can be modified by the user.

3 The igmobil command

3.1 Description

The `igmobil` command provides point estimates, standard errors, and confidence intervals for the three classes of IGM indices discussed in section 2. By default, `igmobil` assumes that data are continuous and computes 10 single-stage indices, 4 indices based on a 5×5 quantile transition matrix, and 5 indices based on inequality measures (including the Fields class applied to the Gini index, generalized entropy measures with parameters −1 and 2, and the Atkinson index with parameters 0.5 and 2). Standard errors are computed with 50 bootstrap replications, and the 95% confidence intervals are based on normal approximation.

The user can modify the program output in the following ways:

- One can omit estimation of some (but not all) classes of IGM indices. Omission of classes can be motivated by lack of interest in that particular class or can be done to decrease computing time.
- One can change the dimension of the quantile transition matrix or switch the program to use a size transition matrix. In the latter, data must be discrete; consequently, single-stage and inequality-based indices will not be computed.
- For inequality-based indices, one can specify that the program use either the Fields or the Shorrocks class and modify the parameters of the generalized entropy measures and the Atkinson index (no more than two scalars for each index).
- Through a proper user-written program, one can include an extra IGM index (see the example in section 4.6).
- One can modify the number of bootstrap replications, the type of confidence intervals (the normal approximation, the percentile method, or the bias-corrected method), and the confidence level.

When accessing results, recall that each index has a progressive number from 1 to 20. The list of indices is reported in table 1 and in the help file. Therefore, to display the standard error of the trace index ($\mathcal{M}_{11}$), we type `display _se[i11]`.

3.2 Syntax

The syntax of `igmobil` is as follows:

`igmobil` *varname1* *varname2* [*if*] [*in*] [, `nosingle` `notrans` `noinequal`
`userwritten(`*userwrittenstr*`)` `classes(`#`)` `discrete` `matrix(`*matname*`)`
`family(`*familystr*`)` `ge(`#[#]`)` `atk(`#[#]`)` `bootstrap(`*bootstrapstr*`)`
`citype(`*citypestr*`)` `format(`*formatstr*`)`]

varname1 and *varname2* denote, respectively, the most and the least recent observation (that is, the child and the parent in the intergenerational setting, or Y_t and Y_{t-1} in the intragenerational setting).

3.3 Options

Main

`nosingle` specifies to not calculate single-stage indices. `nosingle` is appropriate when *varname1* and *varname2* are discrete variables (such as occupational status or income class). This is the default option when the option `discrete` is specified.

`notrans` specifies to not calculate transition-matrix indices. `notrans` is appropriate when the variables are continuous and there is no interest in calculating the transition matrices. `notrans` must not be used with the option `classes()` or `discrete`.

`noinequal` specifies to not calculate indices based on inequality measures. `noinequal` must not be used with the option `family()`, `ge()`, or `atk()`.

`userwritten(`*userwrittenstr*`)` specifies that the output include any IGM index defined in *userwrittenstr*. The program must be r-class and return the IGM index in a scalar named `r(UW)`. An example is provided in section 4.6.

Indices based on a transition matrix

`classes(`#`)` specifies the size of the quantile transition matrix on which transition-matrix indices are to be calculated. The default is `classes(5)`. Quantiles are computed using the `xtile` command (see [D] **pctile**). `classes(`#`)` can be used when *varname1* and *varname2* are continuous and the user wants to specify a quantile transition matrix with a size different from 5. `classes(`#`)` must not be used when variables are discrete or when the option `discrete` or `notrans` is used.

`discrete` specifies that *varname1* and *varname2* are discrete (or already discretized) variables (such as types of jobs, levels of education, or income categories). When `discrete` is used, single-stage and inequality-based indices will not be computed because we are dealing with discrete random variables. `discrete` must not be used with `classes()`.

`matrix(`*matname*`)` saves the resulting transition matrix in *matname*. If the option `notrans` is used, `matrix()` is ignored.

Indices based on inequality measures

`family(`*familystr*`)` specifies what indices will be used to compare inequality measures across generations. *familystr* can be `fields` or `shorrocks` (see Fields [2010] and Shorrocks [1978a]). The default is `family(fields)`.

`ge(`# [#]`)` specifies the values of the generalized entropy measure parameter. The maximum two values of `ge()` can be listed; if only one value is chosen, it is repeated twice. The default is `ge(0 1)`.

atk(#[#]) specifies the values of the Atkinson index parameter. The maximum two values of atk() can be listed; if only one value is chosen, it is repeated twice. The default is atk(0.5 2).

Inference and reporting

bootstrap(*bootstrapstr*) allows the user to customize almost every aspect of the bootstrap procedure. *bootstrapstr* can be any valid option of the bootstrap command (see [R] **bootstrap**), including reps(), strata(), size(), saving(), level(), or seed(). The options notable, nolegend, and nowarn are already "built-in". The computation of the bootstrapped standard error can be avoided using the option bootstrap(off).

citype(*citypestr*) specifies how confidence intervals are to be computed and displayed. *citypestr* can be normal, percentile, or bc, which stand, respectively, for normal approximation, percentile method, and bias-corrected confidence intervals.

format(*formatstr*) displays results accordingly; see [D] **format**.

3.4 Stored results

igmobil stores results in e(). The results stored are the same as in any bootstrap command.

4 Examples

4.1 Preliminary: Artificial dataset

Let's generate Y_i and X_i from a bivariate lognormal distribution with parameters $\mu_Y = \mu_X = 0$, $\sigma^2_Y = \sigma^2_X = 0.25$, and $\rho = 0.5$. Then, let's generate Y_i^{disc} and X_i^{disc} to simulate the case where we have discrete random variables.

```
. clear
. matrix C = (.25, .5*.25 \ .5*.25, .25)
. set seed 12345
. drawnorm u0 u1, n(2000) cov(C)                        /* normal r.v. */
(obs 2,000)
. generate son = exp(u1)                                /* lognormal r.v.*/
. generate dad = exp(u0)
. generate son_disc = irecode(u1, -1, -0.5, 0, 0.5, 1)  /*discrete r.v.*/
. generate dad_disc = irecode(u0, -1, -0.5, 0, 0.5, 1)
. drop u*
```

4.2 Basic use of igmobil

In its simplest form, `igmobil` requires two inputs: the *child* variable and the *parent* variable, both expressed in levels (that is, no logs).

```
. igmobil son dad
(running igmobil_1 on estimation sample)

Bootstrap replications (50)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50

Bootstrap results                               Number of obs     =      2,000
                                                Replications      =         50

Child generation:      son = Y                  Type of variables: continuous
Parent generation:     dad = X
--------------------------------------------------------------------------------
     Type of indices                        IGM    Bootstrap   [95% Conf. Interv.]
                                         estimate  Std. Err.    normal approx.
--------------------------------------------------------------------------------
     Single-stage Indices
(1)  1/N * sum |X - Y|                    0.446     0.010       0.427     0.465
(2)  1/N * sum (X - Y)^2                  0.385     0.022       0.341     0.428
(3)  1/N * sum |ln X - ln Y|              0.395     0.007       0.382     0.408
(4)  1/N * sum |X/mu(X) - Y/mu(Y)|        0.389     0.006       0.377     0.402
(5)  1 - Pearson coef. (on logs)          0.487     0.015       0.457     0.517
(6)  1 - Spearman coef. (on logs)         0.511     0.016       0.479     0.543
(7)  1/N * sum |CDF X - CDF Y|            0.229     0.004       0.221     0.237
(8)  1/N * sum (CDF X - CDF Y)^2          0.085     0.003       0.080     0.090
(9)  1 - OLS(Y,X)                         0.540     0.026       0.489     0.591
(10) 1 - OLS(ln Y,ln X)                   0.499     0.018       0.464     0.534
--------------------------------------------------------------------------------
    Transition matrix Indices (based on 5 quantiles)
(11)  Shorrock/Prais                      0.849     0.013       0.823     0.875
(12)  Bartholomew                         0.269     0.005       0.259     0.279
(13)  1-Second largest eigenvalue         0.522     0.017       0.489     0.556
(14)  Determinant index                   1.000     0.000       1.000     1.000
--------------------------------------------------------------------------------
    Inequality related Indices
(15) Fields - Gini                        0.134     0.010       0.115     0.152
(16) Fields - GE(0)                       0.254     0.016       0.223     0.286
(17) Fields - GE(1)                       0.255     0.018       0.219     0.291
(18) Fields - Atkinson(.5)                0.249     0.017       0.216     0.281
(19) Fields - Atkinson(2)                 0.229     0.014       0.202     0.257
--------------------------------------------------------------------------------
```

The output indicates that the Bartholomew index ($\mathcal{M}_{12}$, the average "jump" of income classes) from our sample is 0.269 with a bootstrapped standard error of 0.005 and a symmetric 95% confidence interval of [0.259; 0.279] obtained with the normal approximation.

In the next example, we want our transition matrix to be based on 10 quantiles. Also, we omit the computation of single-stage and inequality-based indices.

```
. igmobil son dad, nosingle noinequal classes(10)
(running igmobil_1 on estimation sample)

Bootstrap replications (50)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50

Bootstrap results                               Number of obs     =      2,000
                                                Replications      =         50

Child generation:      son = Y                  Type of variables: continuous
Parent generation:     dad = X
--------------------------------------------------------------------------------
     Type of indices                      IGM    Bootstrap   [95% Conf. Interv.]
                                       estimate  Std. Err.    normal approx.
--------------------------------------------------------------------------------

    Transition matrix Indices (based on 10 quantiles)
(11)  Shorrock/Prais                      0.919     0.010       0.900     0.939
(12)  Bartholomew                         0.251     0.005       0.242     0.260
(13)  1-Second largest eigenvalue         0.508     0.016       0.476     0.540
(14)  Determinant index                   1.000     0.000       1.000     1.000
--------------------------------------------------------------------------------
```

We see that our $\mathcal{M}_{11}$–$\mathcal{M}_{14}$ indices change when we double the size of the transition matrix (for example, $\mathcal{M}_{11}$ increases, but $\mathcal{M}_{12}$ decreases). This should warn us against comparing transition-matrix indices derived from transition matrices that have different numbers of classes.

Now assume that the data are already in discrete form, such as income classes or educational achievement. Here we need to use the option `discrete`; otherwise, `igmobil` will assume continuous variables. Note that the option `discrete` automatically sets `nosingle` and `noinequal`, because single-stage and inequality indices cannot be computed from discrete variables.

```
. igmobil son_disc dad_disc, discrete
(running igmobil_1 on estimation sample)

Bootstrap replications (50)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50

Bootstrap results                               Number of obs     =      2,000
                                                Replications      =         50

Child generation:      son_disc = Y             Type of variables: discrete
Parent generation:     dad_disc = X
--------------------------------------------------------------------------------
     Type of indices                      IGM    Bootstrap   [95% Conf. Interv.]
                                       estimate  Std. Err.    normal approx.
--------------------------------------------------------------------------------

    Transition matrix Indices (original categories of X,Y)
(11)  Shorrock/Prais                      0.834     0.016       0.801     0.866
(12)  Bartholomew                         0.195     0.006       0.182     0.207
(13)  1-Second largest eigenvalue         0.547     0.020       0.507     0.586
(14)  Determinant index                   1.000     0.000       1.000     1.000
--------------------------------------------------------------------------------
```

4.3 igmobil and inequality-based indices

`igmobil` allows users to customize both the IGM indices family (from the default Fields to Shorrocks) and the parameters of the generalized entropy and the Atkinson indices (a maximum of two parameters each; if only one parameter is specified, it is repeated twice). These options can be combined with the others previously given.

In the following example, we estimate both 4×4 transition-matrix and inequality-based indices. For the last class, we specify that the Shorrocks family should be used.

```
. igmobil son dad, nosingle classes(4) family(shorrocks)
(running igmobil_1 on estimation sample)
Bootstrap replications (50)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
Bootstrap results                               Number of obs     =      2,000
                                                Replications      =         50

Child generation:      son = Y                  Type of variables: continuous
Parent generation:     dad = X
--------------------------------------------------------------------------------
     Type of indices                      IGM    Bootstrap   [95% Conf. Interv.]
                                       estimate  Std. Err.    normal approx.
--------------------------------------------------------------------------------
    Transition matrix Indices (based on 4 quantiles)
(11)  Shorrock/Prais                     0.807     0.016       0.776     0.837
(12)  Bartholomew                        0.282     0.007       0.269     0.295
(13)  1-Second largest eigenvalue        0.547     0.018       0.512     0.582
(14)  Determinant index                  0.999     0.001       0.997     1.001
--------------------------------------------------------------------------------
    Inequality related Indices
(15) Shorrocks - Gini                    0.140     0.015       0.110     0.170
(16) Shorrocks - GE(0)                   0.265     0.024       0.217     0.313
(17) Shorrocks - GE(1)                   0.266     0.029       0.209     0.323
(18) Shorrocks - Atkinson(.5)            0.259     0.026       0.209     0.310
(19) Shorrocks - Atkinson(2)             0.239     0.021       0.198     0.280
--------------------------------------------------------------------------------
```

In the next example, we modify the parameter of the generalized entropy (2, for which the generalized entropy becomes the half-squared coefficient of variation) and of the Atkinson index (2 and 5).

```
. igmobil son dad, nosingle notrans family(fields) ge(2) atk(2 5)
(running igmobil_1 on estimation sample)

Bootstrap replications (50)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50

Bootstrap results                               Number of obs     =      2,000
                                                Replications      =         50

Child generation:      son = Y                  Type of variables: continuous
Parent generation:     dad = X
--------------------------------------------------------------------------------
     Type of indices                       IGM    Bootstrap   [95% Conf. Interv.]
                                        estimate  Std. Err.    normal approx.
--------------------------------------------------------------------------------

    Inequality related Indices
(15) Fields - Gini                        0.134     0.010        0.114     0.154
(16) Fields - GE(2)                       0.278     0.023        0.233     0.324
(17) Fields - GE(2)                       0.278     0.023        0.233     0.324
(18) Fields - Atkinson(2)                 0.229     0.016        0.198     0.261
(19) Fields - Atkinson(5)                 0.194     0.022        0.150     0.237
--------------------------------------------------------------------------------
```

4.4 Standard errors and confidence intervals

One important feature of `igmobil` is the embedded bootstrap procedure, which allows users to make inference on the estimated indices without the need of extra programming. By default, the number of replications is set to 50 to reduce computational burden. In the next example, we modify the relevant `bootstrap()` option by increasing the number of replications and saving each bootstrap result for later use. We also fix the `seed` number so that results can be replicated, and we omit replication dots to preserve space.

```
. igmobil son dad, nosingle noinequal
> bootstrap(reps(200) seed(12345) saving(myfile, replace) nodots)

Bootstrap results                               Number of obs     =      2,000
                                                Replications      =        200

Child generation:      son = Y                  Type of variables: continuous
Parent generation:     dad = X
--------------------------------------------------------------------------------
     Type of indices                       IGM    Bootstrap   [95% Conf. Interv.]
                                        estimate  Std. Err.    normal approx.
--------------------------------------------------------------------------------

    Transition matrix Indices (based on 5 quantiles)
(11)  Shorrock/Prais                      0.849     0.015        0.820     0.878
(12)  Bartholomew                         0.269     0.006        0.257     0.281
(13)  1-Second largest eigenvalue         0.522     0.019        0.485     0.560
(14)  Determinant index                   1.000     0.000        1.000     1.000
--------------------------------------------------------------------------------
```

Comparing the output with the corresponding part of the table shown in section 4.2, we see that the standard errors and the (normal approximated) confidence intervals hardly changed when we went from 50 to 200 bootstrap replications. However, this might be due to the specific data-generating process used for this example.

In this example, we specify that the program displays the 99% percentile method confidence interval. Note that the confidence level is set as a suboption of the `bootstrap()` option, while the type of confidence interval is chosen with the option `citype()`.

```
. igmobil son dad, nosingle noinequal
> boot(reps(200) seed(12345) saving(myfile, replace) level(99) nodots)
> citype(percentile)
Bootstrap results                                Number of obs     =      2,000
                                                 Replications      =        200
Child generation:      son = Y                   Type of variables: continuous
Parent generation:     dad = X
--------------------------------------------------------------------------------
     Type of indices                      IGM    Bootstrap   [99% Conf. Interv.]
                                       estimate  Std. Err.    percentile method
--------------------------------------------------------------------------------
    Transition matrix Indices (based on 5 quantiles)
(11)  Shorrock/Prais                     0.849     0.015       0.817     0.903
(12)  Bartholomew                        0.269     0.006       0.257     0.288
(13)  1-Second largest eigenvalue        0.522     0.019       0.482     0.568
(14)  Determinant index                  1.000     0.000       1.000     1.000
--------------------------------------------------------------------------------
```

In this case, we can no longer expect the confidence interval to be symmetric with respect to the estimated coefficient. Here the estimated Bartholomew index (0.269) is slightly closer to the left bound of the interval (0.257) than to the right one (0.288).

4.5 Reporting and postestimation

Because `igmobil` is an estimation command, we can easily recall parameter estimates and standard errors by typing `_b[i#]` and `_se[i#]`, where # is the IGM progressive number. We first recall the last estimate of the Eigenvalue2 index ($\mathcal{M}_{13}$), and then we test the hypothesis that this index equals 0.5. We find that the Wald test would not reject the null hypothesis at the 5% level.

```
. display _b[i13]
.52210281
. test i13 = 0.5
 ( 1)  i13 = .5
           chi2(  1) =    1.33
         Prob > chi2 =    0.2485
```

Another advantage of an estimation command is that it can be used with the commands `estimates store`, `estimates restore`, `estimates table`, etc., to manipulate estimation results. Assume that a quarter of the data belongs to country A and the remaining to country B and that we want to estimate the $\mathcal{M}_{11}$, $\mathcal{M}_{12}$, and $\mathcal{M}_{13}$ separately for those countries. We can then apply the following commands:

```
. generate country = cond(_n<= 500, "A", "B")
. quietly igmobil son dad if country == "A",  nosingle noinequal
    Transition matrix Indices (based on 5 quantiles)
. estimate store igm_A
```

```
. quietly igmobil son dad if country == "B",  nosingle noinequal
    Transition matrix Indices (based on 5 quantiles)
. estimate store igm_B
. estimates table igm_A igm_B, stats(N) b(%9.4f) se(%9.4f) keep(i11-i13)
```

Variable	igm_A	igm_B
i11	0.8700	0.8475
	0.0286	0.0153
i12	0.2680	0.2713
	0.0111	0.0070
i13	0.5106	0.5299
	0.0499	0.0235
N	500	1500

legend: b/se

4.6 Advanced use: Adding a new IGM index

Assume that we want to add the Bhattacharya and Mazumder (2011) upward-mobility (UP) index described in section 2 to our standard transition-matrix indices. In particular, we want to estimate the probability that a child's percentile exceeds the one of his or her parent by 10, given that the parent belonged to the lowest quartile. In other words, we seek the sample counterpart of $\text{UP}_{\tau,s} = \text{UP}_{0.10,0.25} = \Pr(r_{Y_i} - r_{X_i} > 0.10 |\, r_{X_i} \leq 0.25)$.

To do this, we write an r-class program that returns the desired IGM estimate in a macro named UW (which stands for user written) and save it as a do-file.

```
* upward-mobility index
capture program drop myindex
        program myindex, rclass
        syntax varlist(min=2 max=2 numeric) [if] [in] [, tau(real 0) s(real 0.25)]

        marksample touse

        tempvar y x ry rx diff
        tempname num den

        tokenize ‘varlist’
        quietly {
                generate ‘y’  = ‘1’ if ‘touse’
                generate ‘x’  = ‘2’ if ‘touse’

                cumul ‘y’, gen(‘ry’)
                cumul ‘x’, gen(‘rx’)

                count if (‘ry’-‘rx’) > ‘tau’ & ‘rx’ <= ‘s’ & ‘touse’
                scalar ‘num’ = r(N)

                count if ‘rx’ <= ‘s’ & ‘touse’
                scalar ‘den’ = r(N)
```

```
                return scalar UW = `num'/`den'
        }
end
```

Finally, we incorporate the myindex command into igmobil:

```
. quietly do myindex.do
. igmobil son dad, nosingle noinequal
> userwritten(myindex son dad, tau(0.1) s(0.25)) classes(4)
(running igmobil_1 on estimation sample)
Bootstrap replications (50)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
Bootstrap results                               Number of obs     =      2,000
                                                Replications      =         50
Child generation:      son = Y                  Type of variables: continuous
Parent generation:     dad = X
--------------------------------------------------------------------------------
     Type of indices                       IGM    Bootstrap   [95% Conf. Interv.]
                                       estimate  Std. Err.    normal approx.
--------------------------------------------------------------------------------
    Transition matrix Indices (based on 4 quantiles)
(11)  Shorrock/Prais                     0.807     0.015       0.778     0.835
(12)  Bartholomew                        0.282     0.006       0.269     0.294
(13)  1-Second largest eigenvalue        0.547     0.019       0.509     0.584
(14)  Determinant index                  0.999     0.001       0.996     1.001
--------------------------------------------------------------------------------
(20) User written program                0.550     0.015       0.520     0.580
--------------------------------------------------------------------------------
```

Among the families whose parents belonged to 25% of the poorest, 55.0% of the children outperformed the rank of their parents by more than 0.1 (or 10%).

5 Acknowledgments

I thank the editor and an anonymous referee for helpful comments; any remaining errors are my responsibility. Any views expressed in this article are those of the author and do not necessarily reflect those of the Bank of Italy.

6 References

Bartholomew, D. J. 1973. *Stochastic Models for Social Processes*. New York: Wiley.

Bhattacharya, D., and B. Mazumder. 2011. A nonparametric analysis of black–white differences in intergenerational income mobility in the United States. *Quantitative Economics* 2: 335–379.

Black, S. E., and P. J. Devereux. 2011. Recent developments in intergenerational mobility. In *Handbook of Labor Economics*, ed. O. Ashenfelter and D. Card, vol. 4B, 1487–1541. San Diego: Elsevier.

Chakravarty, S. R., B. Dutta, and J. A. Weymark. 1985. Ethical indices of income mobility. *Social Choice and Welfare* 2: 1–21.

Checchi, D., and V. Dardanoni. 2002. Mobility comparisons: Does using different measures matter? Departmental Working Papers 2002-15, Department of Economics, Management, and Quantitative Methods at Università degli Studi di Milano. http://ideas.repec.org/p/mil/wpdepa/2002-15.html.

Cowell, F. A. 2000. Measurement of inequality. In *Handbook of Income Distribution, vol. 1*, ed. A. B. Atkinson and F. Bourguignon, chap. 2. The Netherlands: Elsevier.

Cowell, F. A., and C. Schluter. 1998. Income mobility: A robust approach. Discussion paper DARP/37, Suntory and Toyota International Centres for Economics and Related Disciplines, London School of Economics and Political Science. http://sticerd.lse.ac.uk/dps/darp/darp37.pdf.

Fields, G. S. 2008. Income mobility. In *The New Palgrave Dictionary of Economics*, ed. S. N. Durlauf and L. E. Blume, 2nd ed. London: Palgrave Macmillan. http://www.dictionaryofeconomics.com/article?id=pde2008_I000271.

———. 2010. Does income mobility equalize longer-term incomes? New measures of an old concept. *Journal of Economic Inequality* 8: 409–427.

Fields, G. S., and E. A. Ok. 1996. The meaning and measurement of income mobility. *Journal of Economic Theory* 71: 349–377.

———. 1999a. The measurement of income mobility: An introduction to the literature. In *Handbook of Income Inequality Measurement, vol. 71*, ed. J. Silber, 557–598. The Netherlands: Springer.

———. 1999b. Measuring movement of incomes. *Economica* 66: 455–471.

Hart, P. E. 1981. The Statics and Dynamics of Income Distributions: A Survey. In *The Statics and Dynamics of Income*, ed. N. A. Klevmarken and J. A. Lybeck. Clevedon, UK: Tieto.

Jäntti, M., and S. P. Jenkins. 2013. Income Mobility. SOEPpapers No. 607, German Institute for Economic Research. http://www.diw.de/documents/pubikationen/73/diw_01.c.432440.de/diw_sp0607.pdf.

Jenkins, S. P. 2006. Estimation and interpretation of measures of inequality, poverty, and social welfare using Stata. 2006 North American Stata Users Group meeting proceedings. http://ideas.repec.org/p/boc/asug06/16.html.

Shorrocks, A. 1978a. Income inequality and income mobility. *Journal of Economic Theory* 19: 376–393.

Shorrocks, A. F. 1978b. The measurement of mobility. *Econometrica* 1013–1024.

Sommers, P. M., and J. Conlisk. 1979. Eigenvalue immobility measures for Markov chains. *Journal of Mathematical Sociology* 6: 253–276.

Appendix: Generalized entropy and Atkinson indices

In this appendix, I provide some background on the generalized entropy measure and the Atkinson index, relying mostly on Jenkins (2006) and Cowell (2000). The generalized entropy family takes the following form:

$$\mathrm{GE}(a) = \frac{1}{Na(a-1)} \sum_{i=1}^{N} \left\{ \left(\frac{x_i}{\mu_x} \right)^a - 1 \right\} \qquad a \neq [0,1]$$

$$\mathrm{GE}(a) = \frac{1}{N} \sum_{i=1}^{N} \left\{ \frac{x_i}{\mu_x} \log \left(\frac{x_i}{\mu_x} \right) \right\} \qquad a = 1$$

$$\mathrm{GE}(a) = \frac{1}{N} \sum_{i=1}^{N} \log \left(\frac{\mu_x}{x_i} \right) \qquad a = 0$$

The parameter a captures the sensitivity of the $\mathrm{GE}(a)$ family to a particular part of the distribution: a large positive a increases sensitivity to changes in the upper tail, and a negative a increases sensitivity to changes in the lower tail. For specific values of a, the $\mathrm{GE}(a)$ assumes the following known forms: $\mathrm{GE}(0)$ is the mean log deviation, $\mathrm{GE}(1)$ is the Theil index, and $\mathrm{GE}(2)$ is the half-squared coefficient of variation. The default option will have $a_1 = 0$ and $a_2 = 1$.

The Atkinson index is a welfare-based measure of inequality that assumes an explicit formulation of the social welfare function (that is, the way individual utilities are aggregated) and an explicit level of income inequality aversion. At the core of the Atkinson index is the equally distributed equivalent income, $y_e \leq \mu_y$, which is the income value that—if equally distributed—would equal the same level of social welfare as the actual income distribution. The larger the difference between μ_Y and y_e, the higher the cost of inequality. Finally,

$$\mathrm{Atk}(\epsilon) = 1 - \left\{ \frac{1}{N} \sum_{i=1}^{N} \left(\frac{x_i}{\mu_x} \right)^{1-\epsilon} \right\}^{\frac{1}{1-\epsilon}} \qquad \epsilon \geq 0, \epsilon \neq 1$$

$$\mathrm{Atk}(\epsilon) = 1 - \exp \left\{ \frac{1}{N} \sum_{i=1}^{N} \left(\frac{x_i}{\mu_x} \right) \right\} \qquad \epsilon = 1$$

where the parameter ϵ captures the inequality aversion in a society. A larger ϵ means that the society is more inequality averse. The default option will have $\epsilon_1 = 0.5$ and $\epsilon_2 = 2$.

About the author

Marco Savegnago is an economist at the Bank of Italy.