Rivista Italiana di Economia Demografia e Statistica Volume LXVIII n. 3/4 Luglio-Dicembre 2014

IS ITALY A MELTING POT?

Claudio Ceccarelli, Giovanni Maria Giorgi, Alessio Guandalini

1. Introduction

A melting pot is a metaphor for a society where many different types, mainly for ethnicity, race and consequently for culture, of people blend together as one. In an ideal situation it is a society in which these differences do not affect the social status of people. The United States is the classic example of a melting pot. However, there are other several examples in the world such as Afghanistan, Brazil and Israel.

Historically, Italy has always been an emigration country. Only since the seventies has started to become an immigration country. Earlier this shift to immigration was due to its economic situation and, later, mainly, for its position as the entry door of the Eurozone. Therefore, the migration problem and the migration policies are quite recent.

Nowadays, among the European countries, Italy ranks third for absolute number of foreign inhabitants (4.8 million) and eleventh for percentage of foreigners in the total population (5.5%). This work aims to evaluate the integration process of immigrants in Italy and see if our country can be considered a melting pot. Looking at the employee income, an ideal situation in which the foreign inhabitants can be considered integrated, at least for the employee wages, occurs if their incomes overlap with incomes of Italian inhabitants. On the contrary, we could state that the migration policies have been completely erroneous if the foreign inhabitants are the poorest whilst the Italians are the richest. That is, if the population is perfectly stratified.

The peculiarity of the work is represented by the tool used in evaluating the integration process and the migration policies, the analysis of Gini (ANOGI). The ANOGI is similar to the ANOVA (analysis of variance), but it offers an additional parameter: the stratification that enables us to better interpret the results. The work is more focused on the methodological aspects. In the first part, Section 2, the methodological differences between the ANOGI and the ANOVA are investigated. In Section 3, through the application on Italian Labour Force Survey 2007 and 2012 data the differences between the two methods are better clarified. Finally, an analysis of the integration process of immigrants is carried out.

2. Analysis of Gini (ANOGI) and analysis of variance (ANOVA)

2.1 ANOVA

The ANOVA is a well-known method to evaluate the differences between group means and their associated procedure. In the ANOVA setting, the observed variance in a particular variable is partitioned into components attributable to different sources of variation.

In the simplest case, the one-way ANOVA, the data $\boldsymbol{Y_{ij}}$ are assumed to be

$$\boldsymbol{Y_{ij} = \mu + a_i + \varepsilon_{ij}}, \qquad \boldsymbol{i = 1, \dots, k} \quad \boldsymbol{j = 1, \dots, n_i}.$$

In this formulation the values $\boldsymbol{Y_{ij}}$ are expressed in function of a grand mean, $\boldsymbol{\mu}$, that is the common mean level of the treatment (or variable modality), and the unique effect due to treatment (or variable modality) $\boldsymbol{a_i}$, besides the errors $\boldsymbol{\varepsilon_{ij}}$.

The expected value of the errors are assumed to be independent and normally distributed with 0 mean and finite variance $\boldsymbol{\sigma^2}$ equal for all the $\boldsymbol{i}$ (homoschedasticity). In formulas

i. $\mathbb{E}[\boldsymbol{\varepsilon_{ij}}] = \boldsymbol{0}$;
ii. $\mathrm{Var}(\boldsymbol{\varepsilon_{ij}}) = \boldsymbol{\sigma_i^2} < \infty$;
iii. $\boldsymbol{\sigma_i^2 = \sigma^2} \ \forall \boldsymbol{i}$;
iv. $\mathrm{Cov}(\boldsymbol{\varepsilon_{ij}, \varepsilon_{i'j'}}) = \boldsymbol{0}$ with $\boldsymbol{i \neq i'}$ and $\boldsymbol{j \neq j'}$;
v. $\boldsymbol{\varepsilon_{ij} \sim N(0, \sigma^2)}$;

The basic idea of the ANOVA is that the variation is allocated to different sources. In fact, the overall variation of a measurable variable (left-hand side) is decomposed in two terms (right-hand side): between variation due only to treatments and within variation due only to random error, respectively. That is,

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}\right)^2 = \sum_{i=1}^{k} n_i\,(\bar{y}_i - \bar{y})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}_i\right)^2$$

where $\boldsymbol{\bar{y}_i = \frac{1}{n_i}\sum_j y_{ij}}$ and $\boldsymbol{\bar{y} = \sum_j n_i\, y_i / \sum_i n_i}$. The corrected (by degree of freedom) sums of squares, under the ANOVA assumptions, are chi squared random variables. In particular, the left-hand side is distributed as a $\boldsymbol{\chi^2_{N-1}}$ while, under the null hypothesis (equal means among the groups), the right-hand side is the sum of two independent random variables distributed, respectively, as $\boldsymbol{\chi^2_{k-1}}$ and $\boldsymbol{\chi^2_{N-k}}$.

2.2 ANOGI

The ANOGI was firstly proposed by Frick et al. (2006). It is based on the Gini index that in a population $\boldsymbol{P}$ is defined as (Lerman and Yitzhaki, 1989, p. 44)

$$\boldsymbol{G} = \frac{\mathbf{2\ cov}\left(\boldsymbol{y}, \boldsymbol{F}(\boldsymbol{y})\right)}{\boldsymbol{\mu}},$$

that is, twice the covariance between the income $\boldsymbol{y}$ and the rank $\boldsymbol{F}(\boldsymbol{y})$, standardized by mean income $\boldsymbol{\mu}$. When the population is divided in $\boldsymbol{k}$ groups, $\boldsymbol{P} = \boldsymbol{P_1} \cup \boldsymbol{P_2} \cup \ldots \cup \boldsymbol{P_k}$, the Gini index can be expressed as (Yitzhaki, 1994, p. 154)

$$\boldsymbol{G_u} = \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{k}} \boldsymbol{s_i G_i O_i} + \boldsymbol{G_b}, \qquad (1)$$

that is, the Gini index is decomposed in two components: within and between, where

i. $\boldsymbol{s_i} = \boldsymbol{p_i}\, \boldsymbol{\mu_i}/\boldsymbol{\mu}$ is the ratio between the mean of variable $\boldsymbol{y}$ in the group $\boldsymbol{i}$, $\boldsymbol{\mu_i}$, weighted by its share, $\boldsymbol{p_i}$, and the mean of $\boldsymbol{y}$ calculated on the whole population;
ii. $\boldsymbol{G_i}$ is the Gini index within group $\boldsymbol{i}$;
iii. $\boldsymbol{O_i}$ is the overlapping index of group $\boldsymbol{i}$ with the entire population;
iv. $\boldsymbol{G_b}$ is the between-group inequality.

Two elements in (1) must be pointed out: overlapping and between-group inequality. Overlapping should be interpreted as the inverse of stratification (see, e.g., Yitzhaki, 1988, p. 39; Yitzhaki and Lerman, 1991, p. 319). It measures to what extent one group is overlapped by the other. The overlapping index $\boldsymbol{O_i}$ may be expressed as

$$\boldsymbol{O_i} = \frac{\mathbf{cov}_{\boldsymbol{i}}\left(\boldsymbol{y}, \boldsymbol{F_u}(\boldsymbol{y})\right)}{\mathbf{cov}_{\boldsymbol{i}}\left(\boldsymbol{y}, \boldsymbol{F_i}(\boldsymbol{y})\right)},$$

that is the ratio between the covariance of $\boldsymbol{y}$ and the rank of units belonging to group $\boldsymbol{i}$, calculated on their position in the overall distribution, and one-forth of Gini's mean difference of group $\boldsymbol{i}$ (see Yitzhaki and Schechtman, 2009, p. 149).

The overlapping index related to a given group $\boldsymbol{i}$ can be written in terms of the overlapping index between two groups, $\boldsymbol{i}$ and $\boldsymbol{j}$,

$$\boldsymbol{O_i} = \sum_{\boldsymbol{j}} \boldsymbol{p_j}\, \boldsymbol{O_{ij}} = \boldsymbol{p_i O_{ii}} + \sum_{\boldsymbol{j \neq i}} \boldsymbol{p_j}\, \boldsymbol{O_{ji}} = \boldsymbol{p_i} + \sum_{\boldsymbol{j \neq i}} \boldsymbol{p_j}\, \boldsymbol{O_{ji}}$$

where

$$\boldsymbol{O_{ji}} = \frac{\mathbf{cov}_{\boldsymbol{i}}\left(\boldsymbol{y}, \boldsymbol{F_j}(\boldsymbol{y})\right)}{\mathbf{cov}_{\boldsymbol{i}}\left(\boldsymbol{y}, \boldsymbol{F_i}(\boldsymbol{y})\right)}$$

represents the overlapping index of group $\boldsymbol{j}$ by group $\boldsymbol{i}$ (Yitzhaki, 1994). In particular:

i. $\boldsymbol{O_{ji}} = \mathbf{0}$, when no member of group $\boldsymbol{j}$ lies in the range of subgroup $\boldsymbol{i}$;
ii. $\boldsymbol{O_{ji}} = \mathbf{1}$, the distributions of group $\boldsymbol{i}$ and $\boldsymbol{j}$ are identical;
iii. $\boldsymbol{O_{ji}}$ is not symmetrical, that is the higher $\boldsymbol{O_{ji}}$ the lower $\boldsymbol{O_{ij}}$;
iv. $\boldsymbol{O_{ji}} \leq \mathbf{2}$; that is its maximum value, if all the members of group $\boldsymbol{j}$ are included between the members of group $\boldsymbol{i}$ and they are concentrated around the mean of group $\boldsymbol{i}$.

The between group inequality

$$\boldsymbol{G_b} = \frac{\mathbf{2}\ \mathbf{cov}\left(\boldsymbol{y}, \overline{\boldsymbol{F}}_{\boldsymbol{ui}}(\boldsymbol{y})\right)}{\boldsymbol{\mu_u}},$$

which is the ratio between twice the covariance between the mean of variable $\boldsymbol{y}$ of each group and the groups mean rank in the whole population and the mean of $\boldsymbol{y}$.

When the population is perfectly stratified the between-group inequality is equal to the between-group-Pyatt inequality, (Pyatt, 1976, p. 247)

$$\boldsymbol{G_b^p} = \frac{\mathbf{2}\ \mathbf{cov}\left(\boldsymbol{\mu_i}, \overline{\boldsymbol{F}}_{\boldsymbol{i}}(\boldsymbol{y})\right)}{\boldsymbol{\mu_u}}.$$

Yitzhaki and Lerman (1991, p. 322) demonstrated that $\boldsymbol{G_b^p} \geq \boldsymbol{G_b}$. In fact, $\boldsymbol{G_b}$ reaches its upper level as the overlapping index is equal to 0 and, therefore, the amount of total inequality is explained by the between inequality.

Introducing the between-group-Pyatt inequality, (1) can be written as

$$\begin{aligned}\boldsymbol{G_u} &= \sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} \boldsymbol{s_i G_i} + \sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} \boldsymbol{s_i G_i}(\boldsymbol{O_i} - \mathbf{1}) + \boldsymbol{G_b^p} + \left(\boldsymbol{G_b} - \boldsymbol{G_b^p}\right) = \\ &= \boldsymbol{IG} + \boldsymbol{IGO} + \boldsymbol{BG} + \boldsymbol{BGO} \qquad (2)\end{aligned}$$

that is, in terms of the four elements at the basis of ANOGI: the within (IG) and the between-group (BG) components and the effects of overlapping on within and between-group component, IGO and BGO, respectively.

2.3 Similarities and differences between ANOVA and ANOGI

The ANOVA and the ANOGI perform the same task; that is, they decompose a measure of variability, variance or Gini index respectively, and assign it to different sources of variation. Their components are conceptually comparable. As briefly illustrated in Table 1, IG has the same meaning as SSW in the ANOVA and the BG as SSB. In other words, both methods decompose the variability into two quantities: the difference within the groups and the difference between the groups.

Table 1 – *Comparison among components of ANOVA and ANOGI.*

	ANOVA		ANOGI	
Within	$SSW = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2$	$0 \leq SSW \leq SST$	$IG = \sum_{i=1}^{k} s_i G_i$	$0 \leq IG \leq G_u$
Between	$SSB = \sum_{i=1}^{k} n_i\,(\bar{y}_i - \bar{y})^2$	$0 \leq SSB \leq SST$	$BG = G_b^p$	$0 \leq BG \leq G_u$
Overlapping Within			$IGO = \sum_{i=1}^{k} s_i G_i (O_i - 1)$	
Overlapping Between			$BGO = (G_b - G_b^p)$	$-BG - IGO - IG \leq BGO \leq 0$

Moreover, to extra parameters linked to the overlapping, IGO and BGO, are derived with the ANOGI. IGO provides the contribution of each group to within group variability and tell us how much the distributions are intertwined and, therefore, how much the groups are integrated with one another. BGO is related to the effect of overlapping on the between-group inequality. It is always negative, because the overlapping reduce the ability to distinguish between groups.

3. The degree of melting pot

The advantage of the ANOGI with respect to the ANOVA is that it says how much a population is stratified and, on the contrary, how much the groups are intertwined. In this paper the ANOGI is used to investigate the integration of immigrants into the labour market in terms of employee wages. This paper traces out the work by Yitzhaki and Schecthman (2009).

From the Labour Force Survey 2007 and 2012 the employees older than thirty have been selected in order to avoid the effect of different fertility rates between Italians and immigrants. The employees have been split in three main categories, Italians, immigrants and second-generation immigrants, through the variables gathered and in the questionnaire and in accordance with the Italian laws in matter regarding citizenship[1]. Furthermore the immigrants are also classified by geographical areas of origin (Europe, North-America, Center-America, South-America, Africa, North-Africa, Asia, Middle-East, China and Oceania).

The employees classified as second-generation immigrants in one case are aggregate to the Italians (wide classification, W) and, in another case, to the immigrants categorized by their geographical areas of origin (narrow classification, N). In both cases the ANOVA and the ANOGI are applied and the results obtained separately for each classification are compared to derive conclusions on the immigrants' integration.

3.1 ANOVA results

The ANOVA decomposes the total amount of variance in two quantities, between and within (Table 2).

Table 2 – *Results of the ANOVA analysis on Labour Force Survey data of 2007 and 2012.*

		MS between	MS within	Total (df)	SS Between (df)	SS within (df)	F
2007	N	55,640,686	312,283	45,691,375,078 (144,365)	612,047,549 (11)	45,079,327,529 (144,354)	178.17
	W	56,790,114	312,196	45,691,375,078 (144,365)	624,691,252 (11)	45,066,683,826 (144,354)	181.91
2012	N	142,358,333	324,943	44,166,251,741 (131,112)	142,358,333 (11)	42,600,310,082 (131,101)	438.10
	W	4,832,205	336,482	44,166,251,741 (131,112)	53,154,254 (11)	44,113,097487 (131,101)	14.36

Looking at the F ratio the MS between is larger for definition W than for N in 2007 while, in 2012 the contrary occurs. The evidence that the null hypothesis (equal means among the groups) must be rejected is stronger in these cases[2]. This means that in 2007, when the second-generation immigrants is classified as Italians a better stratification is performed while, in 2012, a better classification is reached when the second-generation immigrants is classified as foreigners.

[1] In the 2007's sample the employed were about 145 thousand representative of 12,7 millions in the population: 132 thousand were Italians, 7,5 thousand were immigrants and 4 thousand were second-generation immigrants, representative of 12.3, 0.9 and 0.4 millions of employed in the population, respectively.
In the 2012's sample the employed became about 131 thousand representative of 13,3 millions in the population: 113 thousand were Italians, 13,9 thousand were immigrants and 4,2 thousand were second-generation immigrants, representative of 12.3, 1.6 and 0.4 millions of employed in the population, respectively.

[2] Even considering the Welch's test (Welch, 1947) in the case of non-homogeneity of the variances the evidence is to reject the null hypothesis.

3.2 ANOGI results

Performing the ANOGI on the same data, it is possible to decompose the Gini index into Gini between-groups, Gini within-groups and overlapping. In 2007 the Gini between groups (G_b and also G_b^p) is larger for W – with respect to N – even if the values are close to one another. Instead, in 2012 the Gini between-groups is larger for N than for W. The overlapping index of N definition decreases from 2007 to 2012 whilst that of W definition increases and, therefore, the gap between the two indices becomes larger. This means that in 2007, when the second-generation immigrants are classified as Italians a better stratification is performed, whilst in 2012 a better classification is reached when the second-generation immigrants are classified as foreigners.

In all cases the larger part of the inequality is explained by the within groups inequality (SGO). The overlapping that affected the within inequality is negligible and almost all affects the between-groups inequality. Therefore, the ratio between G_b and G_b^p is crucial to evaluate the stratification of the employee wages. In 2007 a better stratification is obtained for definition W, whilst in 2012 for definition N. This means that in 2007 the second generation of immigrants had employee wages more similar to the Italians, but this is not true for 2012. Therefore, it is possible to state that the integration process had suffered a setback.

Table 3 – *Results of the ANOGI analysis on Labour Force Survey data of 2007 and 2012.*

	Overall Gini	Definition	SGO		G_b		G_b^p	G_b/G_b^p
2007	0.2214 (0.0008)	N	0.2153	97.27%	0.0061	2.73%	0.0172	0.355
		(SE)	(0.0008)		(0.0003)		(0.0005)	
		W	0.2151	97.18%	0.0062	2.82%	0.0152	0.408
		(SE)	(0.0009)		(0.0005)		(0.0002)	
2012	0.2258 (0.0008)	N	0.2137	94.64%	0.0121	5.36%	0.0302	0.401
		(SE)	(0.0008)		(0.0003)		(0.0005)	
		W	0.2254	99.84%	0.0004	0.16%	0.0025	0.160
		(SE)	(0.0008)		(0.0001)		(0.0002)	

4. Conclusion

The ANOVA and the ANOGI perform the same task, but the latter provides an extra parameter, the overlapping, that is useful to better interpret the results. The two methods have been applied to the employee wages from the Labour Force Survey of 2007 and 2010 to investigate the integration of immigrant in the Italian society and, in particular, the labour market but, moreover, to point out the similarities and differences between the two methods. Both the results of the ANOVA and of the ANOGI demonstrate that there was a step back in the integration process from 2007 to 2012. Looking at the ANOGI results, it is possible to state that the second generation of immigrants was better integrated in 2007 than in 2012. However, in the global evaluation of the results it is important to point out that the application refers to employees with regular labour contract who have a higher level of integration in Italian society.

Acknowledgements
The present work has been realized within the grant for the project "Indici classici di disuguaglianza e variabilità: nuove prospettive di ricerca" (Sapienza 2013).

References

FRICK J.R., GOEBEL J., SCHECHTMAN E., WAGNER G.G., YITZHAKI S. (2006). Using Analysis of Gini (ANOGI) for Detecting Whether Two Sub-Sample Represent the Same Universe: The German Socio-Economic Panel Study (SOEP) Experience. *Sociological Methods & Research*, Vol. 34, No. 4, pp. 427-468.

LERMAN R.I., YITZHAKI S. (1984). A Note on the Calculation and Interpretation of the Gini Index. *Economics Letters*, Vol. 15, No. 3-4, pp. 363-368.

PYATT G. (1976).On the Interpretation and disaggregation of Gini Coefficient. *Economic Journal*, Vol. 86, No. 342, pp. 243-255.

YITZHAKI S. (1988). On Stratification and Inequality in Israel. *Bank of Israel Economic Review*, Vol. 63, No. 1-2, pp. 36-51.

YITZHAKI S., LERMAN R.I. (1991). Income Stratification and Income Inequality. *Review of Income and Wealth*, No. 37, No. 3, pp. 313-329.

YITZHAKI S., SCHECHTMAN R.I. (2009). The "Melting Pot": A Success Story?. *Journal of Economic Inequality*, Vol. 7, No. 2, pp. 137-151.

WELCH B.L. (1947). The generalization of Student's problem when several different population variance are involved. *Biometrika*, Vol. 34, No. 1-2, pp. 28-35.

SUMMARY

The immigrants integration process in Italy is investigated through the analysis of Gini (ANOGI). This methodology has an advantage with respect to the analysis of variance (ANOVA) because it provides a further element: the overlapping index, split in overlapping between and within the groups. This enables us to better understand and examine the immigrants integration looking at the stratification of the subpopulation of Italians and immigrants. The ANOGI is compared to the ANOVA and, then, the two methods are applied to Italian Labour Force Survey data of 2007 and 2012.

Claudio CECCARELLI, Italian National Institute of Statistics, Social and environmental statistics department, clceccar@istat.it.

Giovanni Maria GIORGI, "Sapienza" University of Rome, Department of Statistical Sciences, giovanni.giorgi@uniroma1.it.

Alessio GUANDALINI, "Sapienza" University of Rome, Department of Statistical Sciences, alessio.guandalini@uniroma1.it.