Parametric families for the Lorenz curve: an analysis of income distribution in the European countries

Tommaso Lando[1,2], Michaela Staníčková[2], Jiří Franek[2]

Abstract

We investigate differences and similarities of income distribution among European countries on the basis of the Lorenz curve (LC). By performing an empirical analysis, we compare different parametric models for the LC, namely, the power LC, the Lamè LC and a generalized bi-parametric version of the Lamè LC. In particular, besides focusing on the goodness-of-fit, we concern about the effectiveness of multi-parametric models in identifying situations of intersecting LCs.

Key words

Lorenz ordering, income inequality, disparity, stochastic dominance, European countries

JEL Classification: C44, E24, O52

1. Introduction

A fundamental issue facing the global economy is the widening poverty gap between the developed and less developed world and the widening distribution of income within countries and geographical regions. Impoverishment has several causes, including low income as a result of unemployment, under-employment, or low wage employment. It may also be caused by a failure of the government to provide a welfare safety net in the event of the above. Income can be earned from selling labour, including wages, which are the largest source of income, and salaries and commission, which represent a very small fraction of income in comparison with more developed economies. Some income is unearned, such as rents from land ownership and interest from lending money. These sources of income are less available in developing economies. Equity means fairness or evenness, and achieving it is considered an economic objective. Despite the general recognition of the desirability of fairness, it is often regarded as too normative a concept because it is difficult to define and measure. For most economists, equity relates to how fairly income and opportunity are distributed between different groups in a given society (Melecký, 2018; Staníčková, 2017).

The European Union (EU) faces many challenges. On the global stage, the EU has to speak with one voice to counter a plethora of political, military and economic crises. Internally, it needs to foster cohesion in spite of the many events that threaten the EU at its core. In this context, do social issues matter at all? If we look at the EU evolution over the past decades, substantial progress has been made in terms of building an internal market and an economic and monetary union, albeit not without problems, as the 2008 crisis has shown. It looks actually as if the EU and its Member States were mostly thinking in economic terms, hoping that economic solutions will fix all social problems at once. To negate the importance of social

[1] Dr. Tommaso Lando, University of Bergamo, via dei Caniana 2, Bergamo, Italy, tommaso.lando@unibg.it
[2] Dr. Tommaso Lando, Ing. Michaela Staníčková, Ph.D., Ing. Jiří Franek, Ph.D.
VŠB - TU of Ostrava, Faculty of Economics, Department of Finance, Department of European Integration, Department of Business Administration, Sokolská třída 33, 702 00 Ostrava 1, Czech Republic; tommaso.lando@vsb.cz, michaela.stanickova@vsb.cz, jiri.franek@vsb.cz

issues is to undermine the EU foundations (Allmendinger and Driesch, 2014). Many politicians and economists believe that economic growth replaces or diminishes the need for social policies (MacGregor Pelikánová and Beneš, 2017). The strict regulation of corporate non- financial reporting is not the best for good understanding and practice of corporate social responsibility (Pakšiová, 2016). However, the EU growth over the last decades has been accompanied by an increase in inequalities in many countries. Inequalities threaten social cohesion and they threaten growth.

If such concerns are correct, it is essential not only to build institutional structures for European social union but also to map social inequalities in the EU. The low growth performance in the EU over the recent decades has increased concerns regarding an increasing economic dispersion, income inequality at large, and social exclusion. Recent research works have stimulated fierce debate on inequality among academics and policymakers. The recent economic crisis revealed many of the weaknesses of the current European economic policy, not least at the level of its fiscal policy, monetary policy, industrial policy, and social policy, and its inability to address problems related to inequality and the competition fairness (MacGregor Pelikánová, 2017). Inequalities in the EU have been the object of extensive research over the last decade. Several factors can explain this widespread interest; especially the revival of growth theory (Romer, 1990; Aghion and Howitt, 1998) was contemporaneous with a growing empirical literature on economic convergence (Sala-i-Martin, 2006; Quah, 1997; Barro and Sala-i-Martin, 1995).

Therefore, inequality can be quantified by looking at the distribution of income or wealth. The distribution of wealth is likely to be much greater than income because wealth is built up over many decades, and for some families, over centuries. The distribution of income is relatively easy to measure - valuing wealth is more difficult. This is because wealth is often hidden from view, and because it changes its value over time.

The Lorenz curve (LC) is a fundamental tool for the graphical representation of income inequality. The LC evaluated in $p \in [0,1]$ gives the proportion of the total wealth corresponding to the $p100\%$ poorer part of the population. Clearly, perfect equality is represented by a LC with equation $L(p) = p$. If the LC of a distribution is higher than the LC of another distribution, we can rank such distributions with respect to the *Lorenz dominance* (LD). The LD criterion basically reflects the fact that one distribution may exhibit a lesser degree of inequality compared to the other. However, many empirical studies revealed that Lorenz curves often intersect in the practice, that is, it is not rare to find couples of distributions that cannot be ranked based on the LD. In such cases, we can compare the intersecting distributions by relying on weaker orders of inequality, for more information see Lando and Bertoli-Barsotti (2016).

Many different functional forms for the LC have been proposed in the literature. Some of them belong to mono-parametric families, which may be denoted as *ordered* families of LCs, in that the LD is fully characterized by the value of a unique parameter. This property definitely represents an advantage for such kinds of families, although multi-parametric families generally yield better performances in terms of goodness-of-fit. Moreover, by construction, ordered families do not permit LCs to cross, and this lack of flexibility may be inappropriate in a number of cases.

In this paper we study some different families of parametric LCs, namely, the power Lorenz curve (PLC), the Lamè class of LCs (LLC) (Sarabia et al., 2017), which actually consists of two slightly different formulas, and a generalized version of the Lamè curve (GLLC), that has been studied by Sarabia et al. (1999). Whilst the PLC and the LLC are mono-parametric ordered families, the GLLC depends on two parameters, where different parametric combinations may yield intersecting LCs. Therefore, the main objectives of this paper are:

- study the suitability of these different families to model income distribution, in terms of goodness-of-fit. Special attention is given to the advantages/disadvantages of mono-parametric models with respect to multi-parametric models;
- study the usefulness of multi-parametric families to describe non-dominated situations (i.e., cases of intersecting LCs). In particular, a multi-parametric model should be able to identify most of the crossing pairs of LCs.

To this purpose, we perform an empirical analysis of the LCs of 32 European countries in 2015. The data have been downloaded from Eurostat's database.

2. Methods

We recall that a preorder is a binary relation $\leq$ over a set S that is reflexive and transitive. In particular, observe that a preorder $\leq$ does not generally satisfy the antisymmetric property (that is, $a \leq b$ and $b \leq a$ does not necessarily imply$a = b$) and it is generally not total (that is, each pair a, b in S is not necessarily related by $\leq$).

Let F be a non-negative distribution with positive and finite expectation μ_F. The (generalized) inverse or quantile function of F is given by

$$F^{-1}(p) = \inf\{z: F(z) \geq p\} p \in (0,1),. \quad (1)$$

The Lorenz curve $L_F: [0,1] \to [0,1]$ is defined as follows (Gastwirth, 1971):

$$L_F(p) = \frac{1}{\mu_F} \int_0^p F^{-1}(t) dt \, , p \in (0,1). \quad (2)$$

We recall that the Gini index is given by twice the area between the Lorenz curve and the 45° line:

$$\Gamma(F) = 1 - 2 \int_0^1 L_F(t) dt. \quad (3)$$

Actually, for a given percentage p, $L_F(p)$ represents the percentage of "total" possessed by the low $100p\%$ part of the distribution. It is well known that the higher of two non-intersecting Lorenz curves can be obtained from the lower one by a sequence of income transfers from "richer" to "poorer" individuals. This criterion has been called "Pigou-Dalton condition". For this reason, in an economic framework, the higher of two non-intersecting LCs should be preferred, in that it shows less inequality compared with the lower one. This idea defines the LD, defined as follows.

Definition 1. We write $F \leq_L G$ if and only if $L_F(p) \geq L_G(p), \forall \, p \in (0,1)$. (4)

When the LD is not fulfilled, i.e. when LCs intersect, some weaker criteria can be used in order to obtain unambiguous rankings. Muliere and Scarsini (1989) and Aaberge (2009) suggest cumulating LCs from the left or from right: that is, attaching more weighting to low or top incomes. Such ambiguous situations are very important, in that is has been empirically shown that intersecting LCs is an extremely frequent situation in the practice.

Among the different functional forms that have been proposed to approximate the LCs of income distributions, we propose the following ones.

1) *The power LC:* The PLC is defined by the following formula:

$$L_P(p, a) = p^a, \; p \in (0,1), \; a \geq 1. \quad (5)$$

This basic model is clearly ordered with respect to the parameter a, since $L_P(p, a_1) \geq L_P(p, a_2)$ iff $a_1 \leq a_2$. For $a = 1$ we obtain the equality line.

2) *The Lamè LC:* The LLC is defined by the following two different formulas:

$$L_{L1}(p,a) = [1-(1-p)^a]^{1/a},\ p \in (0,1),\ a \in (0,1] \tag{6}$$

$$L_{L2}(p,a) = 1-(1-p^a)^{1/a},\ p \in (0,1),\ a \geq 1. \tag{7}$$

In both cases, for $a = 1$ we obtain the equality line. The LLC has been introduced by Henle et al. (2008) and more recently studied by Sarabia et al. (2017). Both curves are ordered with respect to the parameter a, in particular:

$$L_{L1}(p,a_1) \leq L_{L1}(p,a_2) \text{ iff } a_1 \leq a_2, \tag{8}$$

$$L_{L2}(p,a_1) \leq L_{L2}(p,a_2) \text{ iff } a_1 \geq a_2. \tag{9}$$

3) *The generalized Lamè LC:* The GLLC is defined by the following formula:

$$L_{GL}(p,a,b) = [1-(1-p)^a]^b,\ p \in (0,1),\ a \in (0,1], b \geq 1. \tag{10}$$

For $a = b = 1$ we obtain the equality line. Differently from L_P, L_{L1} and L_{L2}, this family is not ordered, in that different combinations of a and b may yield intersecting LCs.

In the next section, we compare the performance of these different parametric models in terms of goodness-of-fit, and we also analyze the effectiveness of the GLLC in identifying non-dominated cases.

3. Empirical analysis

Data have been retrieved online from Eurostat's website (the data have been studied also in Lando et al., 2017). In particular, Eurostat provides the "distributions of income by quantiles" with two options, in terms of income and living conditions indicator, namely: i) *top cut-off point*, which represents the income of the individual at the right end of the given quantile and; ii) *share of national equalized income*, which is the share of the total income which belongs to a given interval. Eurostat provides i) and ii) for the three quartiles, the four quintiles, the nine deciles and the first (and last) five percentiles. Unfortunately, we note that some countries present negative incomes in the first 2-3 percentiles. The presence of negative incomes hampers the applicability of properties 1, 2 and 3, discussed in the previous section. Moreover, smaller percentile values are generally less reliable and accurate. Then, we decided to consider the LCs starting with $p = 0.05$. Indeed, by properly cumulating the shares of national equalized income we can obtain the values of the LC for

$$p = 0{,}0.05{,}0.1{,}0.2{,}0.25{,}0.3{,}0.4{,}0.5{,}0.6{,}0.7{,}0.75{,}0.8{,}0.9{,}0.94{,}0.95{,}0.96{,}0.97{,}0.98{,}0.99{,}1,$$

that is, a LC with 18 nodes (excluding 0 and 1).

The first step of our analysis consists in analyzing the empirical LCs, computed from the observed data. For the year 2015, we obtained the LCs of 32 countries and compared each pair of LCs (i.e. 32×31/2 pairs) based on the LD relation. We find that the LD can rank only the 56% of the pairs, while the remainder 44% of the pairs present intersecting LCs (i.e. with one or more crossings). Thus, LD seems not to be an effective criterion in comparing LCs of European countries.

Then, as a second step, we estimated the parametric models studied in section 2 by fitting them to the 32 observed LC discussed above. For the sake of simplicity, we used the least squares method, which consists in finding the parametric values that minimize the square of the "distance" between the observed LC and the model considered. Such distance is actually a sum of quadratic differences, evaluated in all the 18 nodes of the observed LCs. By dividing this distance by the number of nodes (18) we obtain the *mean squared error* (MSE), which can be used as a measure of goodness-of-fit. We computed the average MSE over the 32 LCs, for all the models considered, and obtained the following results.

Table 1: Average MSE of the four models considered

Model	Av. MSE
PLC	0.2400
LLC(1)	0.0007
LLC(2)	0.0006
GLLC	0.0002

Source: own elaboration, 2018

The results show that the LLC (in both the two versions) provide an excellent performance in approximating the LC, whilst the PLC is not well fitting at all. The GLLC provides a smaller MSE compared to the LLCs (obviously) but the improvement is questionable, especially if we are just interested in the goodness-of-fit because the LLCs perform extremely well besides having just one parameter.

Now, it is interesting to see if the models considered can identify dominated or non-dominated LCs. As we discussed in section 2, all the ordered families cannot generate crossing LCs, by definition. Then, in the third step of our analysis, we focus just on the GLLC, and compute the number of times when observed LCs cross and estimated LCs also, vs. the times when LCs do not cross but the estimated ones do, and so on. In particular, we find that the GLLC is able to identify the LD in the 82% of the cases (the percentage is computed over the number of cases when the LD holds). On the other hand, the percentage of cases when intersecting estimated LCs correspond to intersecting observed LCs (computed over the set of intersecting observed LCs) is 77%.

Based on the getting results, as proof, right measurement is a powerful instrument for socio-economic progress, which is why efforts are constantly being made to improve their power and precision; wrong or imprecise measurement a source of hazard and even havoc (Turečková, 2015). The essential purpose of economic activity is the promotion of human development, welfare and well-being in a sustainable manner, and not growth for growth's sake, yet we lack effective measures to monitor progress toward these objectives. Advances in understanding, theory and measurement must necessarily proceed hand in hand. Measuring multiple dimensions of socioeconomic progress is indispensable to understanding its components, benchmarking success, and catalyzing improvement. What level have we reached in comparison to others? Are we doing well? Are we going in the right direction? Are we catching-up or lagging behind? Are we meeting benchmarks or are we missing them? Are we using our fair and sustainable share of resources or too much? Is a group of economies converging or not? Just to list a few. At the same time, we are surrounded by an abundance of indicators trying to provide answers to these questions, at different levels of sophistication, in many cases serving as a basis for evidence-based policy decisions. Such indicators often seek to measure much aggregated but also diffuse concepts, rich in value judgements but not always grounded in hard science. The most prominent examples we see are indicators of economic development and performance. In recent years these have been complemented by alternative progress and well-being measurements. These indicators are frequently presented in dashboards and scoreboards, as well as aggregated or model-based composite indicators or indices (CIs). In recent years, international organizations, think-tanks, and the social sciences have contributed to a dramatic expansion in the range of CIs indices measuring concepts such as human development, governance, or social capital. Therefore, a large number of composite indexes of economic and social well-being have been developed. Unfortunately, the methodological issues associated with CI construction have often been neglected or inadequately treated by index developers.

This can be part of further research, i.e. the orientation on income distribution representing by indices.

4. Conclusion

The level of social inequalities belongs to important indicators influencing the socio-economic development and other processes taking place in the social and economic realm. Facilitating rational income distribution and reducing poverty are mentioned among the main goals of public policy. It should be mentioned that such multidimensional phenomena as income disparity and poverty might be analyzed from many different perspectives, including the national and international, also within the EU. Striving for fairness in economic development is crucial in order for societies to be stable and citizens not to feel disenchanted. The economic crisis has put inequalities high on the political agenda and made this an issue of serious public concern. There is an increasing recognition that social policy can reduce inequality and poverty while simultaneously improving the economic functioning of the country as reflected in the idea of inclusive growth in the EU's Europe 2020 strategy, with references to a high-employment economy delivering economic, social, and territorial cohesion in which benefits of growth and jobs are widely shared. Inequality is a key problem facing the EU, and it has significant impacts not only on human well-being but also on economic performance. This study has tried to show that inequalities in the EU are not a recent phenomenon and that they have in general increased over recent times in most of the EU countries. As a general principle, it is important to note that many differences among people in the EU are created by society and systematically linked to life chances. The only way for the EU to meet these challenges is to not only strengthen economic growth policies through broad-based economic programme promoting marketization but also by resolutely pushing for the expansion of social aspects of the EU model (Allmendinger and Driesch, 2014).

The future design of European economic policy must then provide a framework in which the policy instruments essential for a monetary, fiscal, industrial, sectorial, and social policy consistent with full employment and a reduction in inequality play a more prominent role. Europe 2020 is a credible strategy of industrial policy for the future of Europe and has the merits of presenting clear actions, clear targets and a detailed measurement strategy to monitor implementation. Combatting inequality should be considered as an instrumental target for both sustainable and inclusive growth. European policymakers have a long to-do list to foster inclusive growth in Europe (Darvas and Wolff, 2016). In all the countries of the EU, the welfare state has come under intense scrutiny as a result of budgetary pressures and wider societal developments. European social policy responses need national and regional contextualization. Simultaneously, the EU needs a sense of common purpose and a common policy framework in support of national social policies. Its aim should be to create a virtuous circle whereby both pan-European cohesion and national cohesion are enhanced. Cohesion is about income and employment, but also about other dimensions of well-being.

The recent interest in inequality is thus simply the recognition of the centrality of the topic to economic theory, policy, and performance. The recent return of the topic of inequality has been triggered by important contributions to the empirical analysis of inequality (Galbraith, 2009), but these empirical analyses must be combined with an economic theory that is adequate to address the macroeconomic and microeconomic effects of inequality on social welfare. These problems are not always well diagnosed because the empirical measurement of inequality is often unable to take into account the geographical dimension of inequality, which is particularly complex in Europe. In this preliminary analysis, we were able to see some of the advantages/disadvantages of ordered families of LCs compared to bi-parametric families. Ordered families are easy to interpret and may provide an extremely accurate approximation: this is the case, for instance, of the Lamè class, which seems to be able to capture the shape of an LC, at least with regard to the dataset considered. Moreover, the unique parameter contains most of the information about the characteristics and the concentration patterns of the population, and can, therefore, be used as an index of inequality, just like the Gini index. In fact, the Gini index has a one-to-one correspondence with the parameter, in the case of mono-parametric families. On the other hand, the bi-parametric model that we focused on generalizes the Lamè class introducing an additional parameter. As an obvious consequence, the fit is enhanced, although this may be not sufficient to justify the use of a bi-parametric model. In this regard, the main advantage of a generalized model is the possibility of generating scenarios of crossing LCs, which happen to be quite frequent in practical situations. Indeed, our analysis shows that LCs of European countries cross almost half of the times so that it could be inappropriate to describe a partially-ordered set of LCs with a totally-ordered family. Differently, the GLLC address this issue by supporting most of the cases when LC do (or do not) cross. Future work will be aimed at extending this analysis to a larger number of parametric models for the LC.

Acknowledgements

The research was supported through the Czech Science Foundation (GACR) under project 17-23411Y and by VSB-TU Ostrava under the SGS Project SP2018/34.

References

[1] Aaberge, R. (2009). Ranking intersecting Lorenz curves. *Soc. Choice Welf* , 33, pp. 235-259.

[2] Aghion, P. and Howitt, P. (1998). *Endogenous growth theory*. Cambridge: MIT Press.

[3] Allmendinger, J. and Driesch, E. (2014). Social Inequalities in Europe: Facing the challenge. *Discussion Paper P 2014-005*. Berlin: Social Science Center.

[4] Barro, R. and Sala-i-Martin, X. (1995). *Economic growth*. New-York: McGraw Hill.

[5] Galbraith, J.K. (2009). Inequality, unemployment and growth: new measures for old controversies. *Journal of Economic Inequality*, 7(2), pp 189-206.

[6] Gastwirth, J.L. (1971). A general definition of the Lorenz curve. *Econometrica*, 39, pp. 1037-1039.

[7] Henle, J. M., Horton, N. J. and Jakus, S. J. (2008). *Modelling inequality with a single parameter. In Modeling income distributions and Lorenz curves* (pp. 255-269). New York: Springer.

[8] Lando, T., Bertoli-Barsotti, L. and Staníčková, M. (2017). Weak Orders for Intersecting Lorenz Curves: Analysis of income distribution in European countries. *Proceedings of the 35th International Conference on Mathematical Methods in Economics.* Ostrava: VŠB-TUO, pp. 402-407.

[9] Lando, T. and Bertoli-Barsotti, L. (2016). Weak orderings for intersecting Lorenz curves. *Metron*, 74(2), pp. 177-192.

[10] MacGregor Pelikánová, R. (2017). European Myriad of Approaches to Parasitic Commercial Practices. *Oeconomia Copernicana*, 8(2), pp. 167-180.

[11] M. (2017). Does the Full Harmonization of the Consumers´Protection Against Unfair Commercial Practices Via UCPD Fit in Europe 2020? *Czech Yearbook of International Law - CYIL,* 8, pp. 223-231.

[12] Melecký, L. (2018). The main achievements of the EU structural funds 2007–2013 in the EU member states: efficiency analysis of transport sector. *Equilibrium. Quarterly Journal of Economics and Economic Policy,* 13(2), pp. 285-306.

[13] Muliere P. and Scarsini M. (1989). A note on stochastic dominance and inequality measures. *Journal of Economic Theory*, 49(2), pp. 314-323.

[14] Pakšiová, R. (2016). Understanding of corporate social responsibility in large companies in Slovakia within the context of a sustainable development. In *Economic policy in the European Union member countries: proceedings of 14th international scientific conference*. Karviná: Silesian University in Opava, pp. 516-525.

[15] Quah, D. (1997). Empirics for growth and distribution stratification, polarization and convergence clubs. *Journal of Economic Growth*, 2(1), pp 27-59.

[16] Romer, P. (1990). Endogenous technological change. *Journal of Political Economy*, 98(5), pp. S71-S102.

[17] Sala-i-Martin, X. (2006). The world distribution of income: Falling poverty and… Convergence, Period. *The Quarterly Journal of Economics*, 121(2), pp- 351-397.

[18] Sarabia, J. M., Castillo, E. and Slottje, D. J. (1999). An ordered family of Lorenz curves. *Journal of Econometrics*, 91(1), pp. 43-60.

[19] Sarabia, J. M., Jordá, V. and Trueba, C. (2017). The Lamé class of Lorenz curves. *Communications in Statistics-Theory and Methods*, 46(11), pp. 5311-5326.

[20] Staníčková, M. (2017). Can the implementation of the Europe 2020 Strategy goals be efficient? The challenge for achieving social equality in the European Union. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 12(3), pp. 383-398.

[21] Turečková, K. (2015). Income Inequality by Method of Non-weighted Average Absolute Deviation: Case Study of Central and Eastern European Countries. *Equilibrium*. *Quarterly Journal of Economics and Economic Policy,* 10(4), pp. 99-110.