

Determinants of inequality in Italy: An approach based on the Shapley decomposition

Martina Celidoni ¹⁺, Isabella Procidano ² and Luca Salmasi ³

¹ Department of Economics, University of Padua

² Department of Economics, University of Venice

³ Department of Economics, University of Verona

Abstract. *Inequality is a common concern, strongly perceived by people as well as by governments that redistribute resources through taxation and public expenditure. In order to design their redistributive policies, governments need information about the underlying factors explaining how resources are distributed and this paper addresses this topic. We propose an analysis of the inequality determinants in Italy using the Shapley value decomposition on pseudo-panel data and taking into account the geographical specificity of inequality.*

Keywords: Inequality, Gini index, Shapley decomposition, Pseudo-panel

JEL Codes: D31; C10

1.Introduction

The idea that economic developments of the last few years have not been shared fairly is a common thought according to OECD²⁸. Even if people perceive inequality seriously²⁹, looking at the OECD report 2008 about income distribution and poverty, data show that the increase in inequality, although widespread and significant, has not been as huge as most people probably think it has been.

Actually the recorded moderate increase, in developed countries, over the past two decades hides the relevant presence of governments that, redistributing resources, have tried to slow down the growth in inequality. Information about the relationship between poverty and personal characteristics is therefore important to identify those individuals who are vulnerable to poverty and improve the targeting of anti-poverty policy measures (Biewen and Jenkins, 2005).

This paper tries to look behind poverty trends over time and to understand the relationship between poverty and personal characteristics; it presents the dynamic of Italy's inequality in the period 1997-2004 using the Shapley value decomposition (Shorrocks, 1982, 1999).

So far, the methodology has been used with cross-sectional data about income distribution, ignoring the dynamic impact of its underlying sources over the life cycle. In this study we add time dimension by introducing macrocohorts.

As in Italy poverty has a geographical specificity, looking at the national level this heterogeneity will be lost, that is the reason why we consider two different subsamples for the analysis. We will compare the

⁺ Corresponding author. Tel.: (+39) 0498274217 ; fax: (+39) 049 8274211.

E-mail address: martina.celidoni@unipd.it

²⁸ <http://www.oecd.org/dataoecd/45/42/41527936.pdf> [11/06/2009].

²⁹ In Italy, for over 80% of the surveyed, the unfairness is quite strong (OECD Report 2008 about income distribution and poverty).

North-central and Southern Italy³⁰ with the aim of understanding the different impact on inequality of education, occupation and demographic characteristics between 1997 and 2004 using data about households expenditure habits of the Italian National Institute of Statistics (ISTAT). In section two we will present the data used, in section three we will shortly describe the Shapley value decomposition and finally we will conclude with the empirical results.

2.Method

2.1.Data and cohorts specification

The empirical analysis has been developed on the Italian Household Budget Survey (ISTAT), we joined the information of eight independent cross-sections to obtain a cohort reconstruction of the data, grouping by the age of the household's head.

We split the original sample by territorial dislocation of households, distinguishing between those living in the North-central and Southern Italy. The subdivision was decided because there exists a strong heterogeneity between the northern and the southern Italian economic structures.

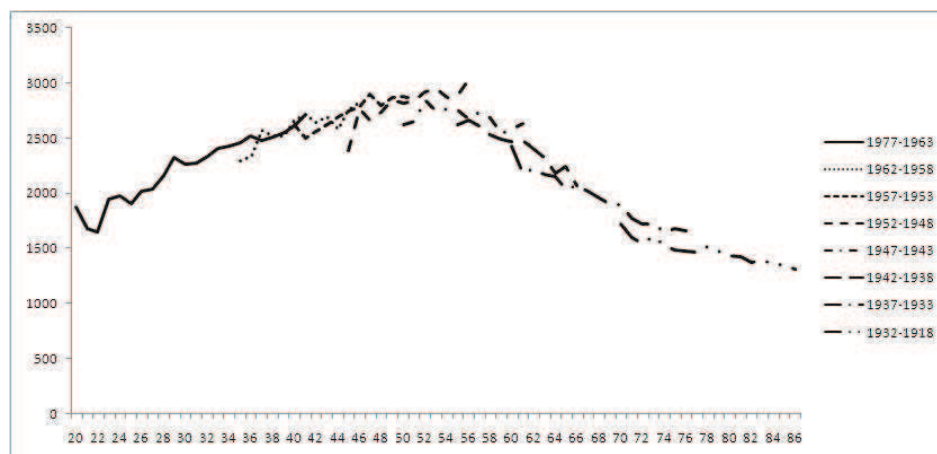


Fig.1a: North-central Italy. Real expenditure by age.

³⁰ We consider as North-central Italy the following regions: Piemonte, Valle d'Aosta, Lombardia, Liguria, Trentino Alto Adige, Veneto, Friuli Venezia Giulia, Emilia Romagna, Toscana, Umbria, Marche, Lazio. The South Italy subsample was defined as residual.

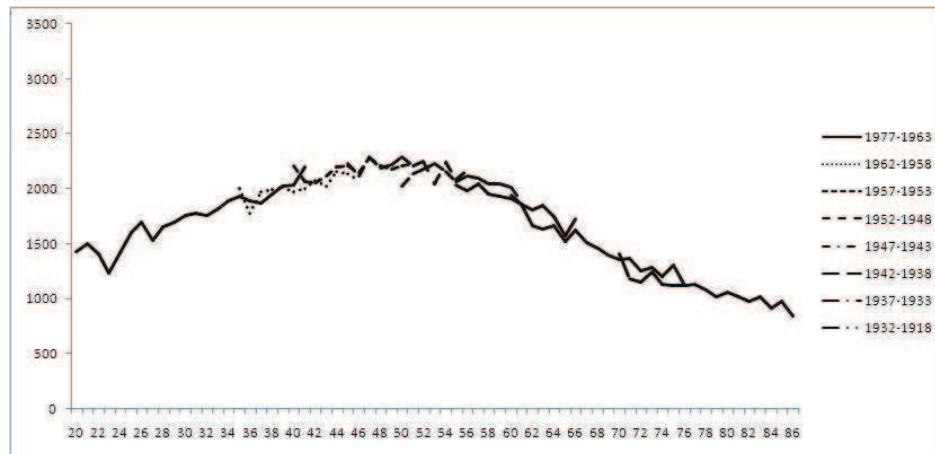


Fig.1b: South Italy. Real expenditure by age.

Figures 1a and 1b represent the dynamic of the real expenditure by age for the two subsamples. We can notice how the North-central expenditure level is higher than the southern and seems to increase more rapidly with the age, for each macrocohort.

The expenditure patterns reflect the different dynamics in the two Italian areas that can be explained as follows: while in the North and Centre the increase in expenditure can be driven by earnings dynamics reaching its maximum over the life cycle when workers usually reach the top of the career, in the South this is less evident, probably because of the structural characteristics of the labour market.

Moreover, as we can see in Figure 2, the Gini index³¹ shows a similar dynamic both in the North-central and Southern Italy over the life cycle, even if the South registers systematically a higher level of inequality for each macrocohort.

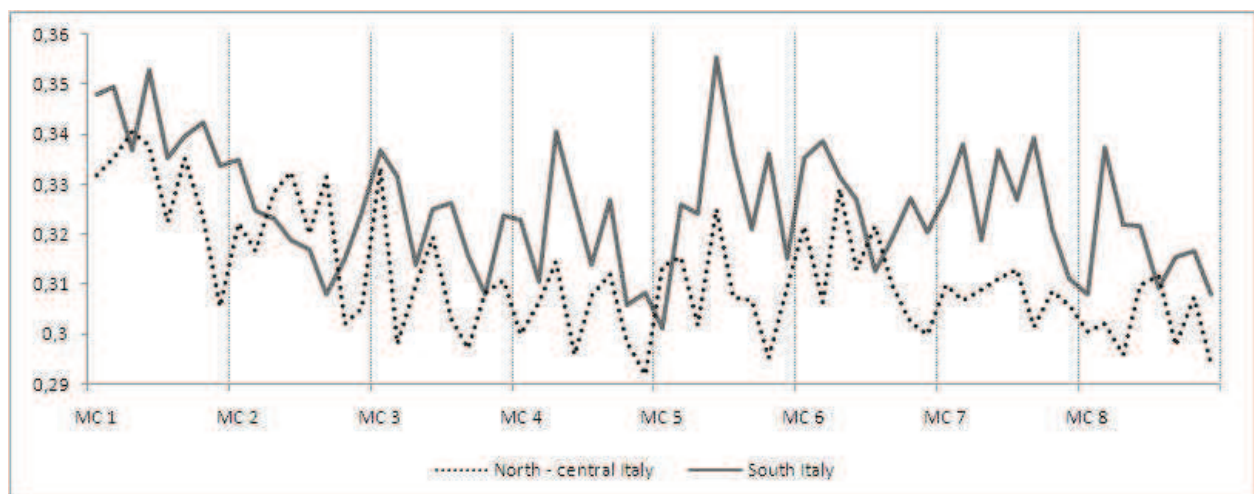


Fig.2: Gini index for macrocohorts.

³¹ The Gini index is computed on the real equivalent expenditure for each macrocohort over the period considered. We used the CPI index and the Carbonaro's scale to obtain the real equivalent expenditure.

However, before using these data as a genuine panel for our estimates, there are some theoretical issues related to the use of cohorts that must be discussed.

A cohort can be defined as a group with fixed membership formed of individuals, who can be identified as they show up in the surveys (Deaton, 1985). The most natural representation is to consider an age cohort formed by individuals (household's heads) born in the same period.

The first problem with cohorts concerns the fact that the unobserved cohort fixed effect is likely to be correlated with the explanatory variables of the model. This problem becomes even more relevant if the average cell size and/or the temporal period considered are small (Verbeek, 1992). However, a cell size of about a hundred individuals is proved to be sufficient to ignore the cohort nature of the data and to treat pseudo-panels as genuine panels³².

Another problem regards non-random attrition, i.e. the population from which the sample is drawn must be homogeneous over time. Following Jiménez-Martin et al. (1998), we found that attrition is not systematically related with the demographic structure of the households and should not be correlated with any unobservable characteristic that affects households expenditure behaviour.

Finally, the aggregation of households observations introduces a systematic heteroscedasticity (Gardes et al. 2005) due to differences in cell sizes across cohorts and over time. Following the approach commonly adopted in empirical studies to circumvent this problem, we weighted each observation by an heteroscedasticity factor that is proportional to the square root of the cell size.

2.2. The Shapley value decomposition

The Shapley value decomposition was first applied as a solution for estimating the power of any given voter in a coalition voting game, then Shorrocks (1982, 1999) proposed this regression-based approach to decompose any inequality index, that can be expressed as the sum of its contributory factors.

Let G for example be the Gini index that depends on n income sources, the Shapley decomposition considers the marginal effect on G of eliminating each of the income sources in sequence, and then assigns to each source the average of its marginal contributions, in all the possible elimination sequences.

The procedure proposed by Shorrocks (1999) treats symmetrically all factors and leads to an exact additive decomposition of the inequality index. Moreover the methodology has the advantage that groups of factors may be considered as a single entity without affecting their total contribution.

To clarify the procedure, let us consider $GTOT$ the value of the Gini index when all the n income sources are used, therefore $GTOT$ will represent the overall level of inequality in the population. The contribution of the x factor to the total inequality is computed removing it from the income generating model and measuring the extent by which the index has changed.

Let $G_{kk}(x)$ be the value of the Gini index when k factors have been dropped, so that only $n-k$ income sources explain the inequality, k denotes also the rank of elimination of x in one of the $n!$ possible dropping sequences. Thus $G_{11}(x)$ denotes the Gini index when only the x factor is dropped as first, while $G_{01}(x)$ corresponds to the Gini index computed with all the factors included. With the same logic, $G_{22}(x)$ will be the Gini index when two factors are dropped and the x factor is the second in the $(n-1)!$ possible elimination sequences. Finally $G_{nn}(x)$ will be the value of the Gini index when all variables have been dropped, where x was removed as last.

³² The cell size used is largely over one hundred for the northern subsample and around ninety for the southern. More detailed information is available upon request.

After having computed the index values, the x factor contribution $C_k(x)$, where k is the rank of elimination of x in the sequence, can be computed as average of the (n-1)! factor contributions.

$$C(x) = \frac{1}{n!} \sum_{k=1}^n C_k(x) \quad (1)$$

Finally the contributions of the x income source to the total inequality will be obtained as an average contribution of all the $C_k(x)$, according to (1), for details see Shorrocks (1999). The methodology also permits to calculate the proportion of total inequality that is not explained.

3. Empirical results: factors contributions

The factors contributions have been computed using a real equivalent consumption generating function that considers variables related to the household's head characteristics: education level, work position (unemployed, blue and white collar, worker, retired) and demographic characteristics (number of adults and number of children aged 0-14 in the household). We grouped the explanatory in three possible sources of inequality, education, demographics and occupation, to compute groups' contributions.

The geographical heterogeneity of the phenomenon can be noticed looking at the Figures 3.a and 3.b where the computed contributions are shown in the two subsamples considered: each shaded area corresponds to the contribution value of one group and their sum gives the total explained inequality. In both cases, North-central and Southern Italy, the selected variables can explain the inequality especially in the younger macrocohorts probably because for older macrocohorts what really determines the level of expenditure is the accumulation of wealth over the life. This decreasing power of the contributory factors justifies the use of macrocohorts to better understand the different impact of education, demographic characteristics and occupation to inequality over the lifecycle.

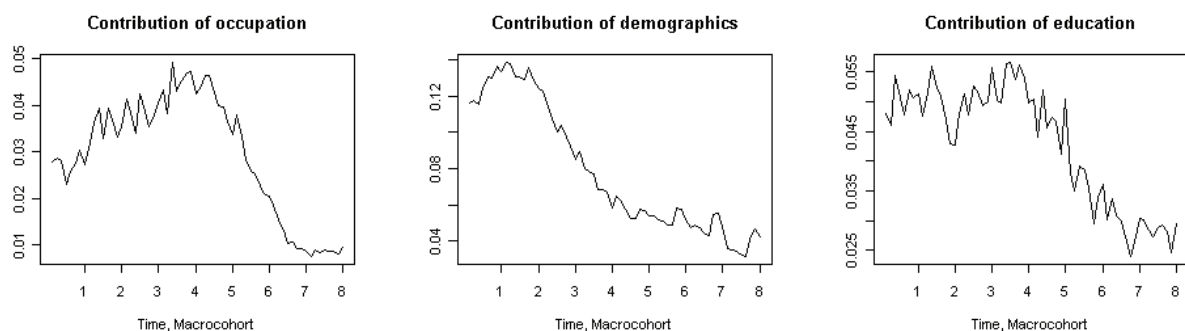


Fig. 3.a: North-central Italy. Gini index and factors contributions.

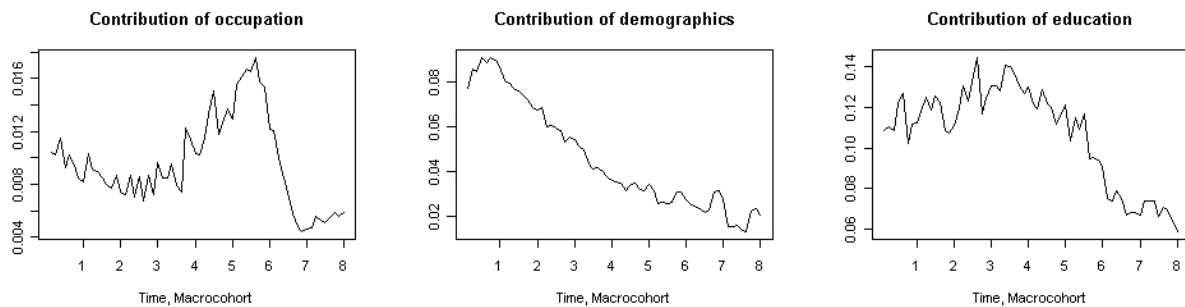


Fig 3.b: Southern Italy. Gini index and factors contributions.

We compare the two subsamples to highlight geographical differences in the explanation of inequality in Italy. In the South the highest contribution is given by education for every macrocohort: probably high levels of education give access to relatively better paid positions in the South compared to the North-Centre where high education is more diffused. In the North-central subsample the contribution of demographic characteristics, related to the composition of the household, seems to be the main determinant of inequality, especially for younger macrocohorts. This can be explained by the fact that in the North-centre the number of adults is more likely to be a proxy for the income sources in the household, since female labour participation is higher being a strong determinant in expenditure levels.

Even if potentially the decomposition could be helpful to identify the sources of inequality and therefore to design targeted policies, the total inequality is related to other factors that are not all included in the explanatories considered, leaving a residual term that needs more attention and further analysis.

4. References

- [1] O. Attanasio, J. Banks, C. Meghir and G. Weber. Humps and Bumps in Lifetime Consumption. *Journal of Business & Economic Statistics*, American Statistical Association. 1999, **17**(1): 22-35.
- [2] M. Biewen and S. Jenkins. A framework for the decomposition of poverty differences with an application to poverty differences between countries. *Empirical Economics*. 2005, **30**: 331-358.
- [3] A. Deaton. Panel data from time series of cross sections. *Journal of Econometrics*. 1985, **30**: 109-126.
- [4] F. Gardes, G. J. Duncan, P. Gaugert, M. Gurgand and C. Stazec. Panel and Pseudo-Panel Estimation of Cross-Sectional and Time Series Elasticities of Food Consumption: The Case of U.S. and Polish Data. *Journal of Business & Economic Statistics*. 2005, **23**: 242-253.
- [5] R. Gunatilaka and D. Chotikapanic. Inequality trends and determinants in Sri-Lanka: a Shapley Approach to decomposition. *Working paper 6/06, Department of Econometrics and Business Statistics, Monash University*. 2006.
- [6] Istat. La povertà assoluta in Italia nel 2007. 2009.
- [7] S. Jiménez-Martín, J.M. Labeaga and A. López. Participation, heterogeneity and dynamics in tobacco consumption: evidence from cohort data. *Health Economics*. 1998, **7**: 401-414.
- [8] L. Shapley. A value for n-person games. in H.W. Kuhn and A.W. Tucker (eds.). *Contributions to the theory of games* 2. Princeton University Press. 1953.
- [9] A. F. Shorrocks. Inequality decomposition by factor components. *Econometrica*. 1982, **52**: 1369-1385.



- [10] A. F. Shorrocks. Decomposition Procedures for distributional analysis: a unified framework based on the Shapley value. *Mimeo, University of Essex*. 1999.
- [11] M. Verbeek and T. Nijman. Can cohort data be treated as genuine panel data?. *Empirical Economics*. 1992, **17**: 9-23.