Camel or dromedary? A study of the equilibrium distribution of income in the EU countries.
Cammello o dromedario? Un'analisi della distribuzione di equilibrio del reddito nei paesi dell'UE.

Crosato L., Ferretti C., Ganugi P.

Abstract We face here the problem of analysing the presence of bimodality of the equilibrium distribution of incomes in the EU countries, using EU-SILC data about 2012-2015. As a first step we visually inspect the kernel distribution and calculate the Sarle's bimodality coefficient. We evaluate also the relationship between bimodality and inequality. As a second step we propose to use some suitable stochastic models to analyse the shape (camel/dromedary) of the estimated the long-run income distribution. The chosen models are the classical Markov Chain and the Mover Stayer model.

Abstract *Questo lavoro affronta il problema di valutare la presenza o meno della bimodalità nella distribuzione di equilibrio dei redditi dei Paesi europei, per mezzo dei dati EU-SILC relativi agli anni 2012-2015. Come primo passo, analizziamo le stime kernel delle distribuzioni dei redditi e calcoliamo il coefficiente di bimodalità di Sarle. Come secondo passo, proponiamo di utilizzare alcuni processi stocastici per analizzare la forma (cammello/dromedario) della distribuzione dei redditi stimata sul lungo periodo. I modelli scelti sono la Catena di Markov classica e il modello Mover-Stayer.*

Key words: long-run income distribution, equilibrium, bimodality, bipolarization, inequality.

Lisa Crosato, Dip. di Economia, Metodi quantitative e Strategie di impresa, Univ. degli Studi Milano-Bicocca; email: lisa.crosato@unimib.it

Camilla Ferretti, Dip. di Scienze Economiche e Sociali, Univ. Cattolica del Sacro Cuore di Piacenza; email: camilla.ferretti@unicatt.it

Piero Ganugi, Dip. Di Ingegneria e Architettura, Univ. degli Studi di Parma; email: piero.ganugi@unipr.it

1 Introduction: Why is Bimodality relevant?

Investigating the presence of bimodality in the equilibrium distribution of income (i.e. the long-run distribution such that frequencies among income classes have achieved stability w.r.o. time) in EU countries is relevant both for political and theoretical reasons.

The political relevance of the topic depends directly from the fact that the achievement of the Welfare Systems after the Second World War in different decades within the European Countries has always found the decisive support in the Middle-Income Population. Given the crucial role of this part of Population in the formation of consensus, its eventual structural reduction in favor of Low and High Incomes introduces the premises of a breakdown in the consensus toward the same System and, in turn, toward the political parties which defend it. The intricacy of the problem is empowered by the fact that an eventual structural bimodality is entirely a different fact from a return to the XIX century Income Distribution with respectively predominant and scanty masses in the Low Incomes and in the tail of the distribution. On the contrary, bimodality involves "polarization" i.e. not only large mass for the working poors but also, even if far more curbed, for the rich employees.

The theoretical reason to study the eventual structural (long period) character of bimodality sources by the fact that just today in some European countries Income Distribution seems to be featured by a noticeable second mode (cfr. Fig. 1). It is then natural to work to ascertain if the same second peak is the result of temporary and now seemingly fading recession and so fated to be reabsorbed by a new period of (equalizing) growth.

Literature on bimodality of Incomes can be divided in two distinctive branches depending on the nature of the data. One of these two branches on Incomes has used the pro capita GDP at World or European level and has aimed to ascertain the eventual polarization of pro capita GDP in the World (Bourguignon and Morrisson, 2002; Pittau, 2005). The second branch of literature on Incomes uses Personal Income micro-data and aims to ascertain bimodality within the single country. Within this second field of research, the problem of measuring bimodality, or bipolarization, has been deepened among others by Chakravarty et al. (2007), Chakravarty and D'Ambrosio (2010), Lasso de la Vega et al. (201), Deutsch et al. (2013).

Our paper contributes to the second strand of the literature, building on the empirical evidence on the bimodality of income distribution in a few European Countries detected using the EU-SILC database. Our goal is to verify whether Income distributions display a "camel" or a "dromedary" shape and if they preserve such shape with respect of time and in the long-run period.

The paper is structured as follows: section 2 illustrates the data source and the bimodality of income distribution, section 3 applies the kernel density estimation and Sarle's coefficient to evaluate the presence of bimodality and to have a look on its relationship with the Gini index, section 4 illustrates the expected results and section 5 concludes.

2 Data Description and evidence of bimodality in EU countries.

Our analysis is based on the data obtained from the European Union Statistics on Income and Living Conditions (EU-SILC) longitudinal study (see Krell et al., 2017, for a full description of the dataset and an analysis of the consistency of the data). EU-SILC has become the EU reference source for statistics on income distribution and social exclusion at European level and supplies, among other variables, the Net employee cash or near cash income (variable PY010N). According to the EU-SILC description, employee income is "the total remuneration, in cash or in kind, payable by an employer to an employee in return for work done by the latter during the income reference period". The net income component is then given by the gross income component but for the tax at source, the social insurance contributions, or both, which are deducted (see the methodological guidelines and description of EU-SILC target variables).

We start by analyzing the net employee income for the last available year, i.e. 2015 income extracted from the longitudinal component of the database, because further on we will need to build transition matrices for the calculation of the equilibrium distribution. Figure 1 reports kernel densities estimates for the countries that, in 2012 and 2015 have at least 1,000 non-NAs and non-null observations for the variable at hand.
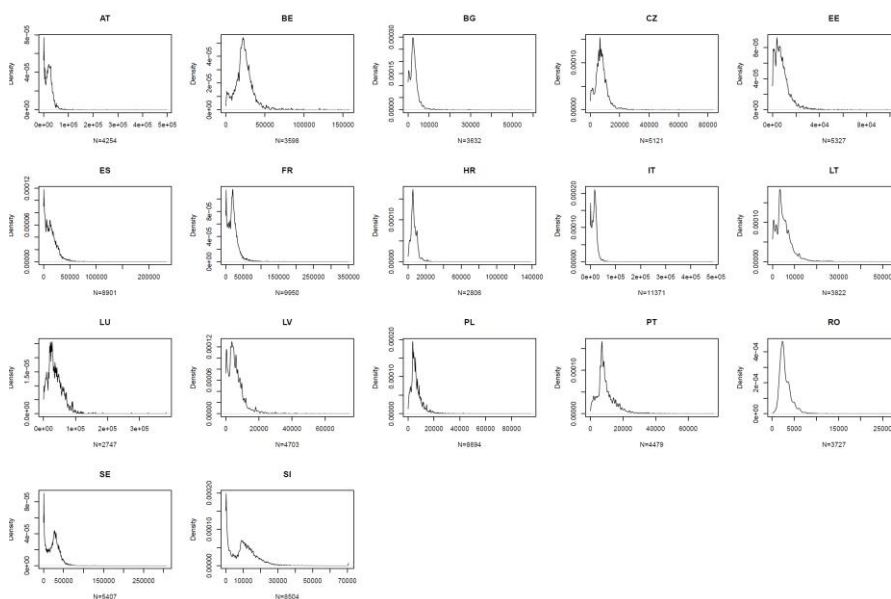

Figure 1: Kernel density estimates for 17 EU countries. Source: Author's elaboration on EU-SILC LONGITUDINAL UDB 2015 – version 1 of March 2017. Variable PY10N, Net employee cash or near cash income greater than zero, countries with at least 1,000 non-null values only.

As can be seen, bimodality can be clearly detected in several countries. The presence of bimodality can be also evaluated through the Sarle's bimodality coefficient (see

Ellison, 1987), whose corrected version for finite samples is given by the following formula:

$$b = \frac{g^2 + 1}{k + \dfrac{3(n-1)^2}{(n-2)(n-3)}}$$

where n is the sample dimension, g is the sample skewness and k is the sample excess kurtosis. In this case however the index neatly confirms bimodality only for Belgium, which is not consistent with the kernel densities of figure 1, suggesting the case for a further assessment of the extent to which bimodality takes place in the different countries.

The coefficient b seems positively related with the Gini index when considering the former communist countries (see figure 2), and, somewhat surprisingly, Belgium. The remaining countries instead show a not so clear relation.
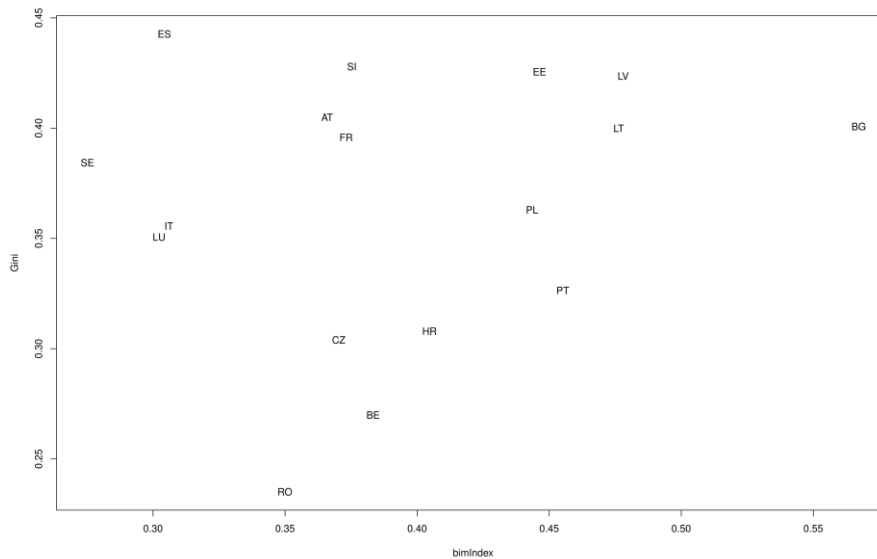

Figure 2: Scatterplot of the Gini concentration ratio (y-axis) versus the Sarle's bimodality index (x-axis) for 27 EU countries. Source: Author's elaboration on EU-SILC LONGITUDINAL UDB 2015 – version 1 of March 2017. Variable PY10N, Net employee cash or near cash income greater than zero, countries with at least 1,000 non-null values only.

3 Our proposal for evaluating the long-run two peaks distribution

Having ascertained the presence of bimodality in a relevant group of EU countries in the years 2012-2015, we tackle the problem to evaluate the long-run behavior of EU net incomes. We propose to estimate it using two stochastic models, Markov Chain

(MC) and Mover Stayer (MS) (see Anderson and Goodman, 1957, and Goodman, 1961). Both are based on the empirical yearly transition matrices, which summarize the probability to move among a set of suitably defined income classes. Such classes should be able to mirror the division among Low-, Middle- and High-Income individuals, as for example in Bourguignon (2002) or more recently in Xuehui and Shang-Jin (2017). The main difference is that the MS model supposes the existence of a bulk of individuals never moving from their starting state. In detail, let $S = diag\{s_i\}$ be the diagonal matrix where s_i denotes the probability that a EU citizen with a given income in 2012 is a Stayer and that consequently he/she will never leave from the corresponding income class. The MS global one-step transition matrix is given by the formula:

$$P = S + (I - S) * M,$$

where I is the identity matrix, and M is the transition matrix for the not-Stayers (Movers), which are supposed to move following a classical Markov Chain ruled exactly by M. Given the starting distribution p_0 (in this case coinciding with the percentage of citizens that in 2012 belongs to each income class), the long-run distribution is given by the equilibrium distribution π evaluated as:

$$\pi = p_0 * \left(\lim_{t \to +\infty} P^{(t)} \right),$$

where $P^{(t)}$ is the global t-steps transition matrix given by $P^{(t)} = S + (I - S) * M^{(t)}$, When $s_i = 0$ for every income class, the previous formula coincides with the classical MC equilibrium distribution. Parameters of both MC and MS can be estimated using the techniques proposed in Anderson and Goodman (1957) and Frydman et al. (1985).

4 Expected results

The long-run distribution gives a glance on what happen to the income distributions if the 2012-2015 economic conditions remain stable also in the future. We can expect only two possible results: 1) a camel or dromedary distribution maintains its starting shape, or 2) a camel/dromedary distribution tends respectively to lose/gain one "humpback". If also the equilibrium distribution remains or becomes a two-peaks one, we can claim that bipolarization has become chronic in some EEC countries, and the income is neatly divided between Low- and High-Income individuals.

Bipolarization and Gini indices have to be calculated also on the equilibrium distributions and their values have to be compared to the values obtained on the initial income distributions p_0. It is in fact intriguing to ascertain if bipolarization involves also a rise in inequality. With the aim to obtain a more robust analysis we will calculated different bipolarization indices as proposed in Chakravarty et al. (2007), Chakravarty and D'Ambrosio (2010), Lasso de la Vega et al. (2010) and Deutsch et al. (2013).

5 Further research

Further research will regard mainly two aspects:

1) the estimation of the long-run distribution through more complex models, as for example a mixture of Markov Chains in which individuals are characterized by different speeds, such as in Frydman et al. (2002);

2) the modelization of the bimodal density distribution through an analytical bimodal density distribution as proposed in Ferretti et al. (2017).

References

1. Anderson, T. and Goodman, L. (1957). Statistical Inference about Markov Chains. *Ann. Math. Statist.*, **28**(1), 89–110.
2. Bourguignon F, Morrisson C (2002). Inequality among World Citizens. *Am. Econ. Re*v., **92**(4):727–744.
3. Chakravarty, S. R., D'Ambrosio C. (2010). Polarization Orderings of Income Distributions, Review of Income and Wealth **56**(1): 47-64.
4. Chakravarty, S., Majumder A., Roy S. (2007). A Treatment of Absolute Indices of Polarization, *The Japanese Economic Review* **58**(2): 273-293.
5. Deutsch J, Fusco A., Silber J. (2013). The BIP Trilogy (Bipolarization, Inequality and Polarization): One Saga but Three Different Stories. *Economics: The Open-Access, Open-Assessment E-Journal*, **7**(2013-22): 1–33.
6. Ellison, A.M. (1987). Effect of seed dimorphism on the density-dependent dynamics of experimental population of *Atriplex triangularis* (Chenopodiaceae). *Am. J. Bot.*, **74**(8), 1280-1288.
7. Ferretti C., Ganugi P., Zammori F. (2017), Change of Variables theorem to fit Bimodal Distributions, in: *SIS 2017 Statistical and Data Sciences: new challenges, new generations*, Conference Proceeding, 417-422.
8. Frydman, H., & Kadam, A. (2002). Estimation in the continuous time Mover Stayer model with an application to bond rating migration. *App. Stoch. Mod. in Business and Industry*, **20**, 155–170.
9. Frydman, H., Kallberg, J., and Kao, D. (1985). Testing the adequacy of Markov Chain and Mover-Stayer Models as representation of Credit Behavior. *Oper. Res.*, **33**(6), 1203–1214.
10. Goodman, L. (1961). Statistical Methods for the Mover-Stayer Model. *J. Amer. Statist. Assoc.*, **56**, 841–868.
11. Krell, K., Frick, J. R., & Grabka, M. M. (2017). Measuring the Consistency of Cross-Sectional and Longitudinal Income Information in EU-SILC. *Rev. of Income and Wealth*, **63**(1), 30-52.
12. Lasso de la Vega, C., A. Urrutia and H. Diez (2010). Unit Consistency and Bipolarization of Income Distributions, Review of Income and Wealth **56**(1): 65-83.
13. Pittau M.G. (2005), Fitting Regional Income Distributions in the European Union, *Oxford Bull. Econ. Stat.,* **67**(2), 135-161.
14. Quah, D. T. (1996). Twin Peaks: Growth and Convergence in Models of Distribution Dynamics, Center for Economic Performance, Discussion paper No. 280. London School of Economics and Political Science.
15. Xuehui H., Shang-Jin W. (2017), Re-examining the middle-income trap hypothesis (MITH): What to reject and what to revive?, *J. Intern. Money and Finance*, **73A,** 41-61