LIS
Working Paper Series

No. 852

Isograph and LaSiPiKa Distribution: The Comparative Morphology of Income Inequalities and Intelligible Parameters of 53 LIS Countries 1967-2020

Louis Chauvel

February 2023

Luxembourg Income Study (LIS), asbl

Isograph and LaSiPiKa Distribution: The Comparative Morphology of Income Inequalities and Intelligible Parameters of 53 LIS Countries 1967-2020[1]

Louis Chauvel louis.chauvel@uni.lu

JEL D31, C16, C46 keywords: inequality, poverty, distributions, polarization, measurement

Contents

Abstract 162 words

The isograph methodology is developed here with associated distributions, indicators of inequality, additional results, and is implemented on 53 LIS countries (with an annex covering 655 LIS country-year samples). The gb2 and other classical distributions (FC, Dagum, Singh-Maddala) are presented along with new proposals, including gb2 subfamilies with p=1/q and p=q, the LaSi distribution to fit the quasi-linear isograph cases of level lambda and slope sigma, and finally the LaSiPiKa that completes LaSi with a polarization term of intensity pi and location kappa. This latest proposal fits better the cases of distributions with sharp flexible profiles in the isograph and provides independent intelligible parameters. The analysis is systematized to 655 samples to show the invariant patterns and significant changes. More complicated distributional shapes can be fitted with hand-tailored additional terms. Working with isograph and LaSiPiKa distributions is a way to diversify and deepen inequality analyses with a larger conceptualization of "morphology of inequality," not reduced to a Gini (or the like) measure.

[1] This paper received the financial support of the Luxembourg FNR Fonds National pour la Recherche project PEARL-IRSEI. I would like to extend my gratitude to the entire LIS team, especially Teresa Munzi, Heba Omar, Piotr Paradowski, and Jörg Neugschwender, etc. for their valuable support in data management and other assistance.

Abstract 500 words

This paper presents new empirical and methodological developments on a recent proposal, the isograph method (Chauvel 2016), and its capacity to improve our comparative knowledge of equivalized income distributions. The isograph is ISO(X)=Y/X, where Y is the logged medianized income, and X the logit of the fractional percentile p (aka logitrank) of the associated c.d.f. When ISO(X) is higher, “local” inequality at percentile p=invlogit(X) is deeper, meaning this percentile income is farther from the median. The isograph can detect significant local singularities in distributions and confirms that income shapes differ between countries. With a systematic analysis of 53 LIS countries (fig3) and 655 country-year LIS samples, a new LaSiPiKa five parameters distribution with four main distribution parameters are identified, on top of a scale parameter:

*lambda (the overall intensity of inequality) is the general level of ISO;
*sigma is the isograph slope, which means richer people are disproportionally richer relative to the median, and the poor are relatively well off.

Lambda and sigma plus a scale parameter b define the LaSi distribution (three parameters), a very parsimonious distribution that can be completed by a hump shaped term PiKa with two additional parameters:

*pi is polarization intensity
*and kappa is the epicenter of the hump (the level of X where polarization is centered).

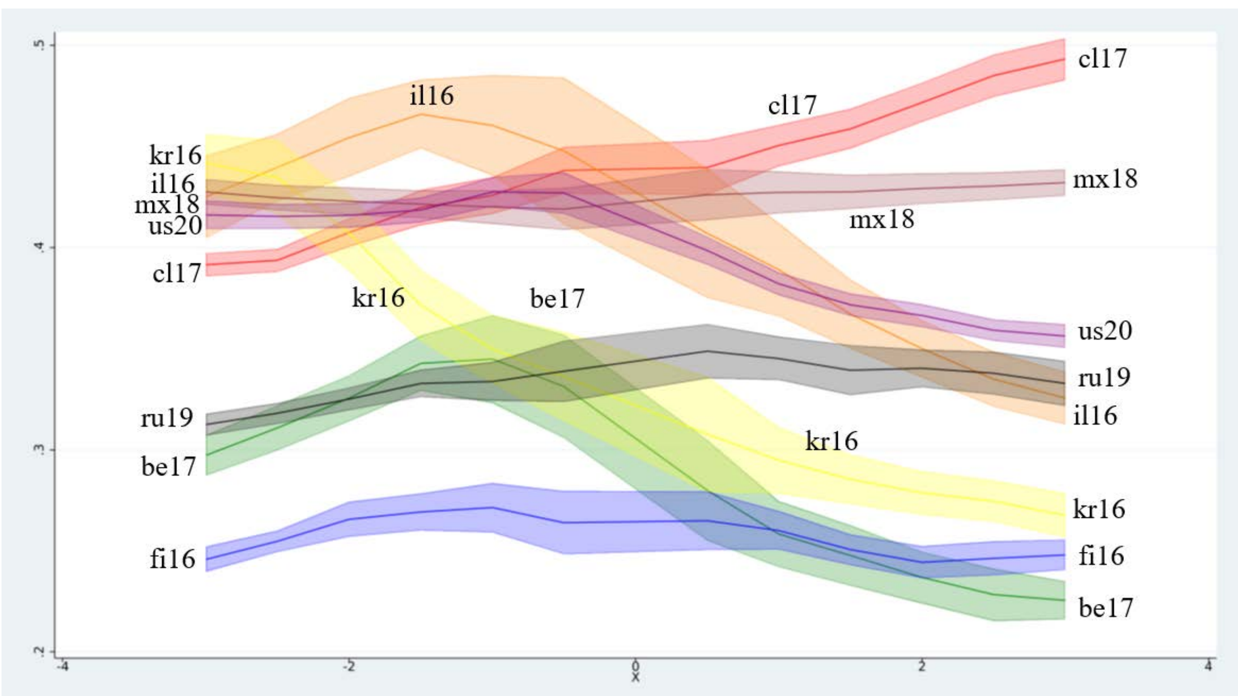
LaSiPiKa, a five parameter distribution, has been systematically estimated for the samples available in LIS and shows the diversity of distributional morphologies, and their change across time, like the shift of the U.S. shape from the early 1970s’ to 2020 (fig17), where increasing inequality is progressively stronger at the top(fig18). Compared to the gb2, which provides not intelligible parameters, LaSiPiKa has parameters with clear interpretation in the isograph and in terms of inequality shapes, with equivalences in terms of the Gini index of relative poverty rates (50% of the median) and other common indicators. Moreover, gb2 appears insufficiently flexible and misses the target distribution in 22 cases over 53, typically when the isograph is not linear. The LaSiPiKa is challenged in only 3 cases, typically when two humps are observed on the isograph. More complicated distributional morphologies can be fitted with hand-tailored additional terms with two polarities.

The analysis is systematized to 655 samples to show the invariant patterns and significant changes by countries. Traditional indicators like the Gini index and poverty rate have immediate meanings in the isograph: the Gini index is the value of the isograph near X=1, and the poverty rate (50% of the median) is the invlogit of X at the intersection between the isograph and the curve Y=ln(.5)/X. Shapes of the isograph detect humps, interpretable in terms of polarization. Comparisons of isograph across time from distribution A to B provide profiles of implicit redistribution from one to the other. Working with isograph and LaSiPiKa distributions is a way to diversify and deepen inequality analyses with a larger conceptualization of “Morphologies of inequality,” not reduced to a Gini (or the like) measure and provides tools to simulate proxies of empirical distributions.

The isograph in a nutshell

The isograph is defined in Chauvel (2016) as a characteristic function representing the shape of income distribution. It is a way to continuously measure inequality across the distribution from the bottom to the top. Contrary to kernel analysis (Jann, 2005, 2007) that provides non-parametric solutions, the aim of isograph is to measure inequalities and to support estimation of parametric distributions like gb2 or LaSiPiKa (see below “Defining LaSiPiKa”). The X axis of the isograph is the logitrank of the distribution, where X=0 stands for the median, X=+1 is close to the top quartile threshold, X=+2 nears the top decile, and X=+3 the top 5%. The isograph represents the ratio ISO=Y/X where Y is the logged medianized equivalized income. The isograph is generally close to a flat line, and when it is precisely a constant, this constant gini is equal to the Gini index of the distribution y, and the distribution y is a two parameters log-logit (Fisk) distribution y=b.exp(gini.X), the simplest gb2(a,b,p,q) where p=q=1. In empirical cases, the isograph of a country-year sample is bent different ways and expresses the “morphology of inequality”.

Figure 0a: isographs of eight empirical samples with confidence intervals CI 95%: be17 (green) cl17 (red) fi16 (blue) il16 (orange) kr16 (yellow) mx18 (brown) ru19 (grey) us20 (purple)



On figure 0, ten samples are presented to show typical diversities in isographs: be17 cl17 fi16 il16 kr16 mx18 ru19 us20, for respectively Belgium, Chile, Israel, Korea, Mexico, Russia, United States. The last two digits stand for the sample year. How can one read the eight isographs? Finland and Mexico are close to flat lines of constant ISO=0.26 and 0.43 respectively. Russia is also close to a flat isograph of constant 0.33. Those constant values are close to the respective Gini indices.

The five other cases are far from a flat line, but two are close to a linear curve: Chile and Korea. A positive sloped isograph means inequality at the top is stronger than at the bottom of the distribution: the rich are relatively far above the median and the poor are relatively closer to the median. Compared to Mexico which is at the same level of overall isograph, Chilean inequalities are stronger at the top (d9/med=2.85 in Chile against 2.56 in Mexico) and lower at the bottom (m/d1=2.40 in Chile and 2.55 in Mexico). The isograph reflects relative inequality across the income scale from the bottom to the top. On the contrary, the negative slopped isograph of Korea means Korean inequality at the top is almost as limited as in Finland: there, the rich have limited resources. Conversely, poor Koreans are far below the median, with a med/d10 ratio similar to the ones of Latin American countries.

The three last cases show specific nonlinearities: Israel, the U.S., and Belgium. Their isographs present a hump, generally close below the median, which means a relative polarization with relatively lower levels of inequality at the extremes, and more inequality near the hump. In Israel, the richer percentiles are not very rich, the poorer ones are comparable to the Korean, American, and Mexican equivalents, but extreme inequality is observed near to X=-1 (near the first quartile threshold), which means a polarization of incomes at this level. The isograph expresses, for instance, that the same country may have the relative poverty levels of neo-liberal economies, and richer incomes similar to countries with extensive, universal equalitarian welfare regimes, or other combinations of interest.

The isograph is related to the Gini index. When the isograph is a constant, this constant is the Gini index. In the general case, the Gini coefficient is close to the ISO(X) value between X=1 and X=2. Therefore, the Gini index over-represents inequality for the richer quartile. Conversely, the relative poverty rate is proportional to the ISO on the left of X=0. The main indicators of inequality have equivalents on the isograph. We will show the value of ISO(0) at the center of the isograph is close to .22ln(d9/d1) (r2=.99). The linear slope of the isograph is (.08)ln(d9.d1/(med)^2) (r2=.88)

The isograph provides an opportunity to test the quality of distributional adjustments in order to detect significant gaps or discrepancies between observed distributions and predicted values. For instance, this contribution challenges the gb2 proposal: when ISO(X) is linear, the gb2 approximation is generally acceptable, but in case of a humped isograph, with significant non-linearity, the gb2 approximation (4 parameters including the scale) tends to systematically miss the target. On figure 0b, gb2 makes a correct approximation for mx, cl, ru, kr, and fi, that have almost linear isographs. On the contrary, samples for il, us, and be largely miss the humped pattern of their non-linear isographs: gb2 significantly underestimates the non-linear pattern in the isograph.

The isograph is used to confirm a theoretical distribution (b2, sm, dagum, etc.) correctly fits an empirical distribution. The systematic study of LIS distributions leads to the elaboration of a LaSiPiKa distribution adjustment (5 parameters including the scale, see chapter “Defining LaSiPiKa”), based on a parameter lambda, overall level of inequality, sigma, a slope in the isograph, and two parameters that model a hump of intensity pi and of epicenter X=kappa where the hump is maximum (as exemplified by figure 11a). When no hump is detected, pi is close to zero, and LaSi distribution is a correct parsimonious approximation. Figure 0c shows the LaSiPiKa estimates outperform the gb2 with a better fit for the countries il us be, with improved curvatures.

Figure 0b: isographs of gb2 predicted distribution and eight empirical samples with CI 95% intervals be17 (green) cl17 (red) fi16 (blue) il16 (orange) kr16 (yellow) mx18 (brown) ru19 (grey) us20 (purple)

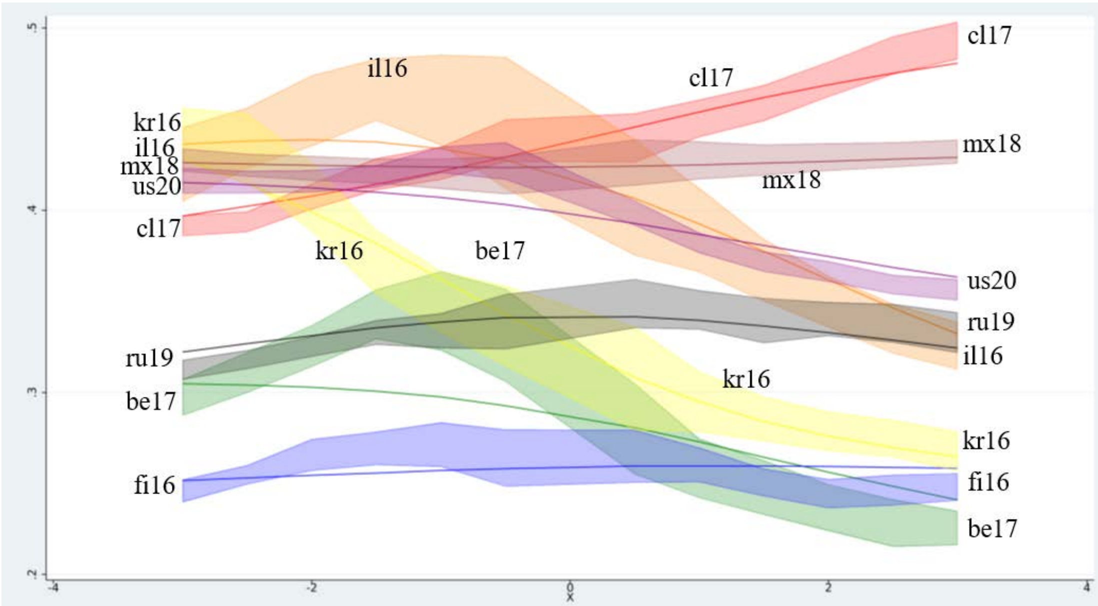


Figure 0c: isographs of LaSiPiKa predicted distributions and eight empirical samples with CI 95% intervals be17 (green) cl17 (red) fi16 (blue) il16 (orange) kr16 (yellow) mx18 (brown) ru19 (grey) us20 (purple)

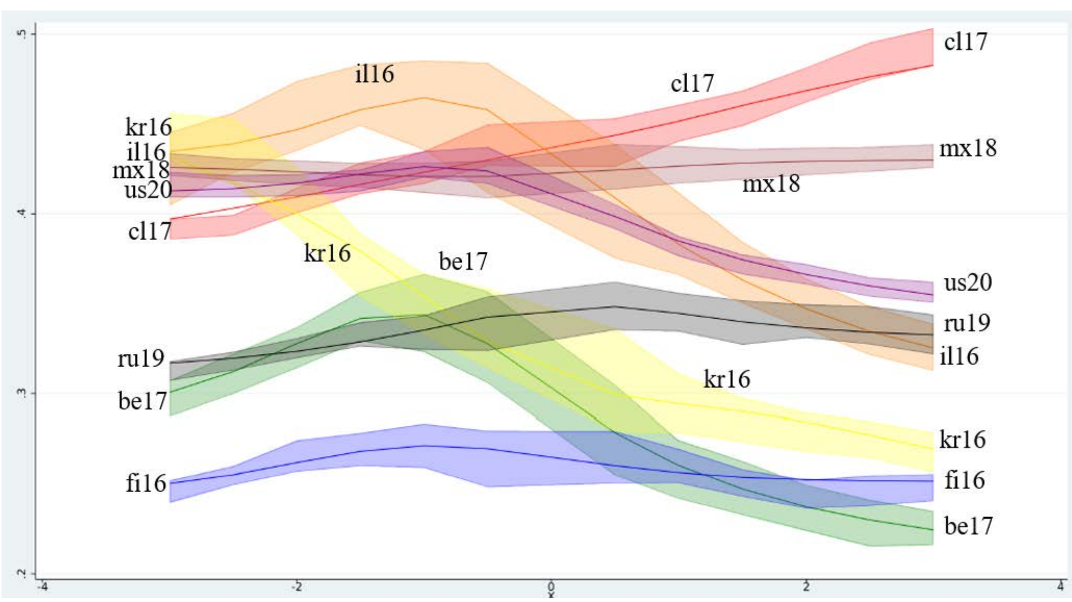


Table 0. Main parameters of the eight distributions, LaSiPiKa parameters, Gini index, relative poverty rate 50%, D9/D5 decile threshold ratio, D5/D1, ln(D9/D1) and ln(D9.D1/(D5.D5))

ccyyyy	la	si	pi	ka	gini	pora	d9/med	med/d1	ln(d9/d1)	Ln((d9*d1)/(med*med))
be2017	0.253	-0.011	0.079	-0.970	0.259	0.102	1.680	2.012	1.218	-0.180
cl2017	0.437	0.016	0.000	-1.545	0.457	0.159	2.853	2.403	1.925	0.172
fi2016	0.249	0.001	0.020	-0.832	0.260	0.059	1.718	1.771	1.112	-0.030
il2016	0.365	-0.020	0.081	-0.796	0.342	0.184	2.089	2.697	1.729	-0.256
kr2016	0.355	-0.033	-0.055	-0.077	0.300	0.143	1.816	2.500	1.513	-0.320
mx2018	0.427	-0.001	-0.004	-0.126	0.428	0.162	2.559	2.548	1.875	0.004
ru2019	0.328	0.005	0.021	-0.390	0.331	0.107	2.128	2.045	1.471	0.040
us2020	0.380	-0.011	0.039	-0.542	0.373	0.160	2.224	2.503	1.717	-0.118

The isograph and its LaSiPiKa (see below "Defining LaSiPiKa") associated parametric distribution clarify the systematic analysis of income distributions, and help to answer a series of general research questions:

Q0: Is the isograph a correct tool for the measurement of inequality? (Yes, except with the Pigou-Dalton principle, since for isograph the monetary gift of a billionaire a to a millionaire decreases inequality at level a and increases it at level b, the millionaire getting richer compared to the median, see Chauvel 2016 on axiomatic properties of the isograph)

Q1: are the shapes of inequality of empirical distributions all the same? (A1=No, and the isograph provides information on where –in terms of percentile level– the distributions show significant variations of inequality)

Q2: is the isograph associated with traditional indicators of inequality or distributional characteristics? (A2=yes for Gini, relative poverty, income ratios, etc.)

Q3: is gb2 sufficient to model all the existing empirical distributions? (A3=No)

Q4: is gb2 a parsimonious approximation to all the existing empirical distributions? (A4= no, LaSi distribution is similarly efficient with only three parameters against four in gb2)

Q5: is LaSiPiKa distribution sufficient to model all the existing empirical distributions? (A5=No, but it outperforms gb2, and hand-designed LaSi(PiKa)$_n$ with n humps provides appropriate solutions)

Isograph and LaSiPiKa are innovative tools providing novel, consistent measurements of inequality useful to compare distributions across time and nations.

The isograph in details

The isograph is defined in Chauvel (2016) as a characteristic function representing the shape of income distribution. The higher an isograph is at a given percentile level, the stronger is inequality, measured by a transformation ratio of income at threshold p and the median. The isograph is a way to have a continuous measure of inequality across the distribution. Consider an income distribution y>0 of density f(y), with cumulative distribution function (CDF) p=F(y) where p, percentile or "the fractional rank" (Van Kerm 2020), is the proportion p of individuals with income below y. The ISO function is defined as ISO=Y/X:

* Y= logmed(y) is the "logmed" transformation of income y, i.e. the logged value of the medianized income. As a general notation, if z is an income distribution of median med(z), y=z/med(z) is the medianized income; if y_m=1 is the median of y, logmed(y_m)=0. So on the median we have Y=0.

* X is the "logitrank" (Copas 1999) of the distribution: X=logit(p)=ln(p/(1-p)) where p is the fractional rank. On the median p_m=0.5, X_0=logit(p_m)=0. X of the median is 0; X=1 is close to the top quartile; X=2 is not far from the top decile; X=3 is near the top vintile (=95%); X=4 approximates the top 2%; X=4.5 the top 1%; X=5 the top 0.5%; X=6 the top 0.25%; X=7 the top 0.1%; etc. (tab 1)

The isograph presents ISO on the vertical axis and X on the horizontal.

Table 1. Conversion tables of p to X and reverse

P	X
0.001	-6.9
0.002	-6.2
0.005	-5.3
0.01	-4.6
0.02	-3.9
0.05	-2.9
0.1	-2.2
0.25	-1.1
0.5	0.0
0.75	1.1
0.9	2.2
0.95	2.9
0.98	3.9
0.99	4.6
0.995	5.3
0.998	6.2
0.999	6.9

X	P
-8	0.0003
-7	0.0009
-6	0.0025
-5	0.0067
-4	0.0180
-3	0.0474
-2	0.1192
-1	0.2689
0	0.5000
1	0.7311
2	0.8808
3	0.9526
4	0.9820
5	0.9933
6	0.9975
7	0.9991
8	0.9997

* The ISO function is defined as ISO(X)=Y/X. This provides a continuous measure of inequality across the distribution from p=0 to p=1 (or $X=-\infty$ to $X=+\infty$). The graph (Y,X) is a convenient transformation of the CDF of y, and the isograph is the slope from the median of any point of (Y,X). The higher the ISO at level X=logit(p), the stronger the ratio of income at percentile p to the median. ISO gives a characteristic profile of inequality across the distribution, from the poorer to the richer percentiles p. The isograph has a set of interesting characteristics.

An apparent issue is the potential discontinuity of the isograph (and estimation instability) near to the median at X=0, since ISO=Y/X. In any case, a sound conjecture by Philippe Van Kerm (see Chauvel et al. 2019, note 3) is to suppose the continuity of density at the median because, considering the L'Hôpital's rule, the limit in X=0 of ISO=Y/X is Y'(0), that is also the density function evaluated at the median. A simple way to approximate ISO at X=0 is to replace this value by the average of ISO(X=-0.5) and ISO(X=+0.5).

A major role of isograph in the methodological landscape of distribution comparisons is to compare empirical distributions with estimated CDF obtained through parametric distributions. Its logic is rather different to kernel strategies of density fit as it is not based on (small number of) parameters' estimation. This is why the main purpose of the isograph is to test the accuracy of well-known distributions like FC, other gb2 type distributions, or any other parametric distributions.

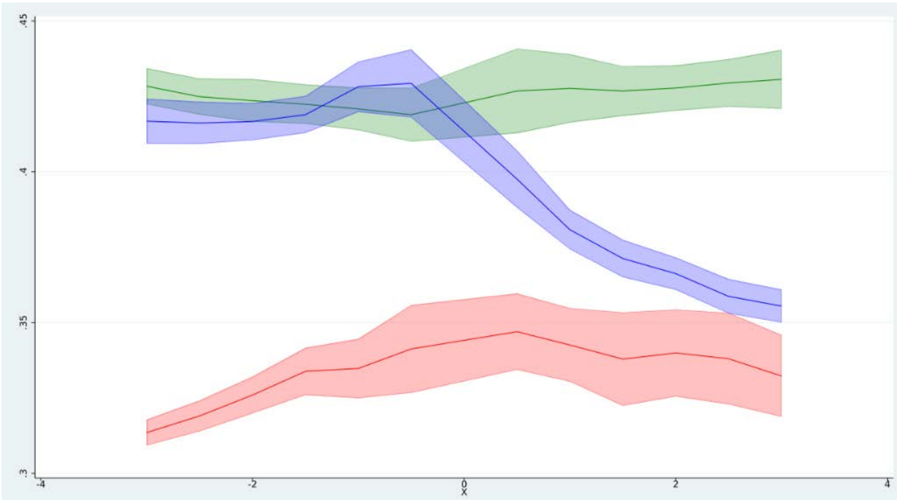
Typical comparison of empirical isographs: Mexico, Russia, and the U.S.

The empirical data used here are derived from the current Luxembourg Income Study's (LIS) 655 country-years collection (Fall 2022) with a focus on the most recent year in each of 53 countries. Strictly positive disposable equivalized, medianized income y is the data concept presented here. A first implementation of the isograph (fig 1) is demonstrated on the 2020 U.S. data compared to the Mexico 2018 and Russia 2019 data. The isographs show specific shapes, depending on the country and the year. Whereas the U.S. isograph shows complex fluctuations (a shaky shape of inequality) that the Gini (1914) coefficient and other distributional indicators cannot represent accurately, the isographs of Russia and Mexico are relatively constant.

When an isograph is constant, this means income y has a Fisk (aka Fisk-Champernowne, FC) distribution. An FC(1/a,b) distribution has 2 parameters, 1/a and b: $x_{FC}=b\ (p\ /\ (1-p))^{1/a}$ where b is a scale parameter and a (=1/gini) is an equality parameter. The Gini coefficient of FC(1/a,b) is 1/a (Dagum 1975, Kleiber and Kotz 2003: 224). Thus, gini is the Gini index of the distribution, and it also the value of the constant on the flat isograph. Mexico, with a flat isograph at the approximate level of 0.425, has a Gini index of 0.427.

The FC(gini,b) has remarkable characteristics. Its isograph has a simple formula: ISO=gini, otherwise: Y=gini.X. This means each gain of one unit of logitrank X (anywhere across the entire X domain) means a gain of income y equal to exp(gini). In an FC distribution of Gini=.427 (Mexico), moving from X=1 (top quartile threshold) to X=2 (near top decile) means an increase of income of a factor exp(.427), i.e. a relative increase of 53.2%, the same increase as between X=2 to X=3 (from fractile p=0.88 to 0.95). In a more income-equalitarian dataset with relatively flat isograph with a Gini index of 0.328 and a relatively constant ISO of approximately 0.33 (the Russian data), the gain of one unit of X means an income gain of only 38.8%. In Mexico, the top 5% (X=3) is 3.56 times the median; in Russia, it is only 2.57. In Mexico, the bottom 5% (X=-3) is 1/3.49 times the median; in Russia, it is 1/2.61, which are logic proportions in relation with FC distributions and isographs.

Figure 1: isographs of U.S. (blue) together with Mexico (green) and Russia (red)



job 1018517 submitted Thursday 24 November 2022 at 11:58 https://www.louischauvel.org/job_1018517.txt

Now the U.S. isograph shows similarity with Mexico below the median and with Russia at the top of the X scale above X=2. In the U.S., at X=2, a gain in one unit of X produces a gain by exp(.36) that is 41% (close to the Russian value). At X=-3, a gain of one unit of X means an income gain of exp(.42), i.e. 52%, as in Mexico. In the U.S., the top 5% income threshold (X=3) is 2.9 times the median (close to the order of the magnitude of Russia), and the lowest 5% are 1/3.4 (as in Mexico) below the median.

This means that the U.S. is like Mexico on the median and below and like Russia at the top. Of course, the interpretation should be cautious: this is what the datasets tell us. Russian apparent equalitarianism is certainly based on the impossibility to survey billionaires, among other hidden realities. The U.S. equalitarianism for top incomes contrasts with the extreme wealth inequalities at the top (Chauvel et al. 2019): a good tax lawyer can transform an income-rich American into a wealthy person with a not-so-high taxed income.

The general purpose of the isograph is to detect variations in the intensity of inequality: in this demonstration, the U.S. appears as less unequal at the top of its distribution than below its median.

Theoretical distributions and their isographs

Comparing isographs of typical distributions (fig 2) helps us to understand their diversity:

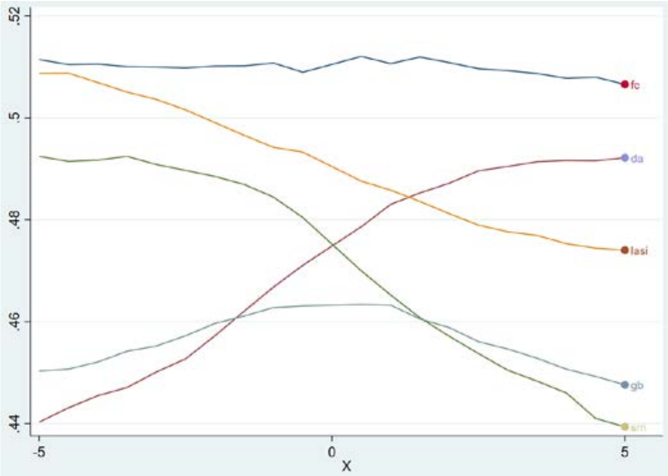
* the FC Fisk (aka Fisk-Champernowne) distribution (=gb2(a,b,1,1)) with parameter a=1.96 (so its Gini index gini=1/a is .51).
* along with other distributions of the gb2 family gb2(a,b,p,q). See Jenkins (2009) for notations and advanced developments on the gb2 family and Kleiber and Kotz (2003: 188) for a more general presentation of the gb2 family. Here the first parameter of the gb2, a, commands the general level of inequality, b, is a scale parameter, and p and q influence the right and left tails. The distributions

below have been randomly generated with Van Kerm's (2017) "grndraw" command[2], with one million observations each:

- Isograph cf: gb2(1.96 1 1 1)
- Isograph da (Dagum): gb2(2 1 1.2 1)
- Isograph sm (Singh-Maddala): gb2(2 1 1 1.2)
- Isograph gb: gb2(2 1 1.15 1.2)
- Isograph ls: a LaSi distribution with parameters lambda=.49 and sigma= -.005.

We consider the isograph of each distribution (fig 2) and their Gini coefficients (gini) along with two indicators (tab 2): the relative poverty rate at level 50% of the median (pora50, abbreviated in the following as pora).

Figure 2: isographs of 5 major types of distribution simulated on N=1,000,000



Source: replication file https://www.louischauvel.org/simu_gb2_dagum_SM.do

Table 2. Gini index and poverty rate (threshold 50% of the median)

fig1 distribution	gini	pora50
fc gb2(1.96 1 1 1)	0.509	0.204
da gb2(2 1 1.2 1)	0.486	0.182
sm gb2(2 1 1 1.2)	0.453	0.194
LaSi l=.49 s= -.005	0.479	0.199
gb2(1.93 1 1.2 1.2)	0.452	0.182

The isograph of FC(a=1.96) is confirmed to be a flat line, constant at level 1/a=0.51, equal to its

2 See installation and specifications in PVK 2017: Philippe Van Kerm, 2017. "_GRNDRAW: Stata module for random number generation from the gb2, Singh-Maddala, Dagum, Fisk and Pareto distributions," Statistical Software Components S458350, Boston College Department of Economics.

Gini, and the poverty rate is 20.4%.

With those parameters, the Dagum distribution, i.e. a gb2 with q=1, is flat on the right at level .49, the value of its Gini shows a lower level of inequality on the left and: the da poor are more equal in the sense they are closer to the median, so the da poverty rate is only 18%. Conversely, the Singh-Maddala distribution, i.e. a gb2 with p=1, is symmetric to the da distribution. It is flat on the left at level .49 and declines on the right below .45. The sm distribution has a Gini of .45 and a poverty rate of 19%. Compared to the da distribution, its Gini is lower but has a higher poverty rate.

The example of gb2 confirms that it is a more flexible distribution on the right and the left because the p and q parameters are able to govern the two tails of the distribution. In this gb distribution example, the Gini is similar to the sm and the poverty rate is close to the da.

The last example of a LaSi(lambda=.49;sigma=-.005) (see below LaSi distribution) has a straight line isograph ("quasi-linear") intersecting X=0 at lambda=.49 and having a slope of sigma=-.005. Its poverty rate is similar to the cf example here, and its Gini is .47, equidistant to the ones of da and sm.

The five examples presented in figure 2 are typical: poverty rates have a link with the behavior of the isographs on the left, and the Gini indices can be approximated on the right of the isographs, somewhere close to X=1.5. These relations will be generalized later on. The isograph is able to detect the differences of morphologies of inequality in different empirical cases (above: mx us ru) and parametric distributions.

Empirical implementation of the isograph to 53 LIS countries

The isograph is implemented to the most recent year sample of each of the 53 countries included in the LIS (fig 3) with y scale adapted to the country/year (see annex 5 when scale is common). The estimates of ISO are presented with the bootstrapped 95% confidence intervals, to differentiate between random fluctuations and significant variations in the isograph. A conservative strategy of bootstrap is implemented this analysis with n (here, 30 in general) repetitions of the random uniform selection of 50% of the samples. It is conservative in the sense that it overestimates the confidence intervals. The first result is the diversity of isographs across the 53 countries (fig 3). Different basic shapes are visible.

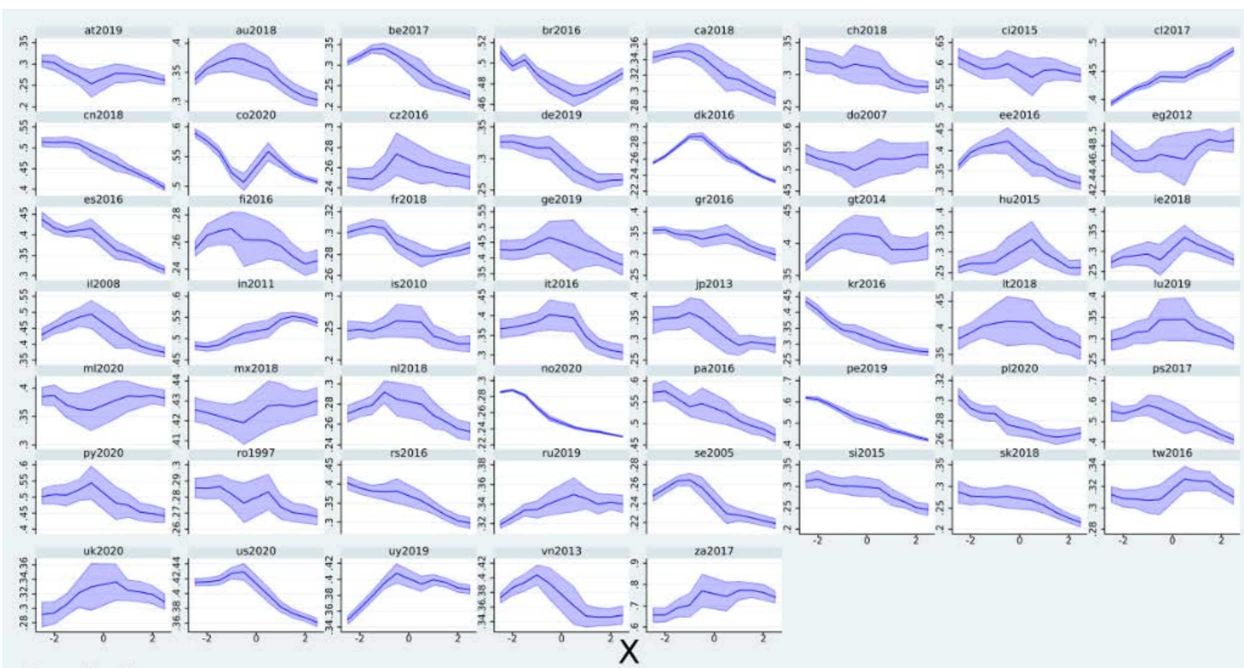
Some isographs are not significantly different to constant value of ISO (characteristic of a Fisk distribution): countries like mx, ml, hu, etc. are characterized as an FC distribution of constant a=1/gini. A larger proportion of countries present a quasi-linear shape, some with positive slope (f.ex. cl), and many more with negative slope and decreasing profile (at, pe, rs, etc.). Depending on the sample sizes (see annex 1), some narrower confidence intervals confirm non-linearity. In some countries (au br do eg it, etc.) these non-linearities are relatively smooth curvatures but in other countries (dk ee il se us, etc.) we observe sharper deviations from the linear trend.

The shape of the 53 countries isographs suggests the general morphology is a combination of a straight line with a constant at X=0 and a slope, plus a non-linear component: a deviation from the linear trend. This implies three main inequality patterns:

- a general level of inequality (higher or lower overall isograph) with an order of magnitude "close" to the Gini index.
- a general slope of the isograph, to be interpreted in terms of quasi-linear variation of inequality from the bottom to the top, a positive slope indicating relative equality at the bottom compared to stronger inequality at the top. The majority of slopes are negative, but positive cases exist. Many countries have more inequality at the bottom than at the top, which looks strange because the elevation of poor incomes (reduction of inequality between the poor) could be "less expensive", socially, than the reduction of higher incomes. Is this the sign of efficient progressive income tax? Is this the consequence of legal tax optimization where incomes are automatically converted into untaxed savings?
- One non-linear residual, with one hump (or wave) shaped profile. We notice a small number of countries show significant double non-linearities (co no).

The isographs of the 53 countries confirm the diversity of distributional morphologies, with general linear patterns and also specific fluctuations. The confidence intervals of 95% confirm that, in the general case, isographs are significantly different from the general shapes of classical distributions.

Figure 3: isographs of 53 countries available in the LIS (most recent year) (y scale variable)



https://www.louischauvel.org/job_1018520.txt job 1018520 submitted Thursday 24 November 2022 at 12:01

gb2 approximation and the 53 isographs

The gb2 distribution (McDonald 1984, McDonald & Xu 1995, Kleiber and Kotz 2003, Jenkins

2009, Jorda et al. 2021, Kot and Paradowski 2022, Sarabia et al. 2021) is often presented as the standard of income distributions for its empirical pervasiveness and theoretical accuracy. We rely on the notations, including the gb2 density $f_{gb2}(x)$ where x is income>0:

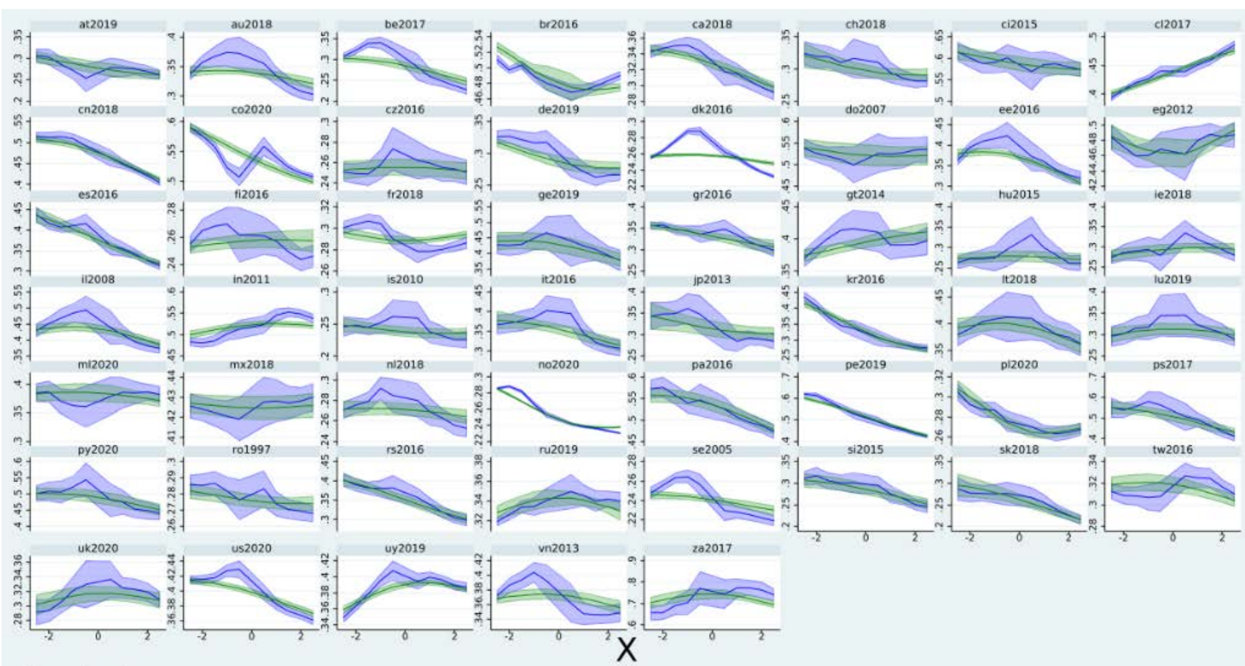
$$f(x) = \frac{a x^{ap-1}}{b^{ap} B(p, q)[1 + (x/b)^a]^{p+q}}, \quad x > 0.$$

(6.5, Kleiber and Kotz 2003: 184)

The quantile function of gb2(aa, bb, pp, qq), provided by Van Kerm (2017), is useful to simulate distributions at percentile p=invlogit(X): y_{gb2}=bb ((1/invibeta(pp,qq,p))-1)$^{(-1/aa)}$

The gb2lfit stata command developed by Jenkins (2014) is a convenient tool to systematically fit empirical income distributions as gb2 distributions. The 53 empirical isographs are presented here (fig 4) along with the respective isographs of their gb2 estimates and their 95% confidence intervals on the domain X=[-3;3]. In 38 of the 53 cases, gb2 is an acceptable approximation of the empirical distribution (almost perfect overlap with no extremely significant divergence). In these cases, gb2 is an acceptable model of the empirical distribution. H, in the 15 other cases, there are significant and even substantial divergences between the empirical isograph and the gb2 solution: in particular in be (co) dk ee fr in (no) se us (uy). In these ten cases on 53, gb2 have obvious difficulties to fit the target.

Figure 4: isographs of 53 countries available in the LIS (most recent year) plus gb2 estimates



https://www.louischauvel.org/job_1018701.txt job 1018701 submitted Thursday 24 November 2022 at 22:45

There is a simple explanation to this limitation of gb2: all the 53 cases show that the gb2 effectively models the monotonous, quasi-linear pattern of empirical isographs, but fails to accurately represent the sharper non-linearities observed, for instance, in the Nordic countries and the U.S.

on the domain X=[-3;3]. In a dozen cases, the sharpness intensity of the observed humps is out of reach of the gb2 capacities of flexibility.

When gb2 is significantly challenged by observed sharp humps, the isograph of the estimated gb2 (four parameters) is generally similar to a simple linear (level+slope.X) equivalent that represents a distribution with three parameters (the third one being the scale parameter b). The gain of the gb2 over a simple quasi-linear approximation of the isograph is limited. A challenge for gb2 here is its lack of parsimony, its excessive mathematical complexity, and the difficulty to interpret its parameters that lack intelligibility. This raises two main questions:

* Q1: because the added value of gb2 relies firstly in the quasi-linear subcase, is it possible to simplify it in an easier intelligible parametric distribution more convenient than the complete gb2?

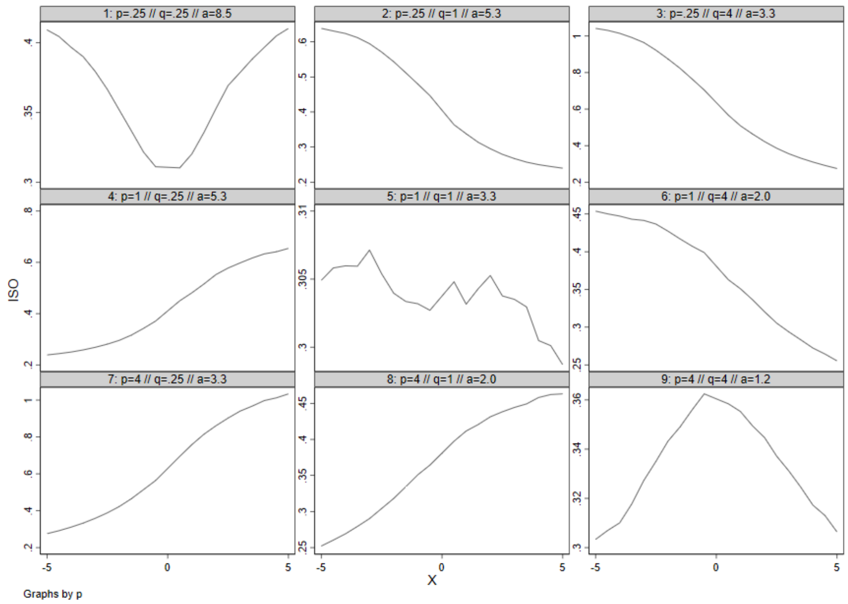
*Q2: because this quasi-linear approximation remains insufficient to correctly fit the target (at least for countries like be co dk ee fr in no se us uy), what improvement should be implemented to substantially improve the quasi-linear approximation with a model of the hump component?

Configurations of gb2 to approximate quasi-linear isographs

There are two important subgroups inside the gb2 distribution, and surprisingly it is not the Dagum versus Singh-Maddala simplification (when q=1 versus when p=1, respectively). A more relevant divide in gb2 shapes are when p=1/q (gb2a) versus when p=q (gb2b). The two types appear clearly with simulations (fig 5)

- The gb2a(a,b,p,q=1/p) pertains to the case where the isograph is monotonous, quasi-linear on the domain X=[-3;3], and more generally a sigmoid shape [tanh(X/4)] on the isograph for X=[-10;10].
- The gb2b(a,b,p,q=p) generates isographs with a curvature similar to a second degree polynomial on the domain X=[-3;3] and a hyperbolic secant (= inverse of hyperbolic cosine) shaped isograph for X=[-10;10] of shape [1/cosh(X)]. The profile of the gb2b on the isograph is a typical smooth hump, with a long wavelength of 4. By comparison, the humps detected in figure 4 are more typically wavelength 1, meaning the best fit of humps in isographs of countries like be or us are of shape [1/cosh(X)], and not [1/cosh(X/4)], which is insufficiently flexible. Simulations with random variations of a and p in the gb2b confirm the weak flexibility of the gb2b that is not sufficient to accurately fit the empirical humped isograph (fig 4), like be dk or us for instance.

Figure 5.1 to 5.9: configurations of gb2 in the isograph with different values of p and q (y-axis is rescaled for each individual graph)



Source: STATA simulation http://www.louischauvel.org/sim_gb2_shape.do

When p=1/q (fig 5.3 5.5 5.7), the gb2 in the isograph shows a monotonous shape, quasi-linear with an almost constant slope from X=-3 to X=3. When p>1, the isograph slope is positive: in this configuration, inequality is higher in top incomes than in bottom incomes, with a ratio (top decile to median) by (median to bottom decile) above 1.

With gb2b, when p=q, the isograph is symmetric (fig 5.1 5.9) or flat (fig 5.5). The variation of p has a role in terms of polarization: when p>1 (and so q>1, fig 5.9), inequality at the center is stronger than at the extremities. This means a relative polarization close to the median: when p>1, the (IQR^2/IDR) (squared interquartile ratio IQR to interdecile ratio IDR) is above 1 (fig 5.9). Polarization at the center exists when the IQR is stretched relative to the IDR. When p<1, density at the center is higher than in the quasi-linear isograph configuration, leading to a ratio (IQR^2/IDR)<1 (fig 5.1). In the general case (fig 5), the gb2 combines these elementary shapes in the isograph, leading to a mixture of lower/higher level, positively/negatively sloped curves, with different relative intensity of polarization. These different isograph shapes of the gb2 family distribution show the relative diversity of their types and also their limitation compared to the observed empirical isographs of the 53 country cases.

LaSi distribution approximating distributions with quasi-linear isograph

Consider, first, the gb2a more systematically. In the isograph, gb2a is primarily a sigmoid shaped curve on the domain X=[-10,10], quasi-linear on X=[-3,3]. To bypass the analytical and computational complexity of the gb2, and the difficulty to find a clear interpretation of its parameters, we develop a new distribution, LaSi, with a shape close to the gb2a (a sigmoid in the isograph):

ISO_{LaSi} = lambda + 4 sigma tanh(X/4) where tanh is the hyperbolic tangent function

The quantile function of LaSi is y_{LaSi} =b exp(X*(lambda + 4 sigma tanh(X/4)))

The LaSi(b,lambda,sigma) distribution y_{LaSi} is a three-parameters distribution including b as a scale parameter, lambda as an overall inequality parameter (lambda the level of the isograph at X=0), and sigma as a shape of inequality parameter (sigma is the slope near X=0). Positive sigma means more inequality at the top and less at the bottom, compared to the LaSi(b,lambda,0)=FC(b,lambda) distribution. Thus, the median is far from the rich and close to the poor when sigma is positive: otherwise, richer rich and richer poor.

Three referential inequality measurements can be related with the LaSi parametrization:

*the gb2a, i.e. the gb2(a,b,p,q) where p=1/q

*the Gini index and pora, relative poverty rate with a poverty threshold at 50% of the median

*the decile thresholds d1 and d9 of the medianized income (d5=1)

The lambda and sigma parameters can be expressed from these three systems.

LaSi and gb2b aa and pp

The LaSi distribution is very close to a gb2a(aa bb pp) distribution (bb is a scale parameter, different to the b of LaSi): simulations show $ISO_{gb2a} \approx ISO_{LaSi}$ with parameters:

lambda≈1/aa+.162(ln(pp))^2 (R2=0.9998 on 100 simulations N=10,000 of LaSi // http://www.louischauvel.org/simu_gb2_slope.do)

sigma≈(.2)ln(pp)/aa (R2= 0.9976 on 100 simulations N=10,000 of LaSi // http://www.louischauvel.org/simu_gb2_slope.do)

The LaSi(b lambda sigma) distribution has three intelligible parameters and is similar to the gb2a(aa,bb,pp,1/pp). The previous equations transform (aa,pp) to (lambda,sigma). The LaSi distribution is an answer to question Q1 above: in the quasi-linear case, the complete gb2 (four parameters) can be replaced by the gb2a (three parameters), or even by the more convenient, intelligible parameters, LaSi (three parameters).

LaSi and Gini/pora

In the quasi-linear case, with Gini coefficients and pora rates, lambda and sigma may be expressed:

lambda≈0.62.gini-0.265/logit(pora) (r2=0.9998) [simu_gb2_slope_b.do]

sigma≈0.29.gini+0.204/logit(pora) (r2=0.9923) [simu_gb2_slope_b.do]

These formulae stand for LaSi distributions, not in the general case where the isograph is not quasi-linear. Yet in the 53 country cases, this formula predicts lambda with an R2= 0.982 and, when 2 outliers are left (za and ee), sigma is predicted at R2=0.789. The same order of magnitude of R2s is obtained on the complete 655 countries LIS database (Annex 4).

Conversely, in the LaSi distribution, the Gini index may be easily approximated by randomized simulation:

gini≈lambda+(1.44)sigma (R2=0.9997 100 simulations N=10,000 of LaSi // http://www.louischauvel.org/simu_gb2_slope.do)

This means the value of the Gini index is the isograph near level X=1.5 or p=0.8, close to the quintile threshold q4.

The estimation of pora (relative poverty rate at level 50% of the median) is more complex. Since pora is the proportion of the poor below income threshold y_{pora}=(-.5).median, its logitranked value X_{pora}=logit(pora) is the solution of polynomial P:

(P): sigma $(X_{pora})^2$ + lambda X_{pora} – ln(.5)=0

then X_{pora}=(sqrt(det)-lambda)/(2sigma) and then pora=invlogit(X_{pora})

where: det=lambda^2+4.ln(.5).sigma, the determinant of polynomial (P).

If sigma=0, pora=invlogit(ln(.5)/lambda)

An approximation of this formula is pora≈invlogit(ln(.5)/(lambda-(2.11)sigma))

[http://www.louischauvel.org/simu_gb2_slope_b.do]

LaSi and d1/d9

In the quasi-linear case, lambda and sigma have immediate expressions in terms of d1 and d9.

lambda=(ln(d0)-ln(d1))/(2logit(.9)) and sigma=(ln(d0)+ln(d1))/(4logit(.9))

LaSi versus Atkinson indices and general entropy

Major indicators of inequality like Atkinson's indices and general entropy measures (see Atkinson 1970, Bourguignon 1979, Jenkins 1991, Cowell 2000) find relatively simple expressions in terms of lambda and sigma through (generally linear) regressions. Analytical translations of these relations might be difficult. Without details [see http://www.louischauvel.org/sim_LaSi_ineqdeco.do] one can obtain:

Atkinson(2)=1.61*lambda-.21 (r2=.9988) and

Atkinson(1/2)= (.931*lambda- 2.18*sigma)^2 (r2=.9987) ln(GeneralEntropy(-1))= 7.26*lambda-3.90 (r2=.9965)

ln(GeneralEntropy(2))= 11.6*lambda+41.2*sigma-4.9 (r2=.9800)

With quasi-linear distributions (LaSi), the systems of inequality indices can be mutually translated. Those relations are complexified in case of polarizations modelled with LaSiPiKa distributions .

Back to research questions

The question Q2 above actually poses two different questions:

* Q2a: is the gb2(aa,bb,pp,qq), with four parameters, able to outperform LaSi that have only three?

*Q2b: when the isograph is not quasi-linear (f.ex., in the U.S.), is it possible to improve LaSi?

NB: As a complementary remark, the LaSi distribution is extremely close to another remarkable, convenient one, the "dissymmetric log-logit" DLL distribution. DLL: $z_{DLL}=b\ p^{(down)}/(1-p)^{(up)}$, where b is a scale parameter, down and up parameters representing the limits of the isograph at X=-∞ and +∞, respectively. LaSi(b,lambda,sigma) is equivalent to DLL(b',down=lambda-5sigma,up=lambda+5sigma). Simulations confirm that LaSi(b,lambda,sigma), L5S(b',lambda,sigma), and gb2a(a,b'',p,1/p) (with lambda≈1/a+.15 $(\ln(p))^2$ and sigma≈.2 ln(p)/a), are similar, with quasi-linear isograph on the domain X=[-3,3].

Defining LaSiPiKa and comparing performance with FC, LaSi, gb2(abpq)

After the systematic observation of the empirical isographs and of the residuals of their LaSi estimates (fig 8), I elaborate a complement "hump" shaped term, "PiKa" based on a smooth hyperbolic trigonometric function [pi/cosh(X-kappa)] where pi is a parameter measuring the intensity of a polarization (stronger inequality at a specific level) and kappa the epicenter of the polarization X_k. Figure 11a exemplifies the logics of LaSiPiKa distribution and its parameters. The two parameters model the intensity of the hump and its location on X. A more complete, three-parameter term can be proposed with a parameter mu of wavelength: PiKaMu [pi/cosh(mu(X-kappa))]. When mu is high, above 1, the hump is sharper with a smaller wavelength, and when mu is below 1, the hump appears on a longer interval of X, with a long wavelength. The value mu=1 is a way to propose a one-size-fit-all term, simplifying the term into only two parameters: PiKa [pi/cosh(X-kappa)].

Thus, an intelligible five-parameter distribution, i.e. LaSiPiKa, is proposed:

y=b exp($ISO_{LaSiPiKa}$ X) where

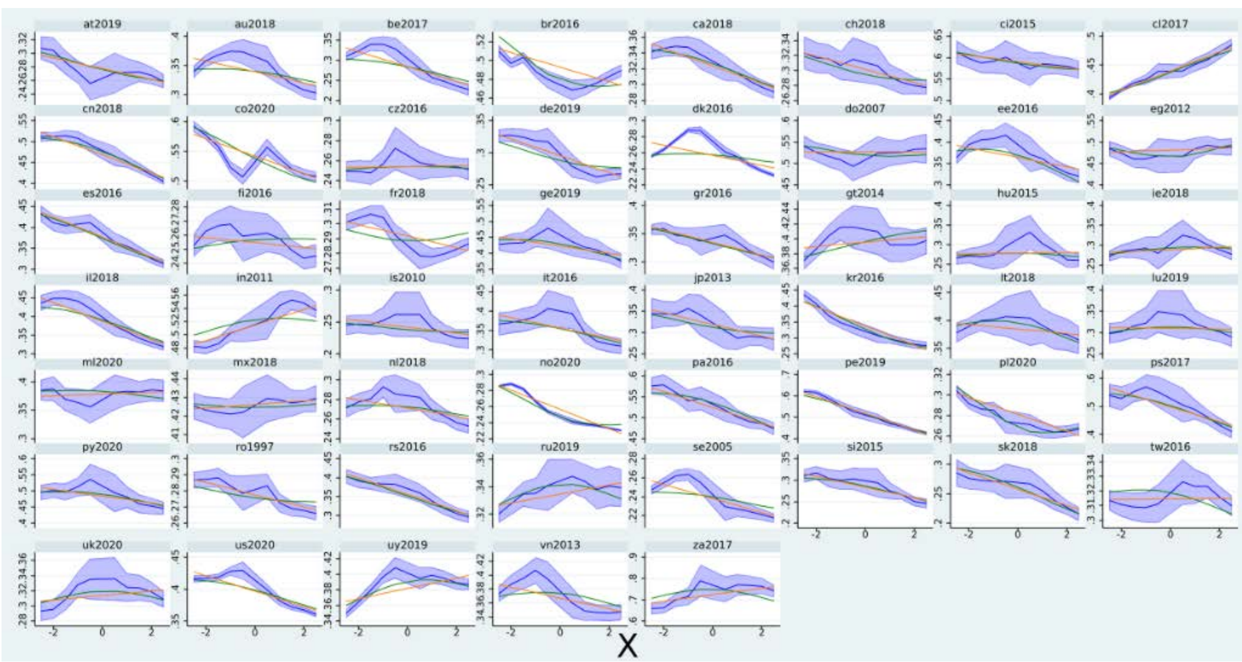
$ISO_{LaSiPiKa}$ = lambda + 4 sigma tanh(X/4) + pi/cosh(X-kappa)

We then compare on 53 LIS countries the LaSiPiKa with four other main types of distribution. This comparison confirms that the gb2 (four parameters) is not ideal:

- Fisk distribution characterized by an isograph with a constant value lambda (2 parameters including the scale parameter b). Its results are generally incorrect.
- gb2a(a,b,p,q=1/p) (three parameters). This quasi-linear shape QLS is sufficient when no hump is observed.
- LaSi distribution with ISO_{LaSi} ≈lambda+sigma*4*tanh(X/4) (with b as a scale parameter, this gives three parameters). LaSi is similar to gb2a, but with intelligible parameters.
- The complete gb2(a,b,p,q) (four parameters). Very close to QLS, and neither parsimonious nor intelligible.
- The LaSiPiKa (five parameters).

The five parameters of LaSiPiKa are (1) b – a scale parameter; (2) lambda – a parameter of the level of inequality; (3) sigma – a parameter of increasing inequality at the top; (4) pi – a parameter of intensity of polarization (or concavity: positive for a sad face); and (5) kappa – a parameter that locates the epicenter of the polarization effect. The LaSiPiKa distribution is easy to program with a non-linear adjustment command like the Stata “nl”.

Figure 6: empirical isographs (blue with 95% CI) LaSi (orange) and gb2 (green)

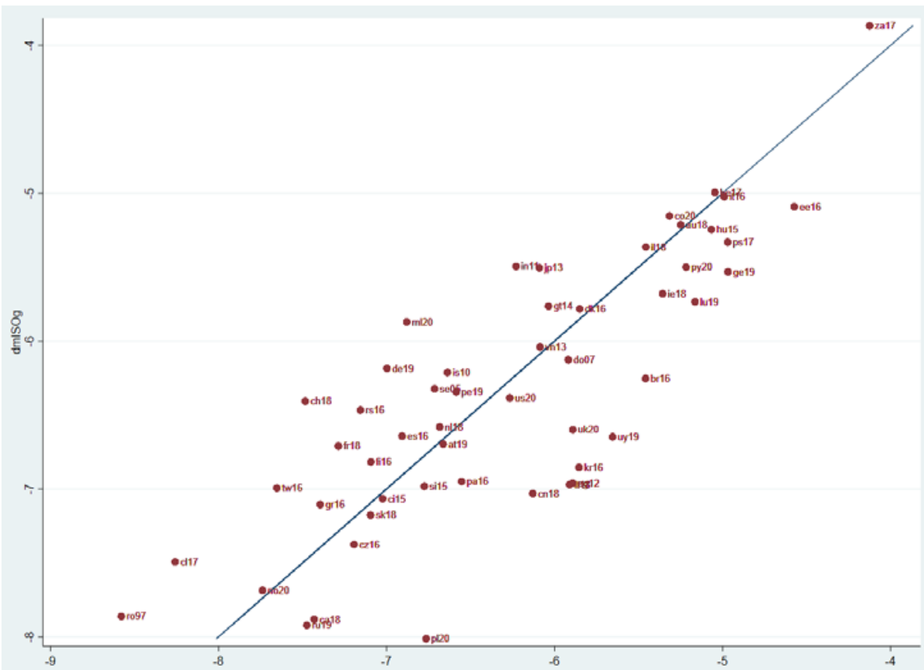


https://www.louischauvel.org/job_1018634.txt **job 1018634 submitted Thursday 24 November 2022 at 16:36**

The comparison of LaSi isographs with gb2, in relation to empirical isographs (fig 6), shows some modest improvements of gb2 on LaSi for br, cn, ee, pl, ru, uy, and za. Conversely, LaSi is not bad, compared to unsatisfying results from the gb2, in countries like be, cz, de, fi, il, in, ro, and tw. Apparently, when LaSi is not good, gb2 misses the target, too. Compared to the Fisk configuration (flat isograph), the LaSi improves the fit (with one parameter), and the added value of gb2 on LaSi is scarce for one additional parameter. The two proposals, LaSi and gb2, are equally successful in the case of monotonic quasi-linear shape, and they almost similarly fail in the case of sharp curvature.

A quantification of those comparisons might be done by comparing the logged distance between the empirical isograph and the different estimates on the interval X=[-3,3]. In this way, the comparison between LaSi and gb2 shows comparable successes and failures (fig 7). The comparison shows the two distributions equally fail for za, be, and it, and are similarly successful for mx (not represented at [-10,-10] on fig 7) and no. At any rate, gb2 outperforms LaSi for pl, cn, eg, and lt, but the inverse is observed for ml in ch, de, cl, and ro.

Figure 7: logged distance between the empirical isographs and their (LaSi versus gb2) estimates. X axis: log-distance between empirical isograph and LaSi estimate; Y axis: log-distance between empirical and gb2 (mx18, not represented, is near [-10,-10])



The comparison shows no strong general advantage of LaSi on gb2 or conversely, apart from the fact that LaSi comprises three intelligible parameters and gb2 has four non-interpretable parameters.

To improve the LaSi curve, we propose to understand the residuals of the empirical ISO by LaSi (fig 8). When significant gaps are observed between the residuals and the baseline zero, sharp curvatures (“humps”) are visible (e.g. be, dk, us, etc.). These humps, i.e. the relatively sharp non-linearities deviating from the linear hypothesis, can be understood as a relative “polarization” of income: at the level X_k (the “epicenter”, at X=κ) where positive humps are observed, the relative density of income is lower (or, otherwise, income scale is stretched at level X_k). The higher the positive residual, the stronger this polarization. The epicenter of this polarization, obviously, is not necessarily on the median X=0. This means an improvement of LaSi with a simple polynomial X^2 should fail because the center is not zero. A polynomial $(X-X_k)^2$ where X_k is the center of the “hump” should be better, but still a problem because polynomials diverge to infinite when abs(X) increases.

The general comparison in 53 cases (fig 9a) between the empirical isograph (blue), the gb2 (green), and the LaSiPiKa estimate (red) confirms the flexibility of LaSiPiKa: the adjustments where gb2 significantly misses the target are improved by LaSiPiKa.

Figure 8: residuals of empirical isographs by LaSi estimates (job 1014498)

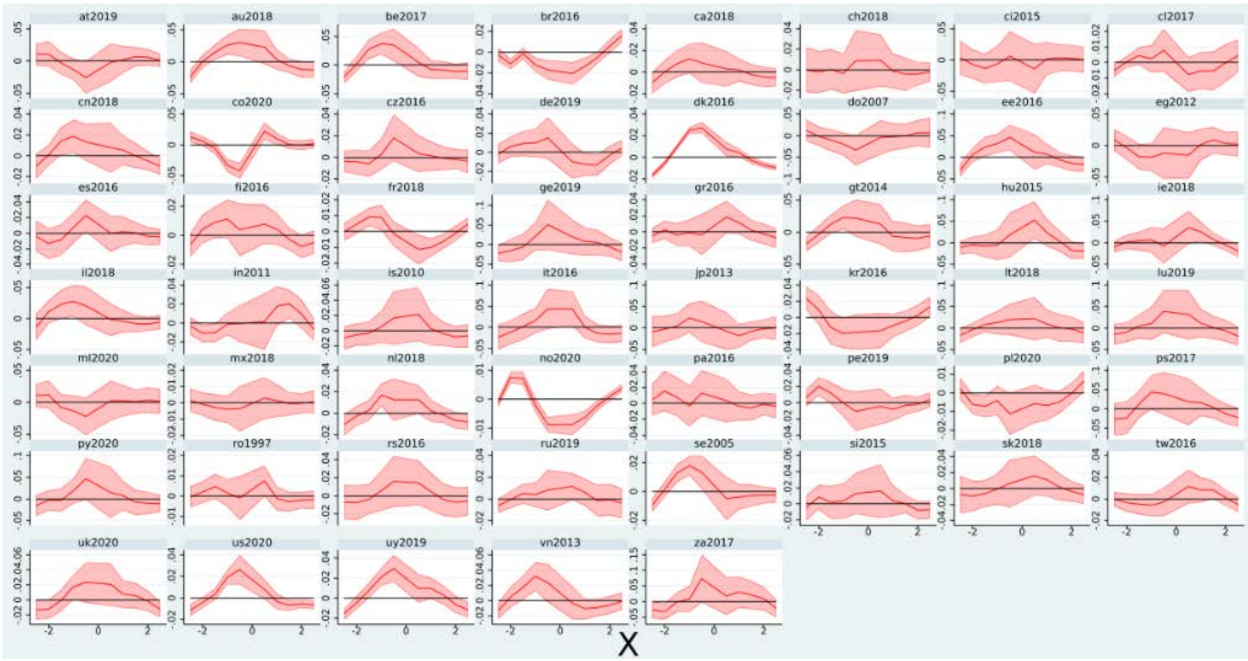
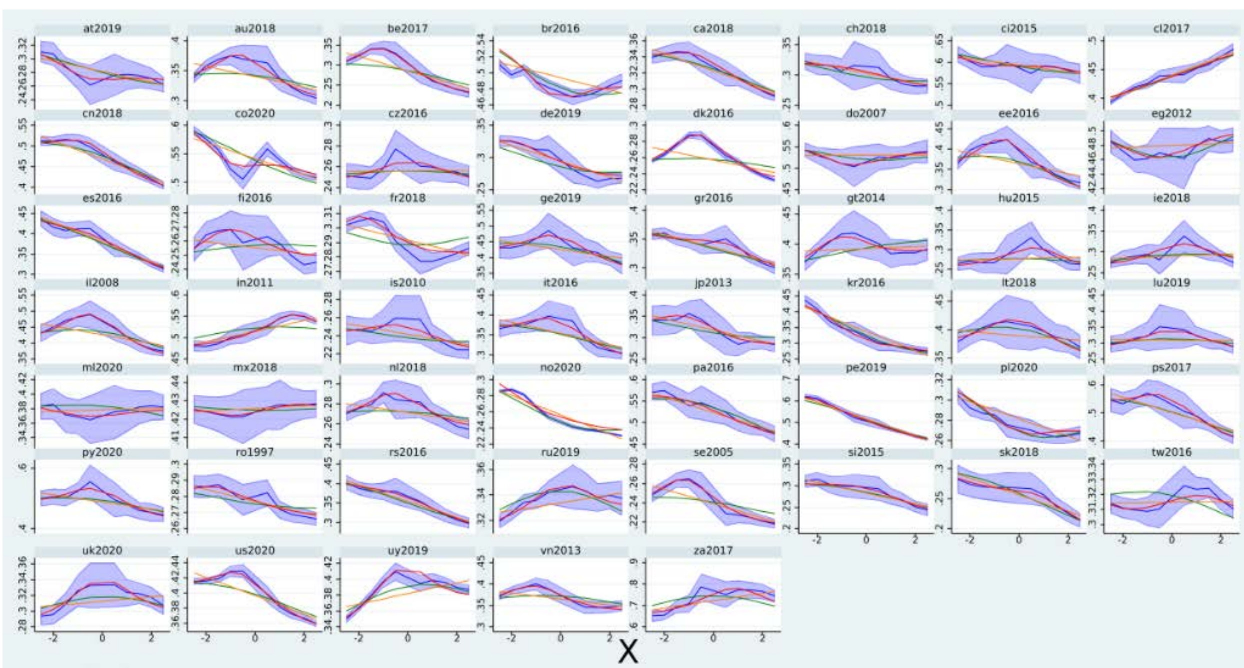


Figure 9a: 53 empirical, gb2, and LaSi isographs (as in fig 6) plus LaSiPiKa (red)



In cases where gb2 is correct (fig 9b), then LaSiPiKa fits correctly the target as well. When gb2 is not sufficient (fig 9c), then LaSiPiKa generally provides a better fit. Clearly, when the isograph is flat (do, ml, mx) or quasi-linear shaped (es, rs, sk, etc.), the performances of gb2 and LaSi are similar (then the added value of LaSiPiKa is limited because LaSi is more parsimonious). In the other cases, when non-linearities appear, then gb2 is not a real improvement compared to LaSi: therefore, LaSiPiKa makes substantial improvements. Conversely, when LaSiPiKa outperforms gb2 (fig 9c), the former seems much more flexible and able to stylize some complex shapes. Even more complicated morphologies should be developed in the specific case of co and no in particular, where a double hump model should be proposed.

An evaluation of gains in terms of accurate fit of the isographs is provided (fig 10) to compare the gains from Fisk (two parameters) to gb2 (four parameters) and then from gb2 to LaSiPiKa (five parameters). The outperformance of gb2 on Fisk is indicated by the green line, with gain situated above the diagonal, and the red line indicates the underperformance of gb2 on LaSiPiKa, and when this line is located below the median it indicates a disadvantage of gb2 relative to LaSiPiKa. The symmetry across the diagonal simply means that the gain in the fit of the empirical isograph from the Fisk to the gb2 distribution is comparable to the gains from the gb2 to the LaSiPiKa distribution. With only one additional parameter compared to gb2, LaSiPiKa offers better fits, even if br, co, and no remain more difficult cases. Another aspect is the reliability of the fit, as expressed by the sample ee16: the gb2 has systematic difficulties with Estonia 2016 and its very problematic values of a, b, p, and q, and aberrant standard errors—ee16 is an obvious outlier. For LaSiPiKa, ee16 is a normal case, correctly fitted.

Figure 9b: empirical isographs (blue with 95 % CI) (as in fig 5) plus LaSiPiKa (red) when gb2 and LaSiPiKa are similarly successful

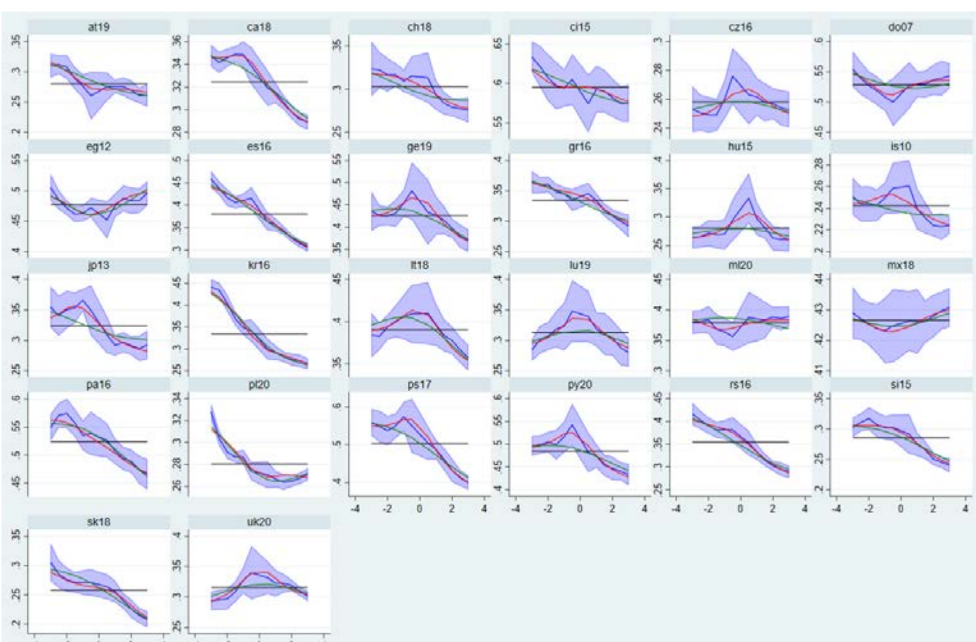
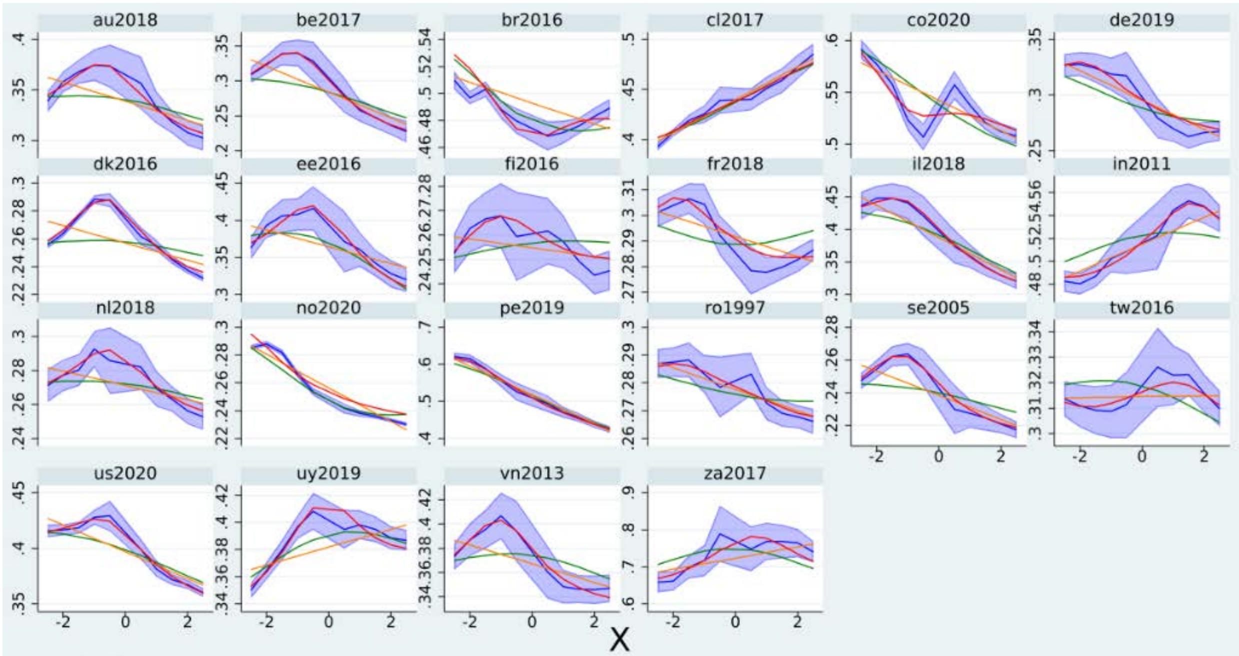
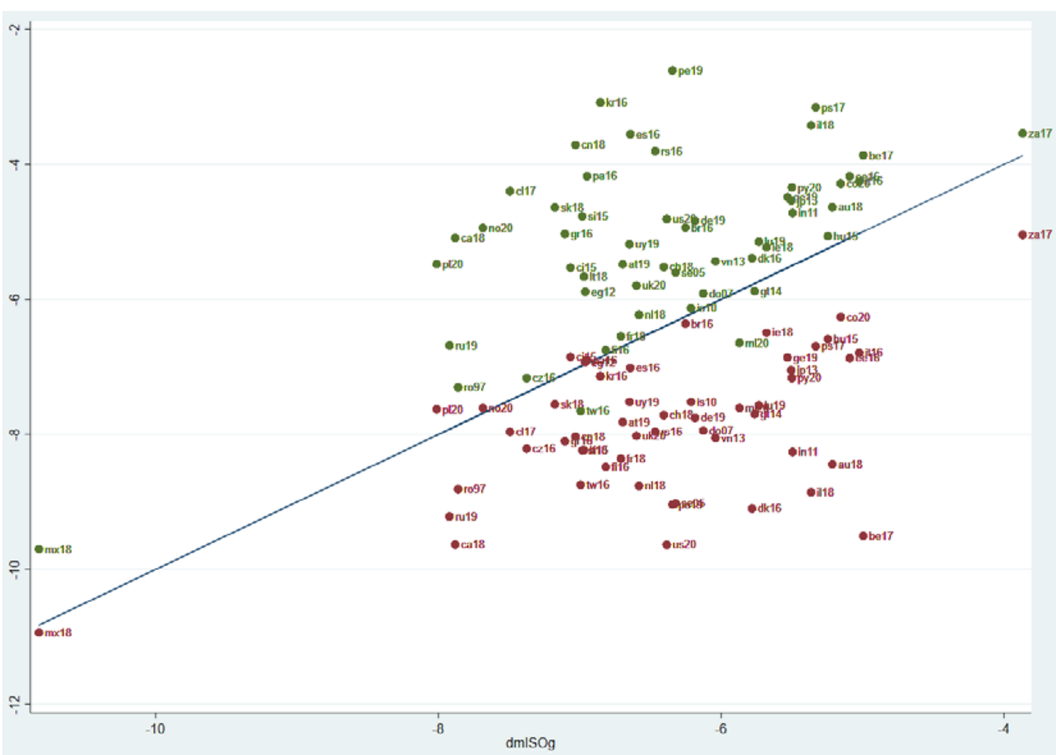


Figure 9c: empirical isographs (blue with 95% CI) (as in fig 5) plus LaSi (orange) and LaSiPiKa (red) when gb2 misses the target at least twice



https://www.louischauvel.org/job_1018634.txt job 1018634 submitted Thursday 24 November 2022 at 16:36

Figure 10: logged gains of distance between the empirical isographs and Fisk (green) and between gb2 and LaSiPiKa (red) with diagonal



At large, LaSiPiKa, an empirical improvement of the quasi-linear case LaSi with a simple non-linear term, generally makes a correct interpolation of the 53 empirical cases, even if co and no could be improved. LaSiPiKa obviously outperforms gb2.

Developments of LaSiPiKa

The parameters of the LaSiPiKa distribution have intelligible interpretations. This completes the LaSi, a quasi-linear shaped distribution of isograph ISO_{LaSi} = lambda + 4 sigma tanh(X/4) where lambda is the level of inequality at X=0 (i.e. the median, p=0.5) and sigma is the slope at X=0. As such, lambda stands for overall inequality and sigma relates to the balance between Gini versus pora, the relative poverty at 50% of the median: when Gini is high relative to pora, sigma is positive; a typical case is Chile. Conversely, a low Gini relative to poverty gives a negative sigma, like in Peru, the U.S., etc. The term PiKa [pi/cosh(X-kappa)] is a complement to LaSi that models a hump, where pi expresses the intensity of polarization (a deviation from the LaSi), located on the X scale at level kappa, its epicenter. When kappa=0, the polarization is on the median. These coefficients can be systematically computed on the 53 countries (tab 3) and on the 655 country-years samples (annex 4).

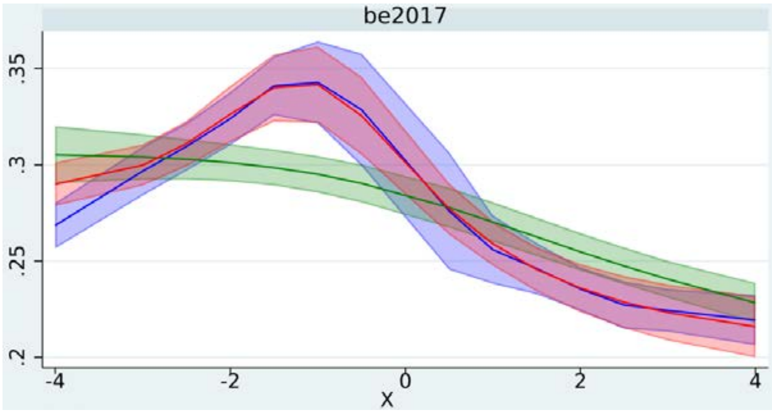
Table 3. parameters for LaSiPiKa and gb2 (with standard deviations, sd)

ccyyyy	La	si	pi	ka	sdla	sdsi	sdpi	Sdka
at2019	0.290	-0.009	-0.025	-0.189	0.006	0.002	0.012	0.531
au2018	0.312	-0.005	0.064	-0.640	0.006	0.002	0.009	0.237
be2017	0.247	-0.011	0.087	-1.019	0.005	0.001	0.011	0.137
br2016	0.515	-0.012	-0.049	-0.256	0.002	0.001	0.005	0.206
ca2018	0.312	-0.010	0.028	-0.686	0.003	0.001	0.006	0.324
ch2018	0.304	-0.008	0.009	0.070	0.011	0.004	0.046	2.049
ci2015	0.596	-0.012	-0.013	-0.148	0.016	0.003	0.035	1.318
cl2017	0.436	0.016	0.007	0.146	0.006	0.002	0.017	1.444
cn2018	0.447	-0.021	0.047	-0.487	0.004	0.002	0.009	0.302
co2020	0.560	-0.020	-0.045	-0.728	0.010	0.001	0.024	0.699
cz2016	0.245	0.000	0.021	0.182	0.004	0.002	0.006	0.498
de2019	0.289	-0.007	0.018	-1.174	0.011	0.002	0.032	1.621
dk2016	0.237	-0.003	0.051	-0.625	0.001	0.000	0.002	0.048
do2007	0.547	-0.005	-0.045	-0.412	0.013	0.004	0.023	0.630
ee2016	0.319	-0.010	0.107	-0.587	0.005	0.001	0.013	0.222
eg2012	0.501	0.000	-0.042	-0.460	0.010	0.003	0.021	0.876
es2016	0.371	-0.028	0.016	-0.117	0.006	0.002	0.014	0.823
fi2016	0.246	0.001	0.025	-1.077	0.002	0.001	0.006	0.458
fr2018	0.285	0.000	0.020	-1.398	0.009	0.001	0.021	1.114
ge2019	0.396	-0.012	0.051	-0.467	0.026	0.005	0.055	0.918
gr2016	0.325	-0.012	0.021	0.439	0.006	0.002	0.008	0.653
gt2014	0.379	0.008	0.035	-0.778	0.013	0.002	0.032	0.796
hu2015	0.256	-0.001	0.048	0.344	0.005	0.003	0.012	0.284
ie2018	0.273	0.002	0.047	0.460	0.006	0.002	0.012	0.421
il2018	0.360	-0.019	0.064	-1.157	0.007	0.002	0.013	0.279
in2011	0.490	0.002	0.059	1.450	0.012	0.001	0.028	0.738
is2010	0.233	-0.003	0.015	-0.031	0.007	0.002	0.016	0.760
it2016	0.326	-0.013	0.071	-0.145	0.009	0.002	0.016	0.260
jp2013	0.302	-0.008	0.045	-1.046	0.013	0.004	0.027	0.781
kr2016	0.356	-0.033	-0.050	-0.156	0.003	0.002	0.006	0.253
lt2018	0.361	-0.005	0.065	0.126	0.008	0.003	0.015	0.382
lu2019	0.281	-0.001	0.063	-0.015	0.007	0.003	0.015	0.446
ml2020	0.382	-0.002	-0.016	-0.182	0.009	0.003	0.034	1.969
mx2018	0.429	0.000	-0.008	-0.377	0.004	0.001	0.010	0.933
nl2018	0.258	-0.002	0.030	-0.444	0.004	0.001	0.007	0.195
no2020	0.249	-0.005	0.037	-2.781	0.001	0.000	0.002	0.060
pa2016	0.512	-0.017	0.028	-0.236	0.011	0.004	0.031	1.599
pe2019	0.502	-0.036	0.032	-1.475	0.012	0.004	0.041	1.630
pl2020	0.295	-0.010	-0.030	-0.367	0.003	0.002	0.005	0.629
ps2017	0.463	-0.027	0.097	-0.322	0.010	0.003	0.025	0.431
py2020	0.455	-0.011	0.073	-0.447	0.009	0.003	0.021	0.462
ro1997	0.274	-0.002	0.010	-1.067	0.004	0.001	0.011	1.405
rs2016	0.343	-0.023	0.026	-0.147	0.008	0.003	0.019	0.752
ru2019	0.322	0.003	0.021	0.291	0.004	0.001	0.026	1.174
se2005	0.225	-0.003	0.036	-1.161	0.002	0.001	0.006	0.145
si2015	0.270	-0.012	0.033	-0.125	0.005	0.002	0.014	0.463
sk2018	0.248	-0.019	0.021	0.310	0.009	0.003	0.022	1.281
tw2016	0.314	-0.003	0.000	-0.639	0.011	0.001	0.026	1.451
uk2020	0.293	0.002	0.051	0.048	0.006	0.001	0.014	0.293
us2020	0.378	-0.010	0.042	-0.548	0.002	0.001	0.006	0.120
uy2019	0.355	0.007	0.062	-0.174	0.002	0.001	0.005	0.131
vn2013	0.340	-0.002	0.058	-1.054	0.005	0.002	0.010	0.118
za2017	0.678	0.000	0.095	0.262	0.033	0.007	0.097	1.317

https://www.louischauvel.org/job_1014473.txt **job 1014473 submitted Saturday 12 November 2022 at 19:40**

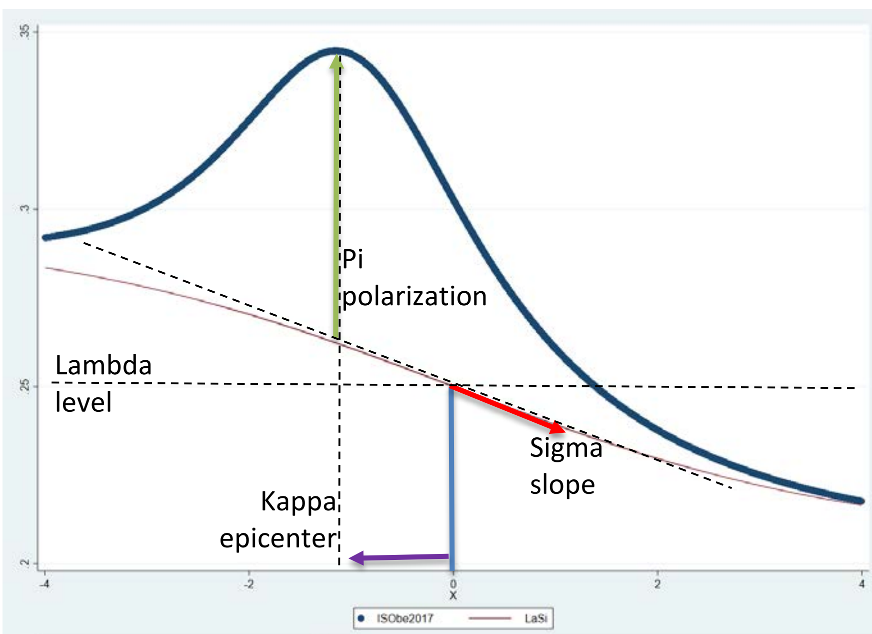
The example of be2017 illustrates the sense of those parameters and their consequences in terms of isograph shape, with lambda=0.250; sigma=-0.011; pi=0.083; kappa=-1.026 (fig 11a and fig 11b).

Figure 11b: empirical (blue), LaSiPiKa (red), and gb2 (green) isographs for be2017



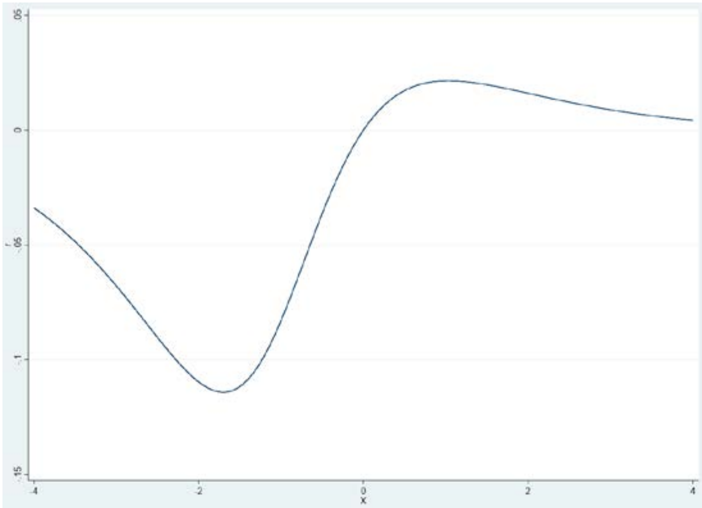
https://www.louischauvel.org/job_1014164.txt job 1014164 submitted Friday 11 November 2022 at 17:02

Figure 11a: LaSi (red) and LaSiPiKa (blue) parameters and their meaning on the isograph



The term PiKa [pi/cosh(X-kappa)] in the formula of LaSiPiKa is a polarization. When LaSi and LaSiPiKa are compared in the case of be2017 (fig 11c), the redistribution from LaSi to LaSiPiKa shows a strong decline of income below X=κ, by 11% in the case of be2017. Above X= κ, incomes decrease less, or even increase above X=0, until +1% near X=1. The polarization term PiKa, when pi>0, means a stretch of the distribution near kappa, where incomes immediately below kappa decline, and relatively increase above kappa.

Figure 11c: Relative income variation between LaSi and LaSiPiKa for be2017



These shapes have a consequence in terms of density of the distribution at different levels of the income scale. As gb2 overestimates the density right below the median of incomes on be2017, LaSi reduces the density in the lower middle class, and finally LaSiPiKa expresses the relative polarization at level kappa=-1.026 (or p=0.264), with consequences in terms of morphologies of income distribution (fig 11d).

Figure 11d: density of distribution across the income scale (vertical). Left distributions: LaSiPiKa (purple), LaSi (orange), and gb2 (green) densities (vertical). Right distributions: LaSiPiKa (purple), and the empirical one (with 95% CI) for be2017 (y-axis=medianized income, x-axis=density)

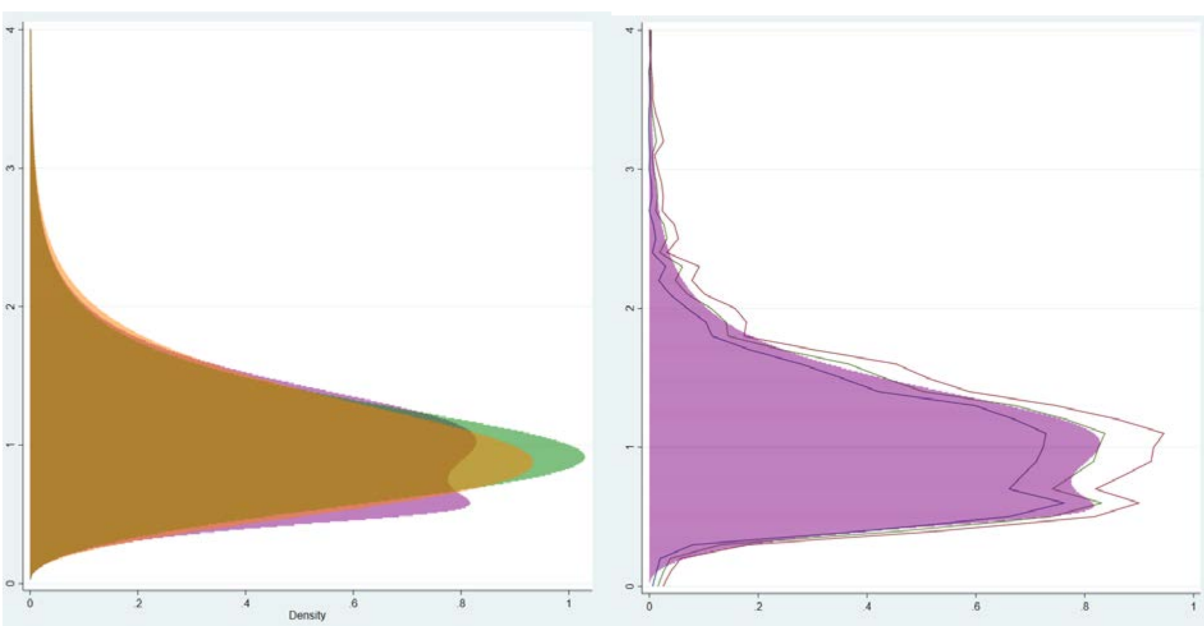
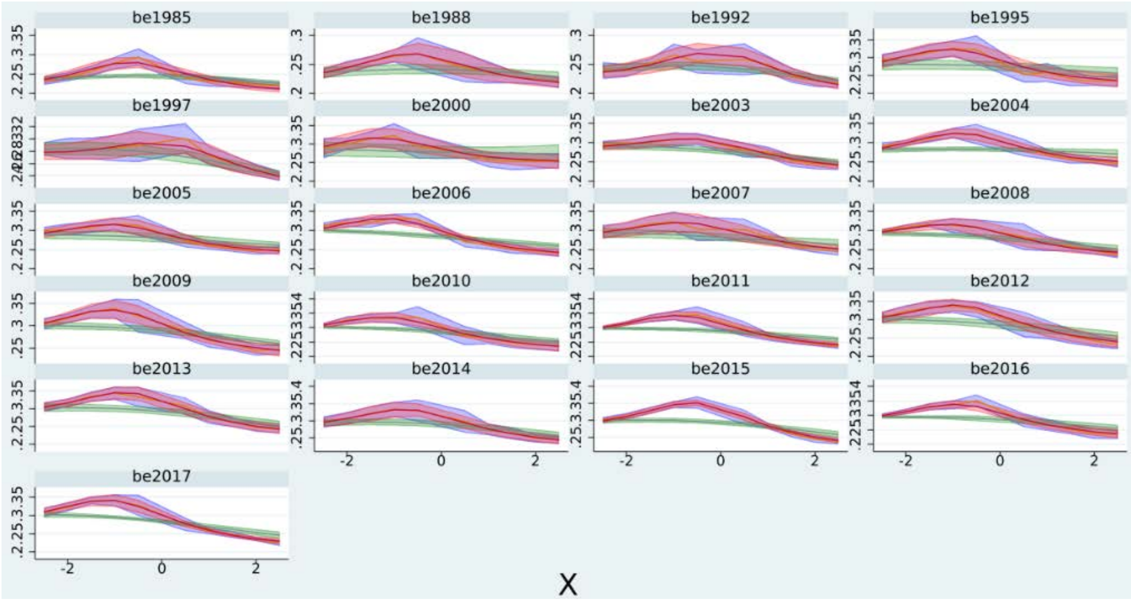


Figure 12: isograph for Belgium 1985-2017 with empirical, gb2, and LaSiPiKa distributions



https://www.louischauvel.org/job_1018768.txt job 1018768 submitted Friday 25 November 2022 at 10:48

More on sigma and the (Gini versus pora) balance, and on pi kappa

Sigma has a meaning in terms of Gini index versus poverty imbalance. A rapid proxy of sigma coefficient is given by the simplified formula (estimated on the set of 53 country-samples):

sigma≈0.2 gini-0.4 pora-0.02 (Adj. R-squared=0.65) (note proxsigma this simplified expression)

For instance, for Peru, gini=.43 and pora=.23, and then the proxsigma is -.026, and the direct LaSiPiKa based estimated value of sigma is -.035, which is graphically confirmed (fig 13). For Chile, the profile is opposite.

The values of sigma and proxsigma are overwhelmingly negative, meaning that the slopes in the isograph are predominantly negative, indicating that the ratio of the upper percentile by the median is lower than the ratio of the median by the symmetric lower percentile. The lambda-sigma map (fig 14) shows the complexity of inequality that is not only a question of level (lambda) but also of imbalance between relative poverty rate and Gini index.

Figure 13: value of slope (sigma) parameter and of prox-sigma

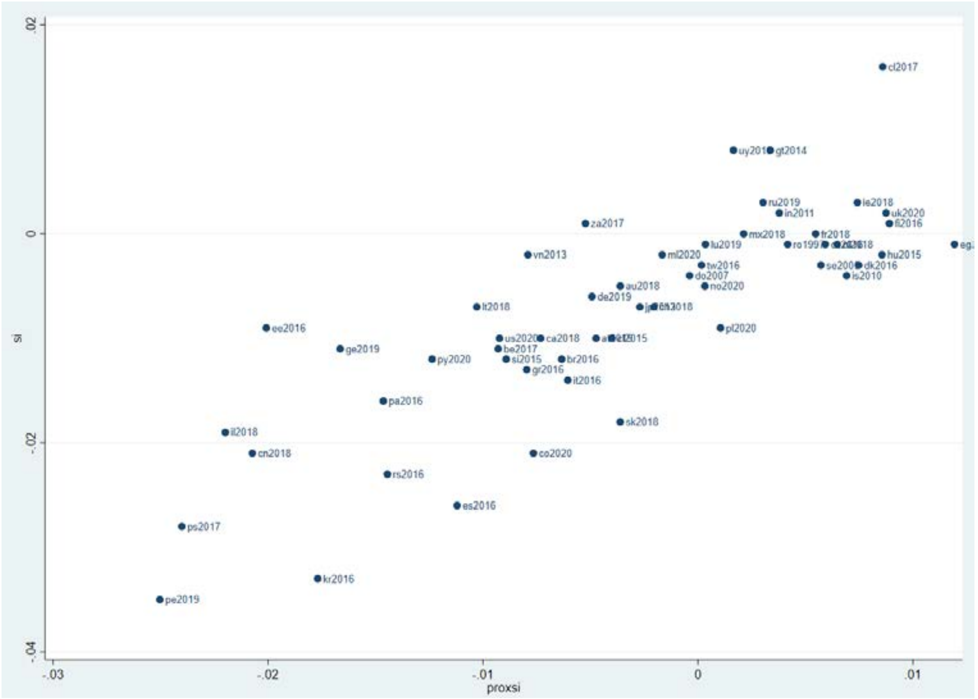
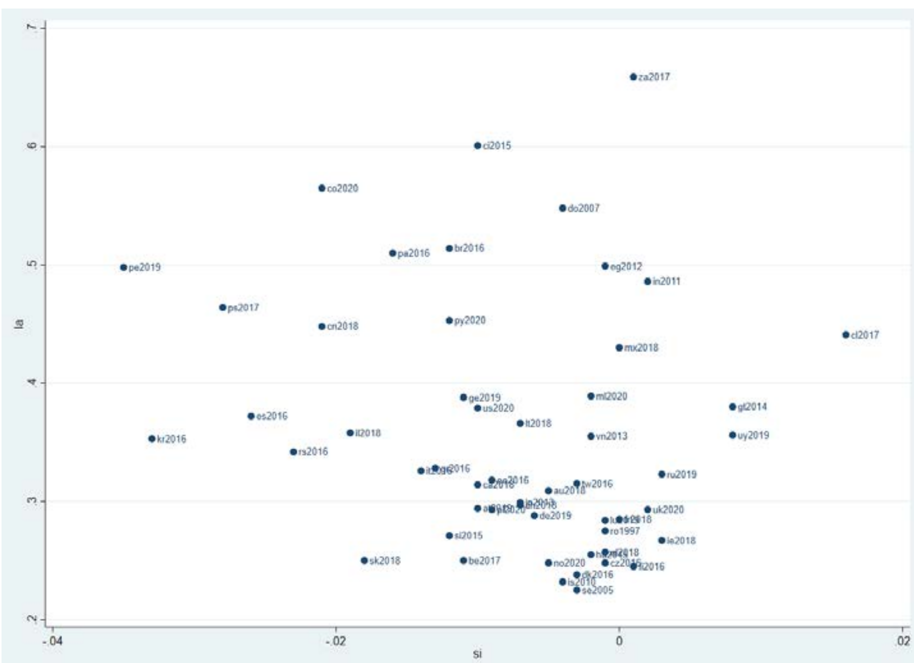


Figure 14: level of inequality (lambda) and slope (sigma) parameters of LaSiPiKa



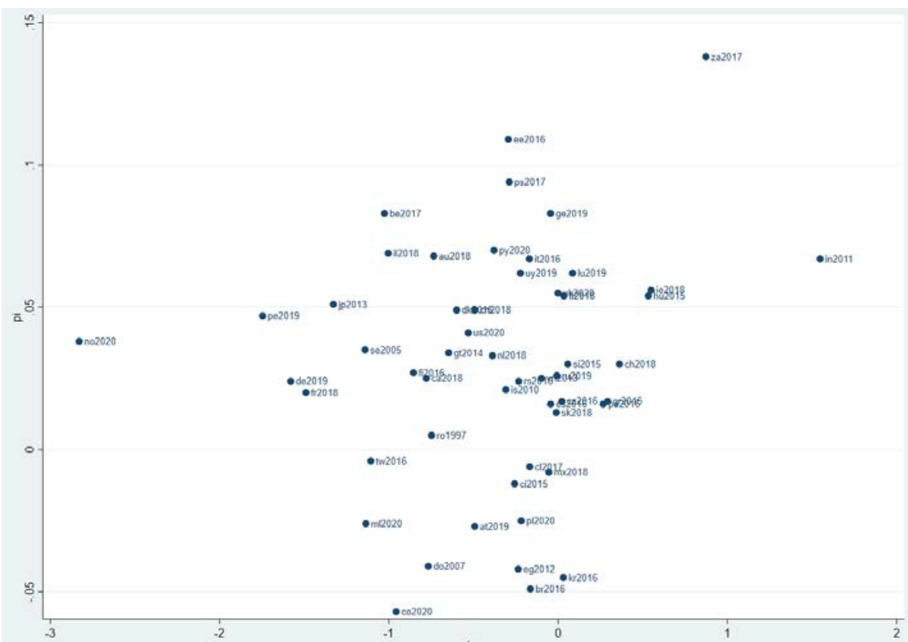
Another approximation of sigma comes from the deciles to the median ratios.

Sigma≈0.075 (ln(d1 d9/d5^2)) (Adj. R-squared=.74) (note prox2sigma this simplified expression) where d1, d5, and d9 are the first decile threshold, the median, and the ninth decile threshold, respectively. Sigma is then positive when logged upper-percentiles are farther to the median relative to the logged lower-percentiles. An important result is that, because sigma is generally negative, in many countries, the median d5 to d1 ratio is higher than the d9 to the median ratio. The poor go farther to the bottom away from the median, compared to the degree to which the rich rise far above the median. This means an important dissymmetry in (logged) distributions.

The parameters lambda and sigma relate to the simple LaSi distribution. The LaSiPiKa is its complexification, where a hump centered on X=κ and of intensity pi is introduced in the LaSi distribution. Then, the parameter pi relates to the intensity of polarization and kappa the location on the axis X of this polarization.

On top of lambda and sigma, the pi and kappa add polarization intensity and the location of polarization on the X scale. The map of pi and kappa help to understand the deviation to the linear trended isograph of the empirical isograph: positive pi means relative concavity on the isograph and negative pi comes with convexity (smiley). Then, kappa expresses whether this hump is located below or above the median (fig 15).

Figure 15: polarization (pi) and center (kappa) parameters of LaSiPiKa



The LaSiPiKa parameters are linearly independent (see the 53 countries based correlation matrix), meaning the capacity of each parameter to catch an independent dimension in the set of distributions. On the contrary, the a, b, p, and q parameters of gb2 (za and ee excuded because they are outliers) show important collinearities (tab 4).

Table 4. Correlation matrix between coefficients LaSiPiKa and between gb2 (a, b, c, d)

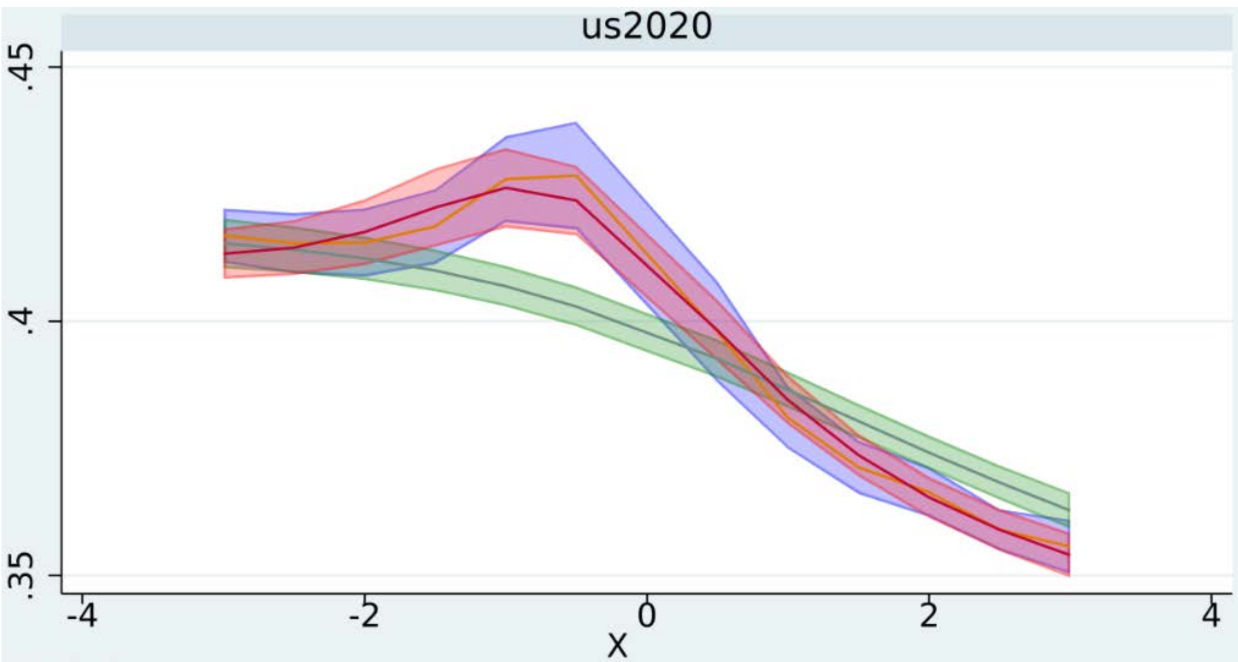
	la	si	pi	ka
la	1.0000			
si	-0.2052	1.0000		
pi	-0.1712	0.0704	1.0000	
ka	0.1834	0.1409	-0.0084	1.0000

	a	b	p	q
a	1.0000			
b	-0.3183	1.0000		
p	-0.6246	-0.0960	1.0000	
q	-0.6248	0.6343	0.6505	1.0000

Empirical implementation: understanding the U.S. distribution and LaSiPiKa parameters

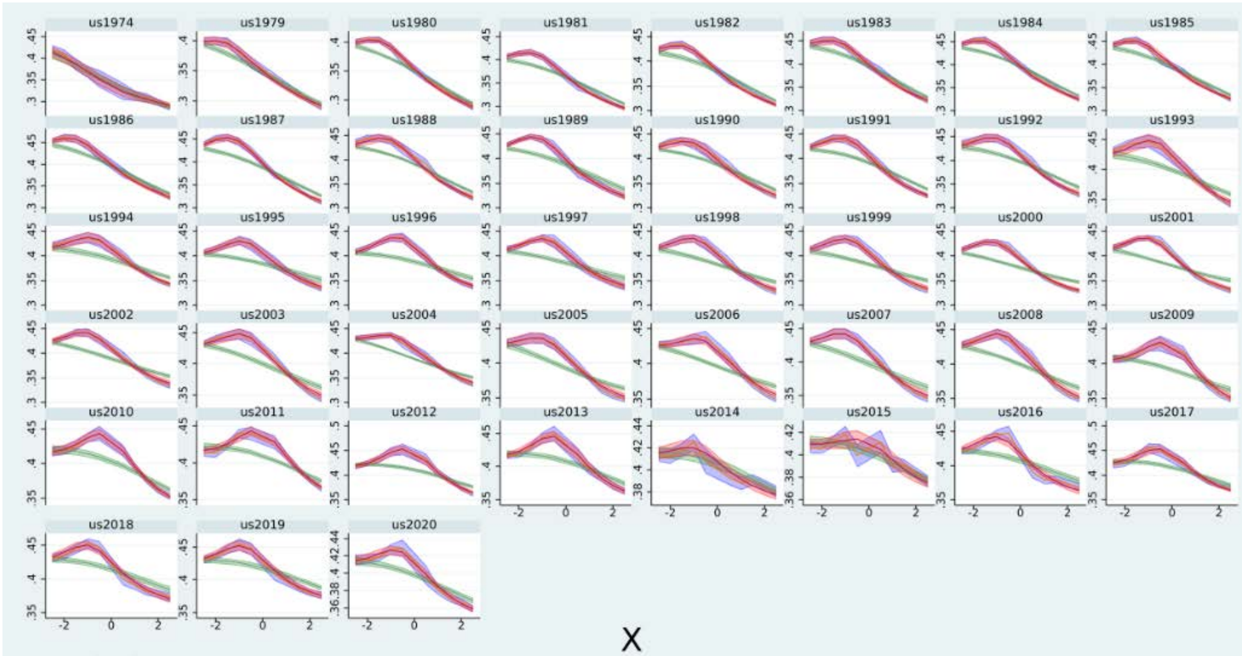
The U.S. is interesting for its specific morphology (fig 16). Like Belgium, it has a negative-sloped shape, with a hump near X=-1. The gb2 significantly misses the target for X between -1 and 0, showing a kappa of -0.53 and a pi of 0.04. The strongest difference is that the U.S.' level lambda is 10 percentage points above that of Belgium. The pertaining standard errors of estimates (see fig 10 above) confirm the significant difference to zero of sigma, pi, and kappa. This U.S. shape is a stable trait, at least after the end of the 1970s (fig 17). Apart from 1974, where the isograph was quasi-linear, and from potentially mistaken years (2014-2015), the general shape remains stable. If American inequality is relatively high for a developed country, the most specific trait is extreme poverty in the lowest deciles, particularly at level X=-1, the lowest quartile threshold. The U.S: isograph confirms the existence of the "other America", specific for its extreme poverty.

Figure 16: U.S. 2020 empirical isograph (blue with 95% CI), gb2 (green), and LaSiPiKa (red)



job 1012123 submitted Thursday 3 November 2022 at 08:53

Figure 17: U.S. 1974-2020 empirical (blue with 95% CI), gb2 (green) and LaSiPiKa (red)



job 1018542 submitted Thursday 24 November 2022 at 13:02 https://www.louischauvel.org/job_1018542.txt

A closer analysis of the parameters LaSiPiKa (fig 18) shows some relevant trends: lambda accelerated in the late 1970s and then increased with a regular trend from 2000 onwards, meaning that overall inequality has been increasing in the U.S. Sigma has always been negative but gets closer to zero over time. The isograph, with a very negative slope in the past, becomes flatter because inequality has increased faster at the upper level than at the bottom.

Polarization parameter pi declined moderately, apart from during the outlier years 2014-2015, but remain significant, and the epicenter of this polarization (or hump) measured by kappa was below X=-1 in the early 1980s and is now closer to X=-.5. This confirms polarization is not necessarily centered on the median X=0.

The series, U.S. 1979-2020, confirms the relative stability and significance of the estimated coefficients that express intelligible inequality transformations across decades. Another important feature is the American (relative) exceptionalism, with strong negative-shaped isograph: compared to the “relative” equality at the top of the distribution with a Gini of 0.373 (i.e. 5 percentage points below Mexico which is at 0.423), us2020 comes, with a relative poverty rate of 15.8%, not far away from the Mexican one that is at 16.3%: The U.S. is more exceptional for its income-poor than for its income-rich, because inequality above the median is less extreme for income not wealth.

Figure 18: U.S. 1979-2020 LaSiPiKa coefficients

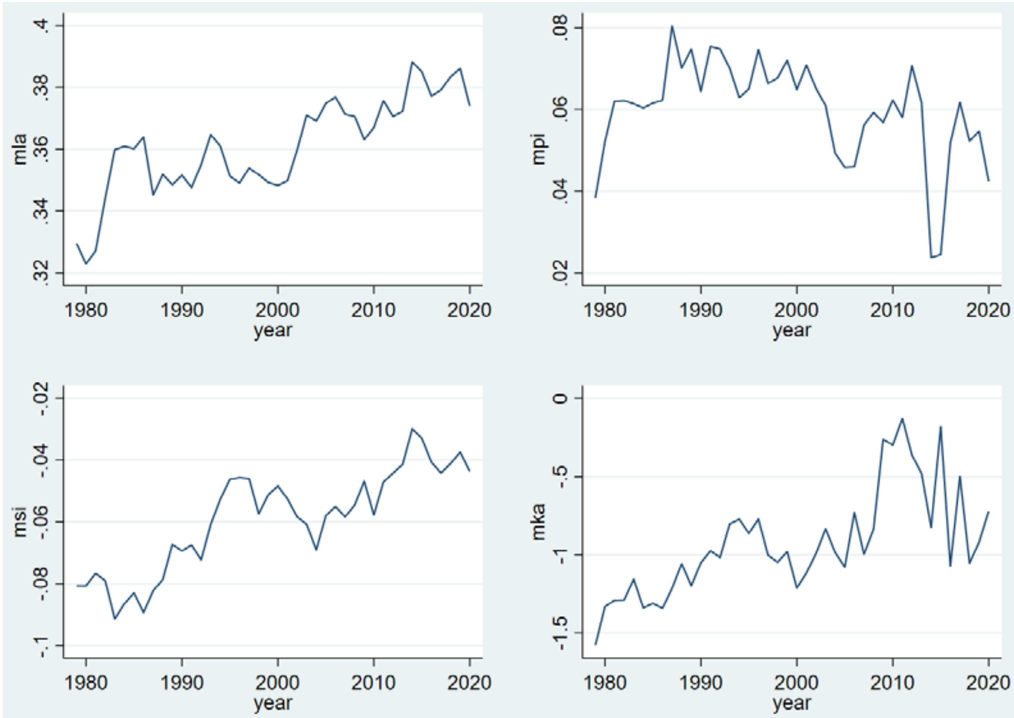
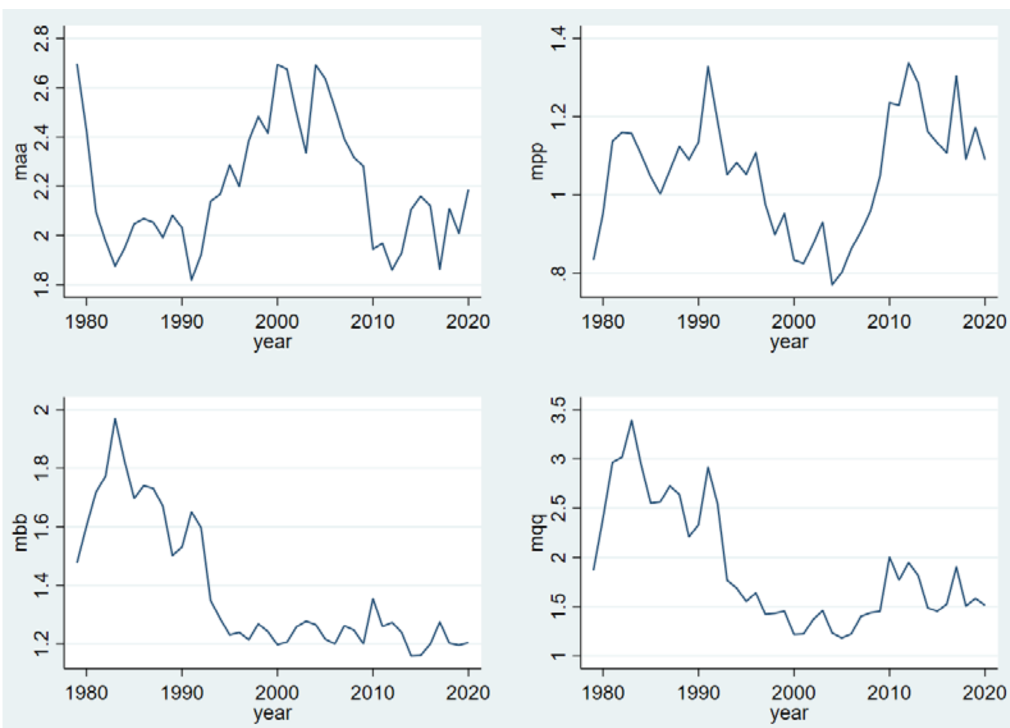


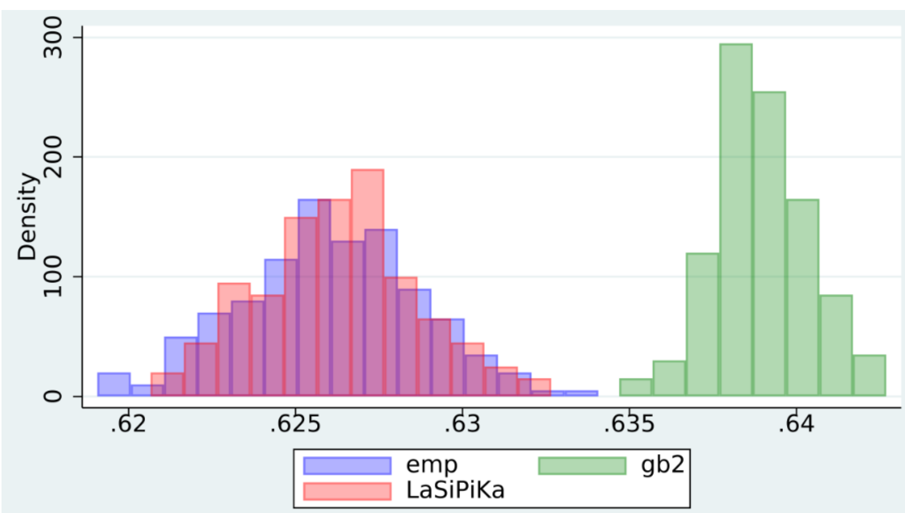
Figure 19: U.S. 1979-2020 gb2(aa, bb, pp, qq) parameters



The intelligible coefficients lambda, sigma, pi and kappa provide these elements of diagnosis. Regarding the gb2 parameters (aa, bb, pp, qq), interpretation of American transformations become more difficult (fig 17). The first parameter aa shows relative stability, and bb, that is, a scale parameter, expresses a decline and then stabilization after 1995; the meaning of which is not evident. Parameter pp is very negatively correlated with parameter aa (r=-0.94), and parameter qq is strongly and positively related to bb (r=+0.96). It is difficult to be conclusive: gb2 estimated parameters are not independent; the maximum of correlation in LaSiPiKa parameters is between sigma and kappa (r=+0.66), meaning a stronger independence between the different parameters. But the attempt to understand something from the gb2 variations might be limited: the estimated isographs of gb2 are missing, often significantly, the empirical isographs of the U.S., far outside the confidence intervals, as the LaSiPiKa isographs generally make a much better estimation.

Looking closer at the LaSiPiKa and gb2 estimated isographs (in comparison to the empirical isographs), we observe a systematic lack of accuracy between X=-1 and X=-0.5 because the gb2 makes a straight line estimate of a sector of the isograph where we detect a significant hump (fig 15). For the rest, the different isographs tend to overlap. When one focuses on the percentile p=25 (first quartile level, X=-1.1) of the medianized distribution, with 200 bootstrapped estimates, income levels are close to 0.625 times the median income in LaSiPiKa estimation and in the empirical data (bootstrapped confidence intervals overlap); in the gb2 estimates, the first quartile income levels are closer to 0.64 times the median, with clear separation of the confidence intervals. Even if the absolute gap of 1.5 percentage points in empirical and gb2 estimated incomes is not massive, it is sufficiently systematic to make a real difference in accuracy of the estimates. At first quartile, gb2 overestimates the income level (i.e. underestimates inequality as measured on the isograph): in the empirical distribution, the empirical first quartile misses by 2.5% (relative gap) the income estimated by gb2. This distortion is rather modest, but significant enough to make a difference in the shape, so that one should prefer LaSiPiKa to gb2.

Figure 20: U.S. 2020 first quartile level (empirical and estimates, 200 bootstraps) in terms of fraction of the median income empirical (blue), LaSiPiKa (red), and gb2 (green) isographs



job 1012133 submitted Thursday 3 November 2022 at 10:25

At large, gb2, that has relevant theoretical foundations, shows empirical difficulties: its parameters are not all interpretable, not independent, underlining problems of parsimony, and the isograph shows its shape is not flexible enough to fit empirical cases. LaSiPiKa provides the better solution.

Other national cases and parameters

Other examples of countries with longer trends are available in Annex 2. They confirm that each country has a specific, relatively stable empirical and LaSiPiKa isograph profile. The profile of France may be the flatter example, close to a Fisk distribution; Germany is not too different but with stronger, somewhat complicated fluctuations and is well fitted by LaSiPiKa. The profile of Chile is a LaSi positive sigma distribution, and Canada has a LaSi with negative sigma and increasing pi over time. The cases of Israel and the UK underline episodes of strong polarization that evolved over time. A general analysis of LaSiPiKa parameters should be developed to determine the domain of variation of the different parameters and to better understand the cases where LaSiPiKa isographs significantly differ from the empirical ones.

See Annex 2 for other national cases and parameters. Looking at these, one can notice:

- France’s flat (Fisk style) distribution
- Germany’s structure that is equal at the top and less at the bottom
- U.K.’s expansion of inequality (and polarization near to the median) in the 1980s followed by a progressive reduction of inequalities
- Canada’s similarity to the U.S. in terms of shape, but .05 lower levels on the isograph
- Chile’s specific positive slope quasi-linear shape with extreme inequalities at the top and less at the bottom.
- Israel’s polarized structure that is closer to the U.S.’s shape in recent years
- Sweden’s variations from 1967
- Denmark’s very specific, highly polarized approximate X=-1, structure
- Norway’s quasi-linear negative slope shape that is difficult to fit perfectly through LaSiPiKa
- Columbia’s highly (i.e. the most) complicated structure

The LaSiPiKa quantile function provides the possibility to graph predicted densities (See Annex 3) depicting the diversity of country distributions.

Designing complicated isographs

In cases like co and no, the LaSiPiKa significantly misses the target. The first reason for this is the size of the samples in these countries: large samples come with smaller confidence intervals. The second reason derives from the obvious complexity in the isograph that differs from a simple quasi-linear shape completed by one hump term.

Solutions to improve the interpolation of these distributions consist of:

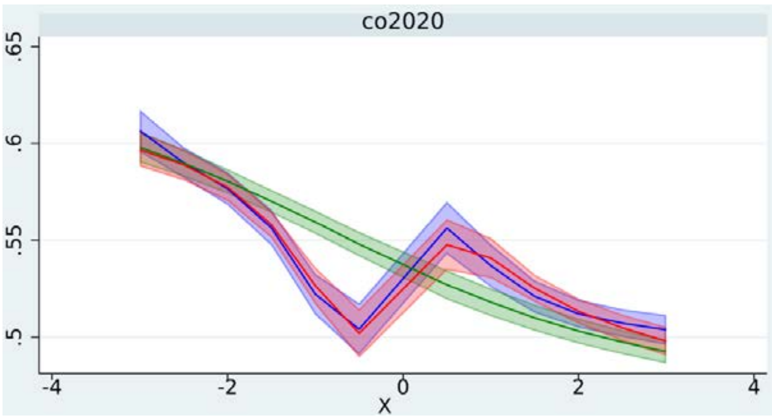
- Manually improve the mu parameter in the hump term PiKaMu [pi/cosh(mu(X-kappa))] that measures the wavelength of the bump, simplified as mu=1 in the usual LaSiPiKa
- Multiply the number of humps

Or, even a combination of the two. This means the possibility to handmade humps composition to improve the adjustment of a LaSiPiKa.

The solution for co2020 consists in a double hump, shortwaved (mu=2), one positive and one negative, centered on X=-0.5 and X=0.5 respectively.

$y=b\ \exp(X\ (\lambda + 4\ \sigma\ \tanh(X/4) + pi_1/\cosh(2X-1) + pi_2/\cosh(2X+1)))$

Figure 21: Colombia 2020 empirical (blue), LaSiPiKa (red), and gb2 (green) isographs



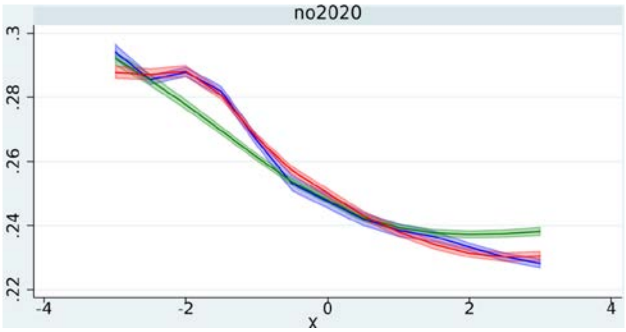
job 1014320 submitted Saturday 12 November 2022 at 09:09

```
nl(ISOm={la}+4*tanh(contX/4)*{si}+{pi1}*(1/cosh(2*X+1)) +{pi2}*(1/cosh(2*X-1)))
      Source |       SS            df       MS
-------------+----------------------------------    Number of obs =     47,248
       Model |  33.577359           3  11.1924531    R-squared     =     0.9548
    Residual |  1.5882146       47244  .000033617    Adj R-squared =     0.9548
-------------+----------------------------------    Root MSE      =    .005798
       Total |  35.165574       47247  .000744292    Res. dev.     =  -352596.6
------------------------------------------------------------------------------
        ISOm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         /la |   .5415547   .0000401 13496.39   0.000      .541476    .5416333
         /si |  -.0177952   .0000182  -977.34   0.000    -.0178309   -.0177595
        /pi1 |  -.0539501   .0000984  -548.45   0.000    -.0541429   -.0537573
        /pi2 |    .029236   .0000896   326.16   0.000     .0290603    .0294117
------------------------------------------------------------------------------
```

The case of Norway 2017 is solved with a similar strategy to module wavelength and epicenters of three different humps: $y=b\ \exp(X\ (\lambda + 4\ \sigma\ \tanh(X/4) + pi_1/\cosh(2(X+1.8)) + pi_3/\cosh((X-2.5)/4)))$

```
nl(ISOm={la}+4*tanh(X/4)*{si}+{pi1}*(1/((X+1.8)*2)) +{pi2}*(1/ cosh((contX-2.5)/4)))
      Source |       SS            df       MS
-------------+----------------------------------    Number of obs =     97,525
       Model |  57.775054           3  19.2583512    R-squared     =     0.9907
    Residual |  .54389662       97521  5.5772e-06    Adj R-squared =     0.9907
-------------+----------------------------------    Root MSE      =   .0023616
       Total |   58.31895       97524  .000597996    Res. dev.     =  -902982.3
------------------------------------------------------------------------------
        ISOm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         /la |   .3269872   .0001215  2692.31   0.000     .3267491    .3272252
         /si |   -.001345   .0000179   -75.19   0.000    -.0013801   -.0013099
        /pi1 |   .0186713   .0000282   661.80   0.000      .018616    .0187266
        /pi2 |  -.0952321   .0001545  -616.53   0.000    -.0955348   -.0949293
------------------------------------------------------------------------------
```

Figure 22: Norway 2020 empirical (blue) LaSiPiKa (red) and gb2 (green) isographs



job 1014706 submitted Sunday 13 November 2022 at 18:06

Potential further implementations of isograph & LaSiPiKa

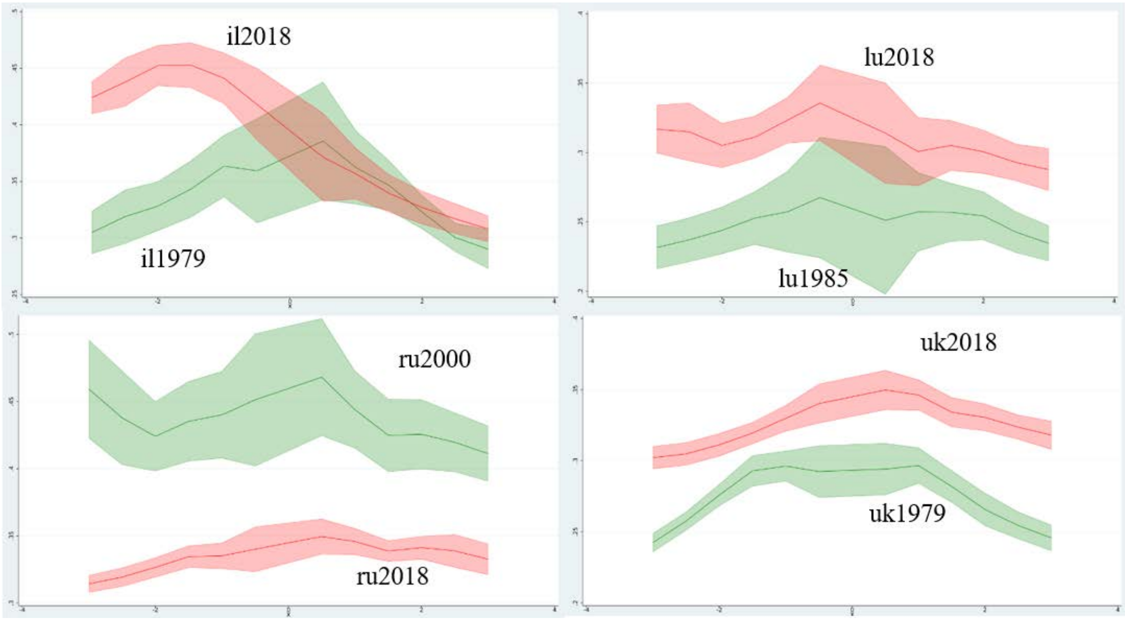
Comparing times

All in all, the isograph and LaSiPiKa are powerful tools to understand inequality change. From 1980 to 2020 (fig 23), the change in the American inequality regime affected the percentiles on the median and above: this explains the sharp increase in the Gini index (from 0.31 to 0.37) as the relative poverty rate remained almost unchanged (from 15% to 16%). With a strong increase in ISO(0), at the median, we confirm the stretching median class in the U.S. (Gornick and Jantti 2013).

Figure 23: U.S. 1980-2020 change in isograph (95% CI)

job 1035863 submitted Sunday 22 January 2023 at 18:52 https://www.louischauvel.org/job_1035863.txt

Figure 24: Change in isograph in Israel, Luxembourg, Russia and the U.K. (95% CI)

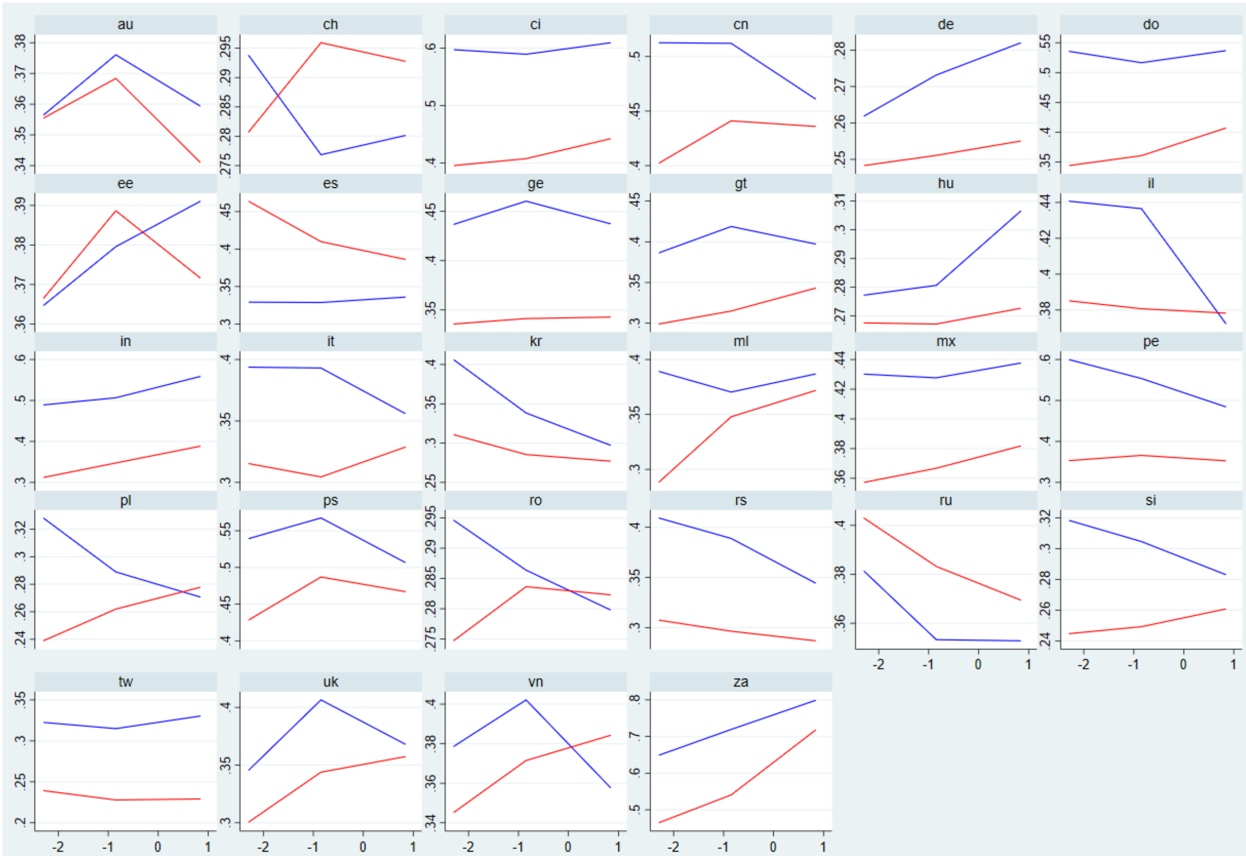


Similar implementations on Luxembourg, Israel, the U.K. or Russian data (fig.23) of the same tools show the diversity potential changes. This approach provides opportunities to continue and develop earlier estimations on the same countries: the explosion of poverty in Israel (Bleikh 2016), the significant increase of income tails in Luxembourg (Allegrezza et al 2004), the LIS data decline in inequality at large in Russia (Hlasny, 2022) or in the increase in U.K. inequality (Jenkins 2009).

Comparing income and consumption

The same way, the isograph clarifies the complex nexus between income inequality and consumption inequality, with obviously diverging patterns between countries (fig 25). In a country like Switzerland (ch) the two facets of inequality are the same, we notice a general pattern of stronger income inequality than consumption inequality (the blue isograph for income is generally higher that he red one for consumption), with a gap that culminates in Ivory Coast (ci), but one should consider exceptions like Spain and Russia, where consumption inequality is higher, potentially due to underreported incomes of richer households. Anyway, in many countries cn gt il lt kr ml pe pl ps ro rs si uk vn za, this gap is smaller for the rich than for the poor, a pattern that makes sense if the poorest make use of other resources (auto production, kins' and neighbors' solidarity, or even debt). This result means the link between consumption and income differs from one country to another and from a percentile to another: a one-size-fit-all conversion from consumption inequality to income inequality is an imprudent choice.

Figure 25: Equivalized income (blue) and consumption (red) isographs in 28 countries (latest available year, rescaled y axis)



Job_1032986

Comparing policies

The isograph may be implemented in order to compare before/after redistributive social and tax policies. The isograph confirms:

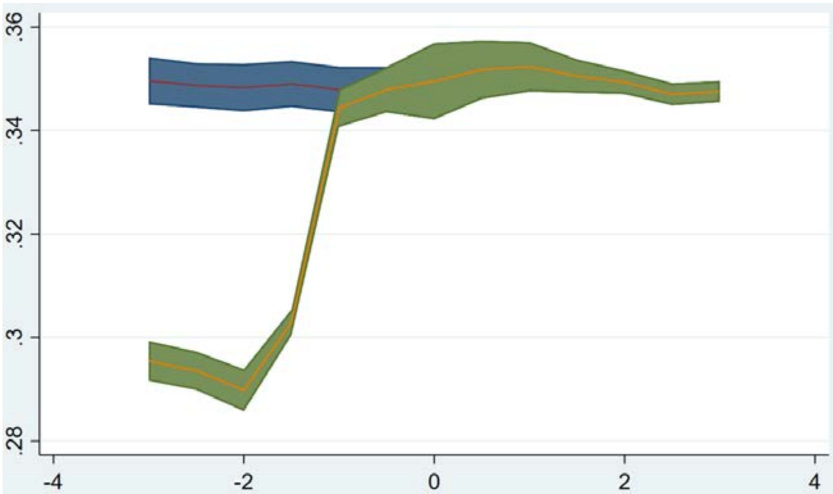
* a flat percentage tax (or a proportional to income redistribution of a public loan) has no effect in isograph

* a uniform decline of the isograph by 1 percentage point (f.ex. from a Gini index of g_1=.31 for distribution y1 to g_2=.30 for y2) means a progressive redistribution profile defined by: y2= y1 exp((g_2- g_1)X)

* a change in the slope (resp. polarization) supposes a second (resp. third) degree polynomial redistribution profile

* any other redistributive profile can be analyzed through observation or theoretical design. For instance, on an initial FC distribution of Gini=.35, a (not necessarily wise) redistribution of an amount of 5% the median to individuals below 50% of the median, and paid by a flat rate tax on the rest of the population, can be easily modeled (fig 26).

Figure 26: Before/after isograph of a gift of 5% the median income to the poor (50% below the median) from an initial Fisk(.35) distribution



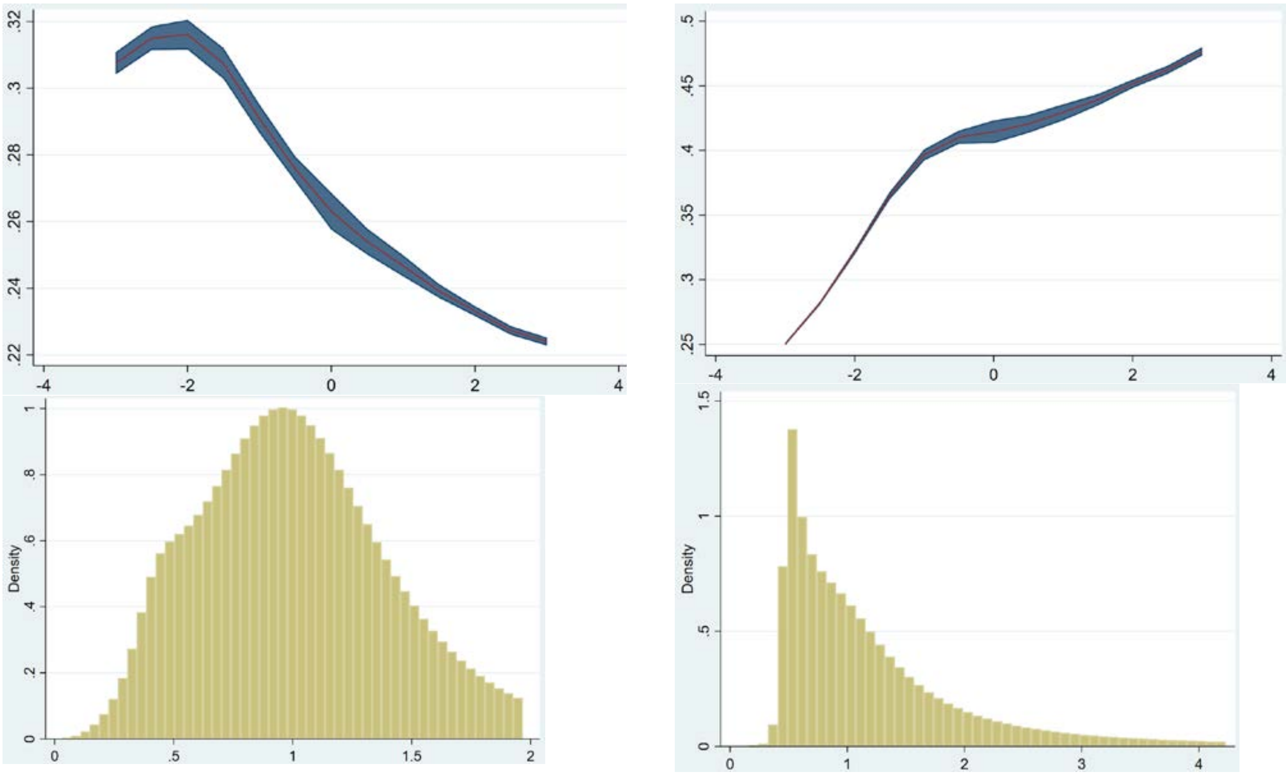
Thinking typical distributions, utopias and dystopias

Distributions are complex, and involve the balance between (at least) three groups: the rich, the poor and the median class. From a middle-class point of view, the best society for its own stability would be low isograph at the median (for homogeneity), high for the almost-poor (in order to buy their service for cheap) but no extreme poverty (so low isograph at the far poor, for social stability) and low isograph for the rich (so that they are not too far above). With parameters lambda=.25, sigma=-.01, pi=.05, and kappa=-2, we have a typical Nordic country distribution.

Conversely, a middle class nightmare would be high isograph, increasing at the top, and low at the bottom, with may be a strong polarization at level X=-.5 for a loss of homogeneity and deep risks of socioeconomic below the median in case of decline in socioeconomic rank. This distribution may be obtained with parameters: lambda=.35, sigma=.05, pi=.1, kappa=-1.

As such, the isograph and LaSiPiKa provide tools to understand the complexity of social justice between three classes of incomes: the median class, the rich and the poor, characterized by different interests.

Figure 27: two societies, one propitious to the middle class (left) versus unpropitious (right) with their isograph (above) and density (below)



Conclusion

With the association of the isograph, to understand the potential complexity of empirical income distributions, with the LaSiPiKa distribution that provides interpretable parameters of inequality, a novel strategy of distributional analyses is proposed. The gb2, with its four difficult to interpret parameters, is a good approximation of certain cases, but then the LaSi does almost the same with only three intelligible parameters: scale, level of overall inequality, and slope, that is, the balance between inequalities above and below the median. The LaSi parameters have clear correspondences with the Gini coefficient and the relative poverty rate. When gb2 misses the target, LaSi does the same, but LaSiPiKa is able to model specific polarization of intensity pi and epicenter kappa.

As a consequence, even if gb2 is useful in certain cases, LaSi versus LaSiPiKa appear also as a more parsimonious versus more flexible set of solutions with interpretable parameters. Additionally, the isograph provides a visualization tool, able to provide diagnoses on the complexity of empirical distribution and accuracy of their fits.

Isograph and LaSiPiKa novel technology confirm the morphology of inequality is not a universal constant: the general level of inequality (measured by the Gini index, D9/D1 or lambda, or whatever) hides other patterns like the variation of inequality for the rich and the poor (measured by sigma) and polarization, measured by its intensity pi and epicenter kappa. Those parameters have correspondence with traditional measurements and indicators of inequality. In the general case, the LaSiPiKa distribution is generally sufficient to model many empirical samples. When this LaSiPiKa fit is not sufficient, through a diagnosis provided by the isograph on the gap between the empirical distribution and the predicted one, LaSiPiKa can be complexified through a handmade design of PiKaMu hump terms, where mu eventually improves the wavelength of the polarization. Any kind of distribution can be hand-tailored on demand so that empirical and modeled isographs coincide. The handmade, realistic ISO function is then the base of a quantile function of y=b.exp(ISO.X) that will help to simulate distributions. In the same way, density functions may be systematically derived and compared (see Annex 3 reporting derived densities).

As such, Isograph and LaSiPiKa are innovative tools providing novel, consistent measurements of inequality useful to compare distributions across time and nations. The fields of implementation of these tools are very large: long-term changes in the variation of inequalities, analysis of transfers (taxes and social redistributions), or even comparison of income/consumption inequalities (fig 25), are only some of the potential domains where fine grain comparison of inequalities below, above, or near the median, are essential to understand the comparative morphology of inequalities.

References

Allegrezza, Serge, Georges Heinrich, David Jesuit, Poverty and income inequality in Luxembourg and the Grande Région in comparative perspective, Socio-Economic Review, 2:2, pp. 263–283, https://doi.org/10.1093/soceco/2.2.263

Atkinson, AB (1970) On the measurement of inequality. Journal of Economic Theory, 2 (3), pp. 244–263, doi:10.1016/0022-0531(70)90039-6

Atkinson, Anthony B., and Francois Bourguignon. 2000/2015. Handbook of Income Distribution. Amsterdam: Elsevier, vol. 1&2.

Bourguignon, F. (1979). Decomposable Income Inequality Measures. Econometrica, 47(4), 901–920. https://doi.org/10.2307/1914138

Bleikh, Haim (2016) : Poverty and Inequality in Israel: An International Perspective, LIS Working Paper Series, No. 661, Luxembourg Income Study (LIS), Luxembourg

Chauvel, L. (2016). The intensity and shape of inequality: the ABG method of distributional analysis. Review of Income and Wealth, 62(1), 52-68.

Chauvel, L., Hartung, A., Bar-Haim, E. and Kerm, P.V. (2019), "Income and Wealth Above the Median: New Measurements and Results for Europe and the United States", Decancq, K. and Kerm, P.V. (Ed.) What Drives Inequality? (Research on Economic Inequality, Vol. 27), Emerald Publishing Limited, Bingley, pp. 89-104. https://doi.org/10.1108/S1049-258520190000027007

Copas, J. "The Effectiveness of Risk Scores: The Logit Rank Plot Quick View," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 48, 165–83, 1999.

Cowell, F.A. 2000. Measurement of inequality. In Handbook of Income Distribution Volume 1, eds A.B. Atkinson and F. Bourguignon. Amsterdam: Elsevier Science, 59-85.

Cowell, F., Flachaire, E.: Statistical Methods for Distribution Analysis. In: Atkinson, A.B., Bourguignon, F. (eds.) Handbook of Income Distribution, vol. 2. Elsevier, North-Holland, Amsterdam (2015)

Duclos, J.-Y., & Taptué, A.-M. (2015). Polarization. In A. B. Atkinson, & F. Bourguignon (Eds.), Handbook of income distribution (Vol. 2) (pp. 301–358). Amsterdam: Elsevier. [Google Scholar]OpenURL

Gini, C., Sulla misura della concentrazione e della variabilità dei caratteri. Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti, LXXIII, parte II, 73, 1203–48, 1914.

Gornick, J. And M. Jäntti (Eds.), 2013, Income Inequality: Economic Disparities And The Middle Class In Affluent Countries, Stanford, Ca, Stanford University Press.

Hlasny Vladimir, 2022, Household Earnings in Putin's Russia: Distributional Changes across Socioeconomic Groups, 2000–2016, LIS working papers series - No. 847.

Jann, B., 2005, kdens: Stata module for univariate kernel density estimation. Available on: http://ideas.repec.org/c/boc/bocode/s456410.html

Jann, B., 2007, "Univariate kernel density estimation," Boston College Department of Economics, Statistical Software Component No. S456410 (Cambridge, MA: Boston College).

Jenkins Stephen P., 1991. The measurement of income inequality. In L. Osberg (ed.) Economic Inequality and Poverty: International Perspectives. Armonk, NY: M.E. Sharpe.

Jenkins Stephen P., 1999. "INEQDECO: Stata module to calculate inequality indices with decomposition by subgroup," Statistical Software Components S366002, Boston College Department of Economics, revised 15 Feb 2021.

Jenkins Stephen P., 2009, "Distributionally-Sensitive Inequality Indices and the gb2 Income Distribution," Review of Income and Wealth, 55, 392–98, 2009.

Jenkins Stephen P., 2014. "gb2LFIT: Stata module to fit Generalized Beta of the Second Kind distribution by maximum likelihood (log parameter metric)," Statistical Software Components S457897, Boston College Department of Economics.

Jorda, Vanesa, Sarabia, José María and Jäntti, Markus, (2021), Inequality measurement with grouped data: Parametric and non-parametric methods, Journal of the Royal Statistical Society Series A, 184, issue 3, p. 964-984. https://doi.org/10.1111/rssa.12702

Kerman Sean C. & James B. McDonald (2015) Skewness-Kurtosis Bounds for EGB1, Egb2, and Special Cases, Communications in Statistics - Theory and Methods, 44:18, 3857-3864, DOI: 10.1080/03610926.2013.844255

Kleiber, C. and S. Kotz, *Statistical Size Distributions in Economics and Actuarial Sciences*, John Wiley, Hoboken, NJ, 2003.

Kot, S. M., & Paradowski, P. R. (2022). The atlas of inequality aversion: theory and empirical evidence on 55 countries from the Luxembourg Income Study database. Equilibrium. Quarterly Journal of Economics and Economic Policy, 17(2), 261–316. https://doi.org/10.24136/eq.2022.010 and Kot, S.M. and P.R. Paradowski (2022). The Atlas of Inequality Aversion, https://www.lisdatacenter.org/resources/other-databases/.

Luxembourg Income Study (LIS) Database. 2023. http://www.lisdatacenter.org (multiple countries; waves for 1967-2020, microdata runs completed in Sept 2022-Jan 2023), Luxembourg: LIS.

McDonald, J. B. and Y. Xu, "A Generalization of the Beta Distribution with Applications," *Journal of Econometrics*, 66, 133–52, 1995.

McDonald, J., "Some Generalized Functions for the Size Distribution of Income", *Econometrica*, 52, 647-63, 1984.

Sarabia, José María, Vanesa Jordá, Faustino Prieto, and Montserrat Guillén. 2021. "Multivariate Classes of gb2 Distributions with Applications" Mathematics 9, no. 1: 72. https://doi.org/10.3390/math9010072

Shoukri, M. M., I. U. M. Mian and D. S. Tracy, "Sampling Properties of Estimators of the Log-Logistic Distribution with Application to Canadian Precipitation Data," *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*, 16, 223–36, 1988.

Sitthiyot, T., Holasut, K. A simple method for measuring inequality. Palgrave Commun 6, 112 (2020). https://doi.org/10.1057/s41599-020-0484-6

Van Kerm Philippe, 2017. "_GRNDRAW: Stata module for random number generation from the gb2, Singh-Maddala, Dagum, Fisk and Pareto distributions," Statistical Software Components S458350, Boston College Department of Economics.

Van Kerm Philippe, 2020. "SGINI: Stata module to compute Generalized Gini and Concentration coefficients, Gini correlations and fractional ranks," Statistical Software Components S458778, Boston College Department of Economics.

Yitzhaki, S., "Stochastic Dominance, Mean Variance, and Gini's Mean Difference," *American Economic Review*, 72, 178–85, 1982.

Zheng, B. Almost Lorenz dominance. *Soc Choice Welf* **51,** 51–63 (2018). https://doi.org/10.1007/s00355-017-1106-0

Annexes:

Annex 1: sample sizes (53countries latest available year)

Annex 2: other national cases and parameters—empirical, LaSiPiKa, and gb2 isographs

Annex 3: derived densities

Annex 4: 655 countries parameters

Annex 5: isographs of 53 countries available in the LIS (most recent year)

Annex 6: An anthology of long-term national isographs

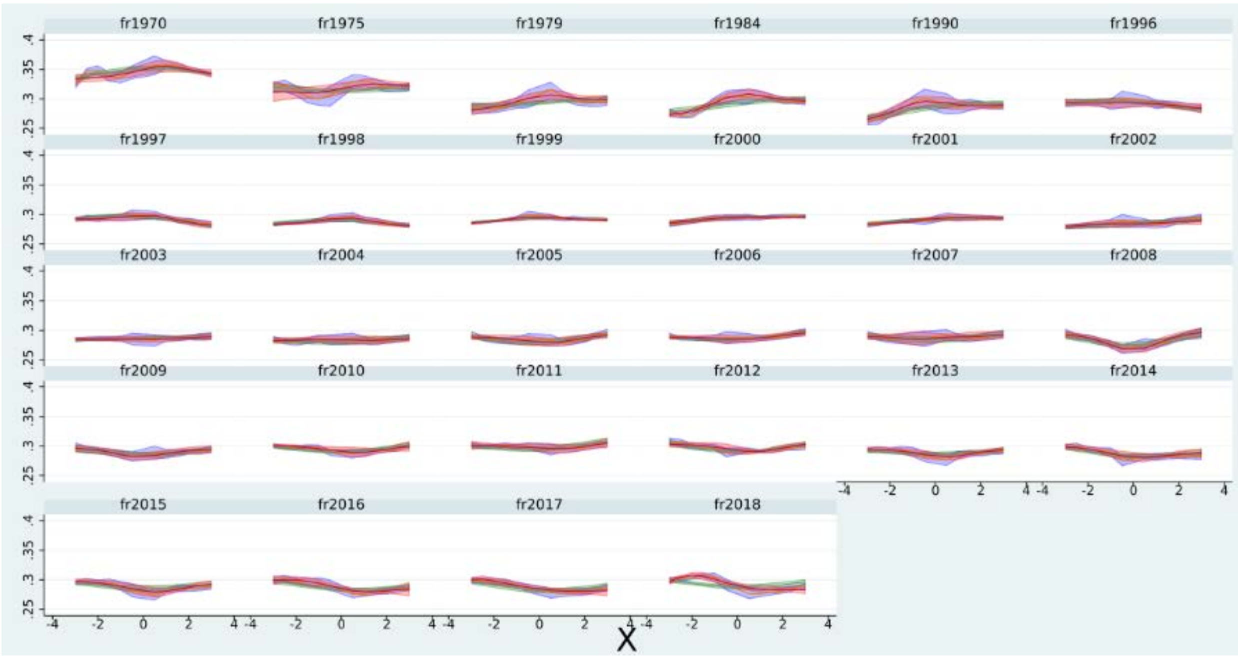
Annex 1: sample sizes (53countries latest available year)

Sample sizes (sorted by ISO code and by size)

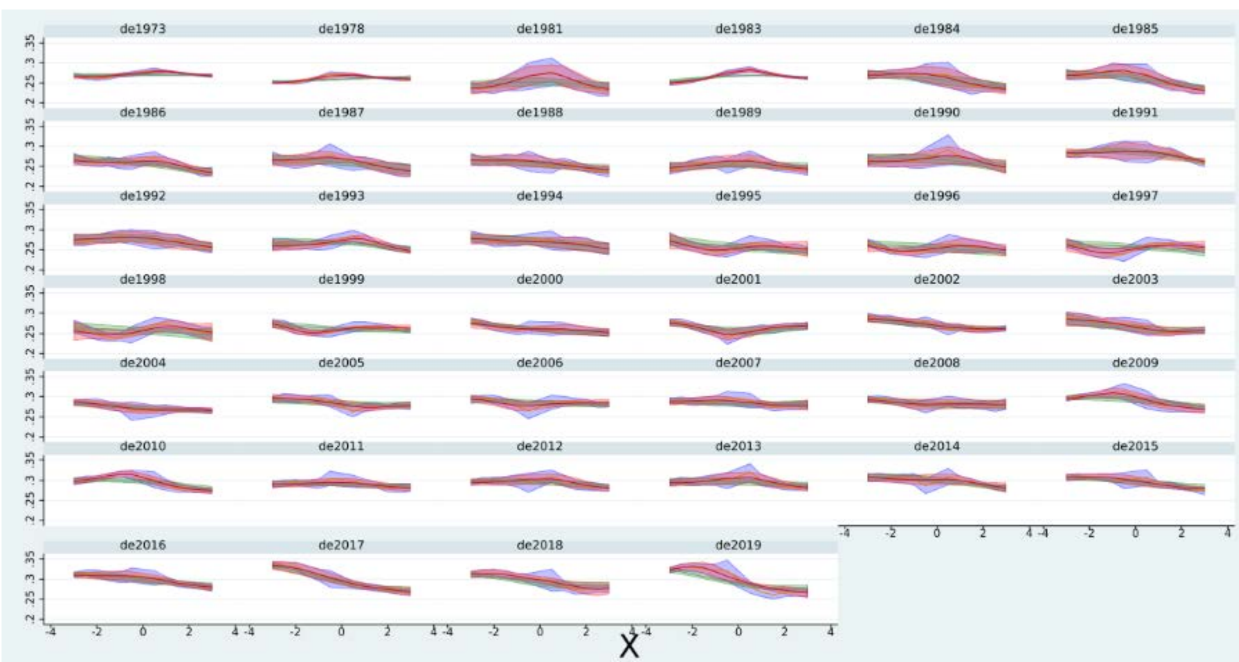
ISO	N	ISO	N
at	5,988	no	258,993
au	13,966	br	144,192
be	5,926	co	131,609
br	144,192	dk	89,098
ca	40,828	mx	74,429
ch	7,333	cl	70,574
ci	11,931	us	62,613
cl	70,574	ru	59,556
cn	20,307	fr	50,704
co	131,609	uy	42,459
cz	8,701	in	41,742
de	18,749	ca	40,828
dk	89,098	pe	33,968
do	8,321	pl	32,787
ee	6,137	ro	32,187
eg	11,838	gr	22,362
es	13,672	cn	20,307
fi	10,209	de	18,749
fr	50,704	tw	16,511
ge	3,175	se	16,252
gr	22,362	au	13,966
gt	11,512	nl	13,719
hu	2,771	es	13,672
ie	4,183	ci	11,931
il	5,954	eg	11,838
in	41,742	pa	11,541
is	3,015	gt	11,512
it	7,284	kr	11,373
jp	2,026	za	10,838
kr	11,373	fi	10,209
lt	5,101	uk	9,954
lu	2,735	vn	9,392
ml	6,685	cz	8,701
mx	74,429	do	8,321
nl	13,719	ch	7,333
no	258,993	it	7,284
pa	11,541	ml	6,685
pe	33,968	rs	6,362
pl	32,787	ee	6,137
ps	3,736	at	5,988
py	4,834	il	5,954
ro	32,187	be	5,926
rs	6,362	sk	5,586
ru	59,556	lt	5,101
se	16,252	py	4,834
si	3,750	ie	4,183
sk	5,586	si	3,750
tw	16,511	ps	3,736
uk	9,954	ge	3,175
us	62,613	is	3,015
uy	42,459	hu	2,771
vn	9,392	lu	2,735
za	10,838	jp	2,026

Annex 2: other national cases and parameters—empirical (blue), LaSiPiKa (red), and gb2 (green) isographs

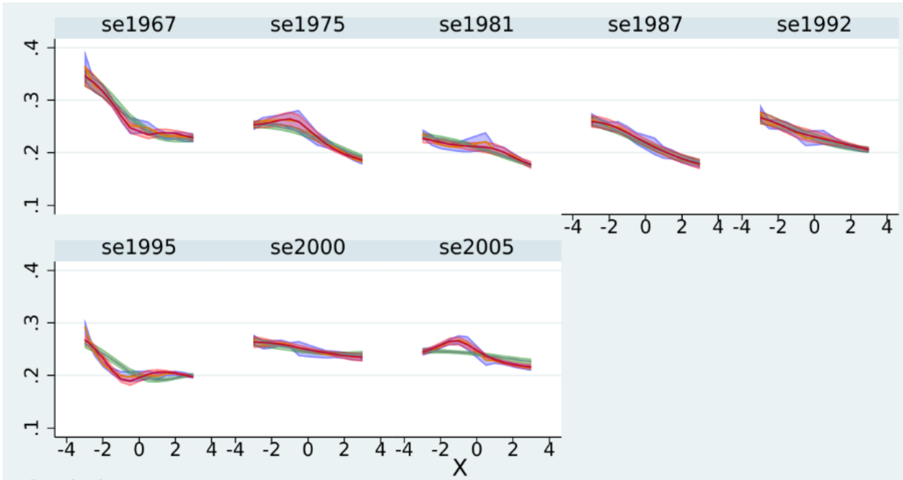
France 1970-2018



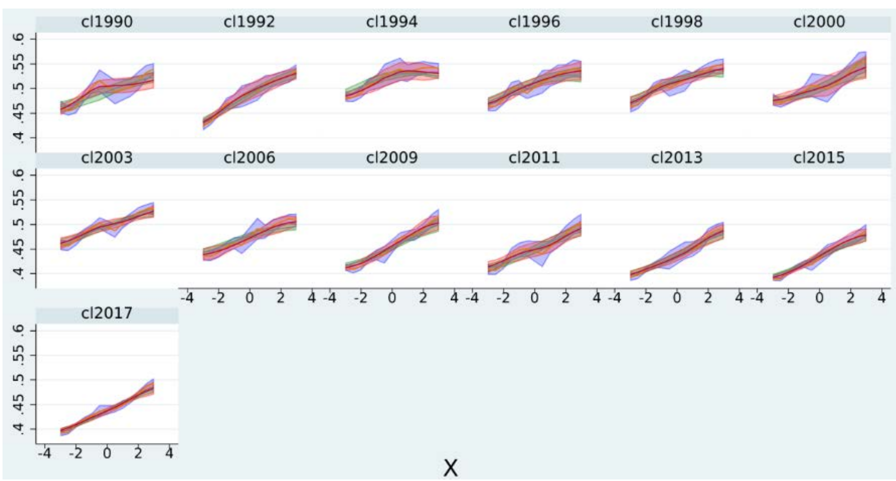
Germany 1973-2019



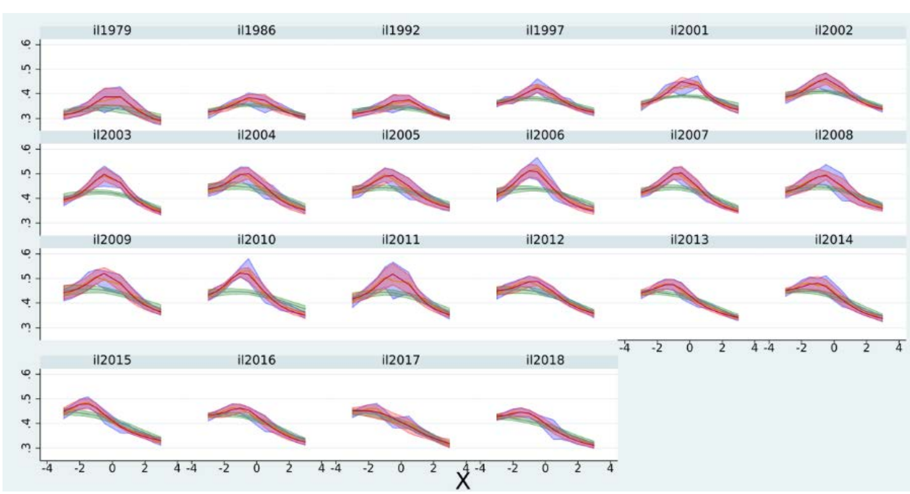
Sweden 1967-2005



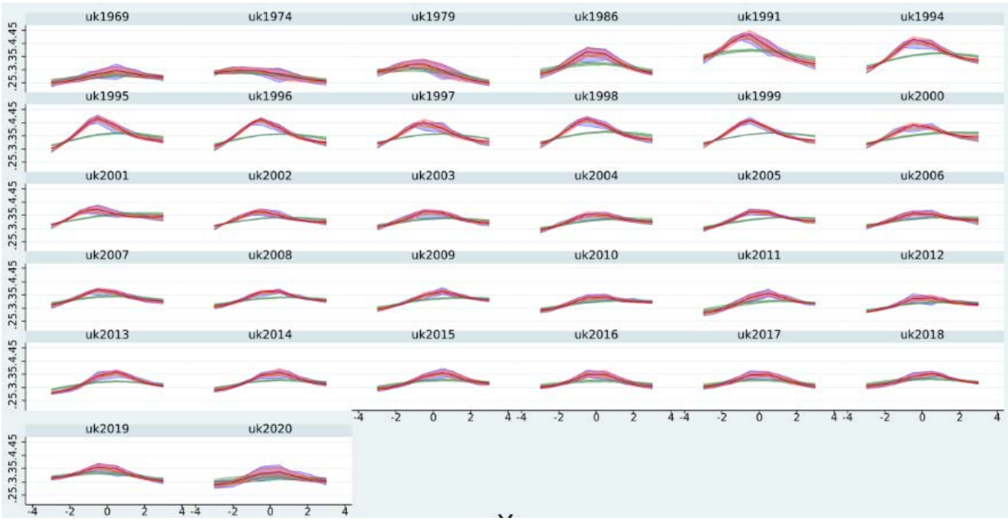
Chile 1990-2017



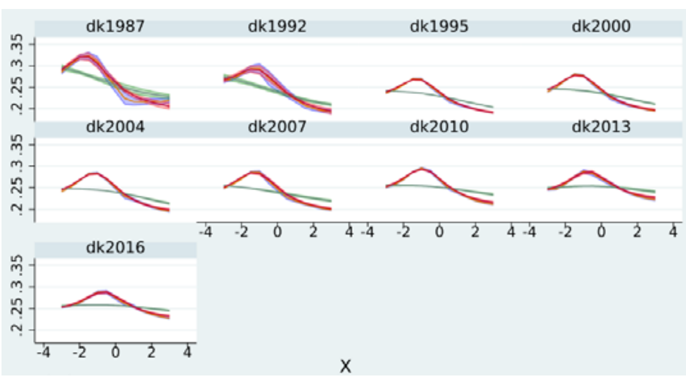
Israel 1979-2018



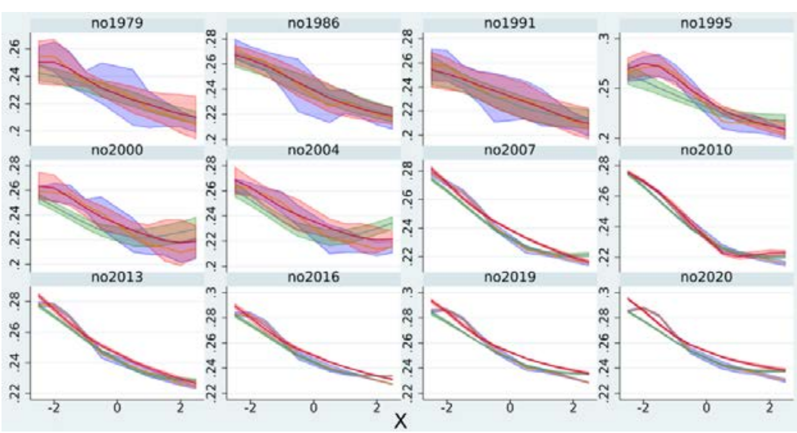
U.K. 1969-2020



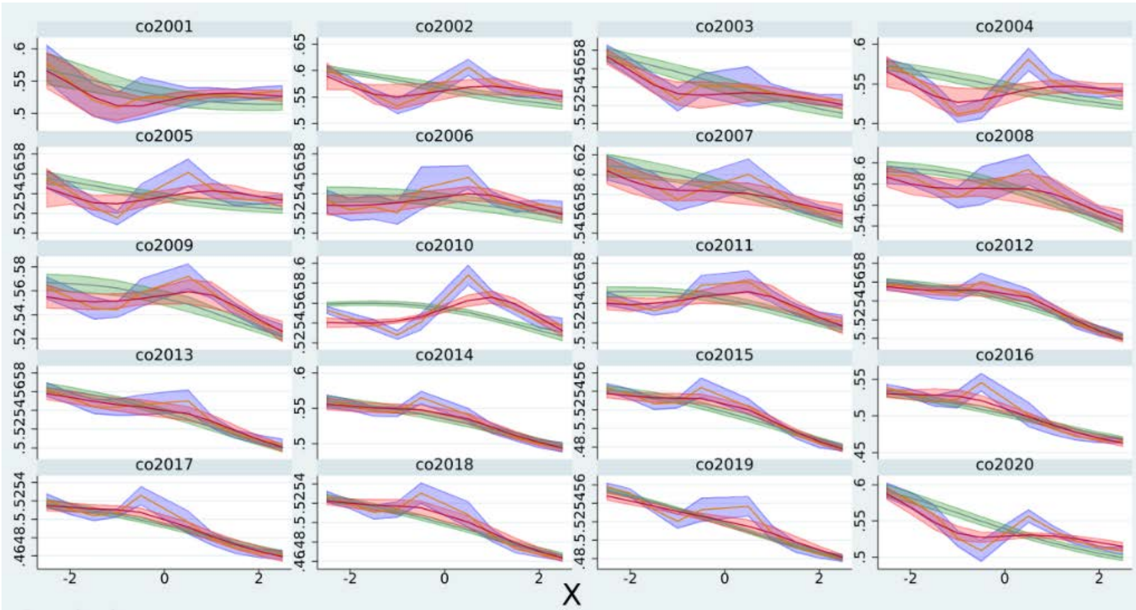
Denmark 1987-2016



Norway 1979-2020



Colombia 2001-2020



Annex 3: derived densities

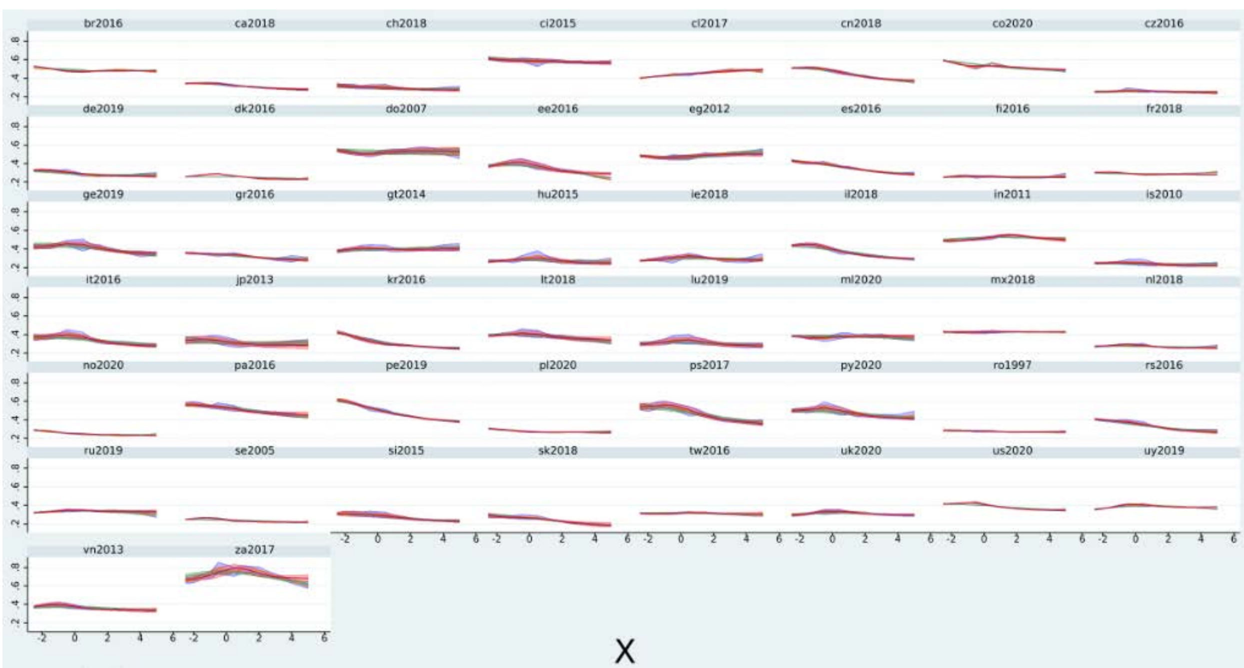
Annex 4: 655 countries parameters (three bootstraps…)

ccyyyy	mla	msi	mpi	mka	sdla	sdsi	sdpi	sdka
at1994	0.277671	-0.008209	0.018998	1.094757	0.009245	0.002050	0.006202	0.177075
at1995	0.261201	-0.006974	0.037458	0.587243	0.009058	0.001314	0.013983	0.217058
at1996	0.251549	-0.009669	0.055540	-0.298018	0.001983	0.001408	0.014082	0.404816
at1997	0.267612	-0.012053	0.020268	-0.174755	0.005743	0.002513	0.020106	0.702625
at1998	0.261584	-0.008415	0.033668	-0.370163	0.003733	0.001847	0.002458	0.397156
at1999	0.257185	-0.004453	0.024106	-0.063605	0.003397	0.001414	0.006597	0.480818
at2000	0.258746	-0.006123	-0.006596	-0.675233	0.009815	0.001672	0.017021	0.476631

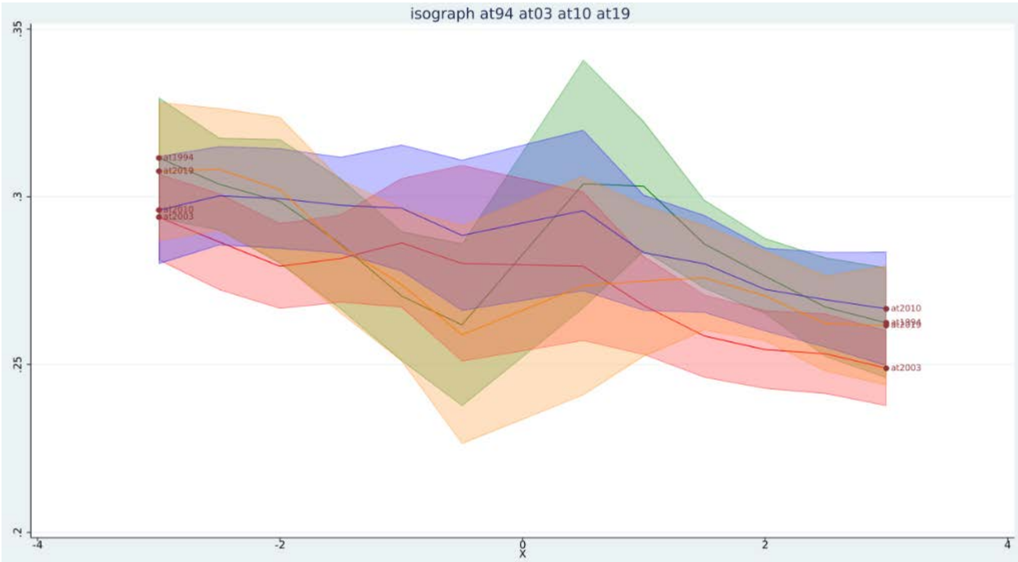
See the rest here https://www.louischauvel.org/LIS_655_lasipika_etc.xlsx

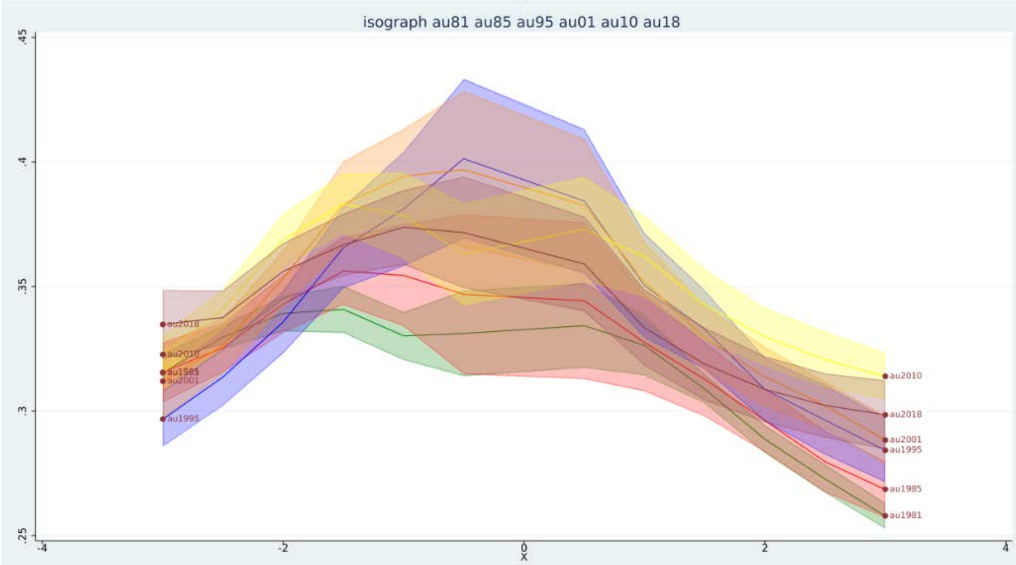
Annex 5: isographs of 53 countries available in the LIS (most recent year) (y scale common)

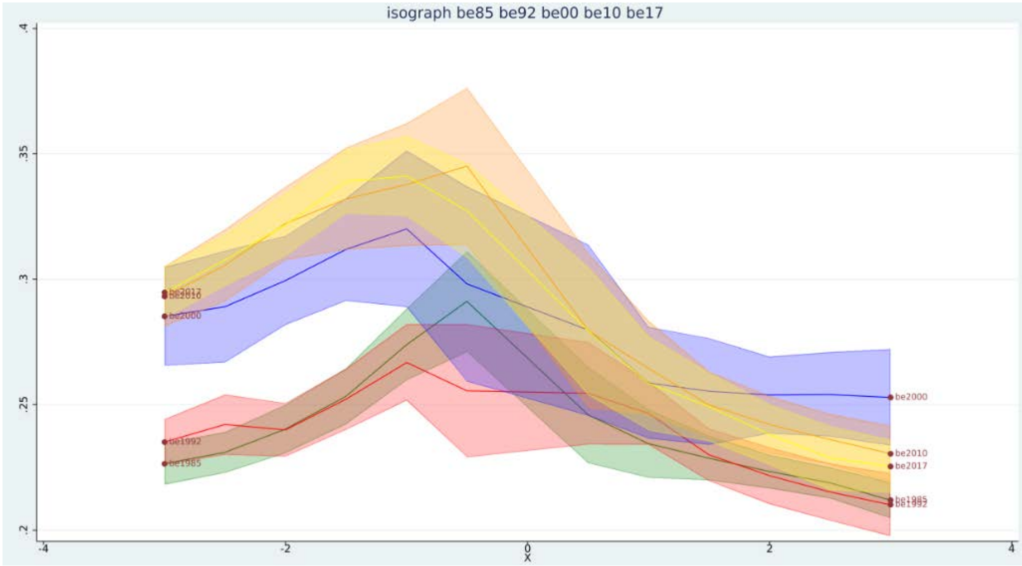
job 1012366 submitted Friday 4 November 2022 at 06:34

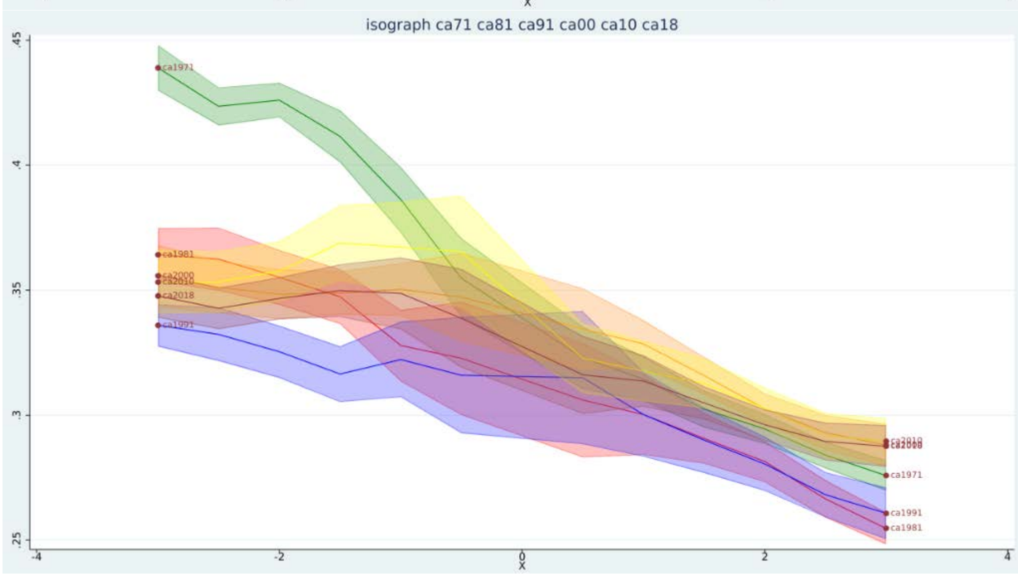


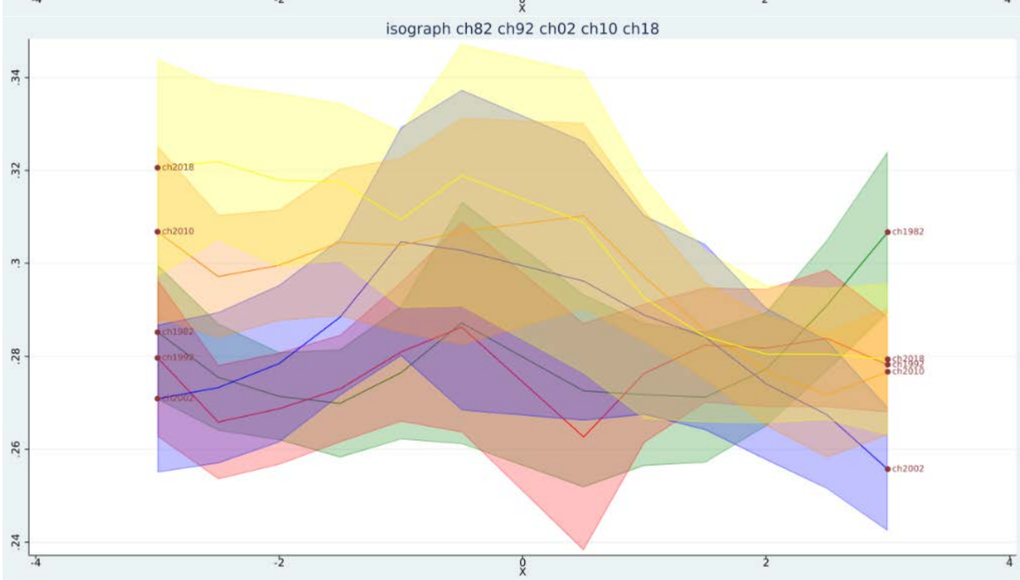
Annex 6: An anthology of long-term national isographs

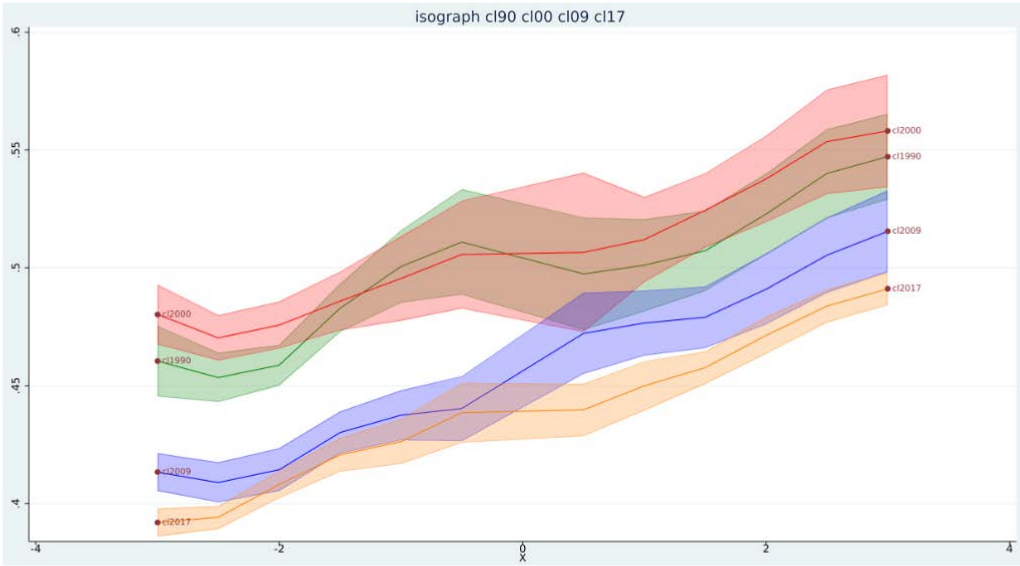


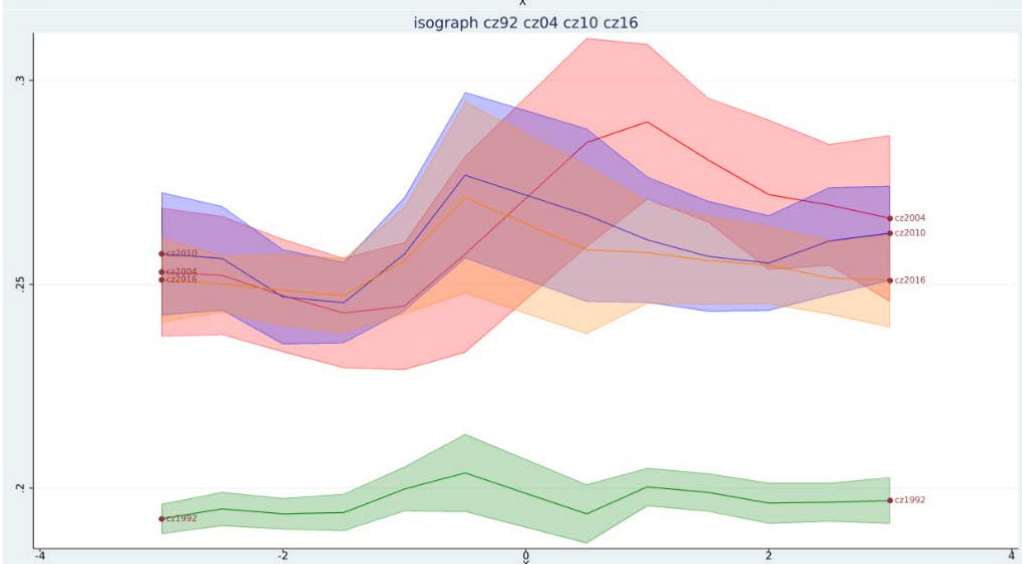


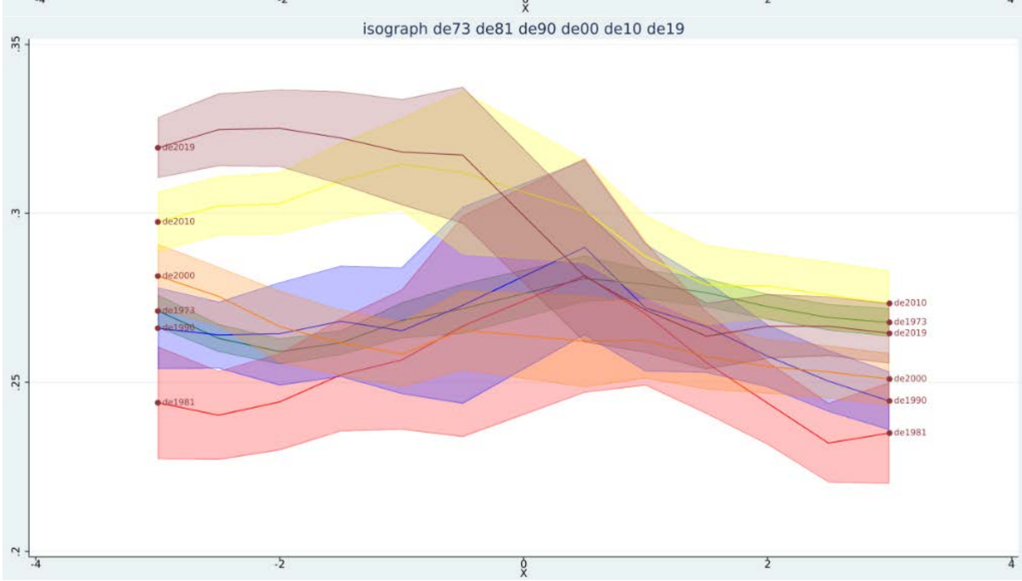

isograph be85 be92 be00 be10 be17
.2
.25
.3
.35
.4
-4
-2
0
2
4
x
be2000
be2010
be2017
be1985
be1992

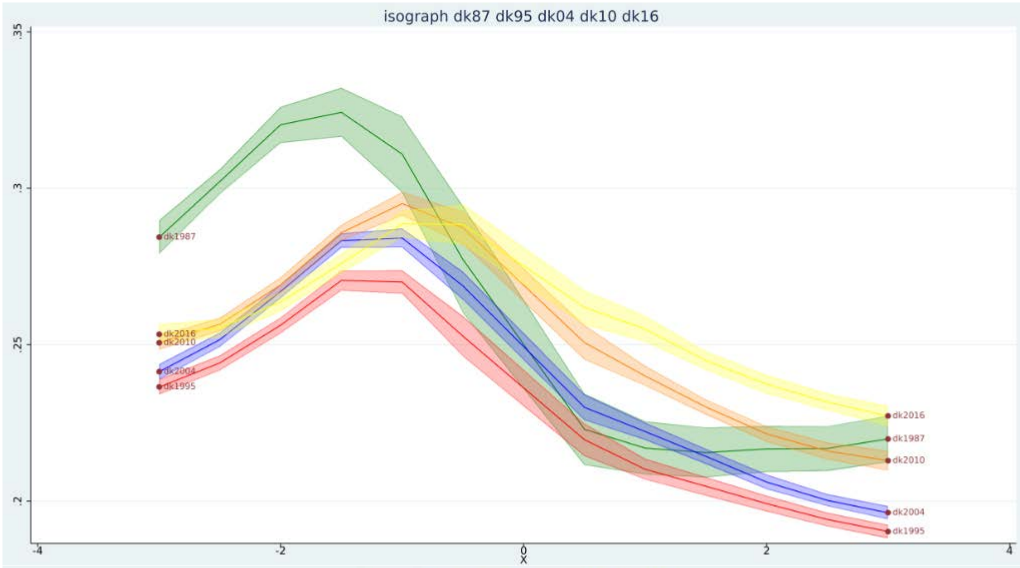

isograph ca71 ca81 ca91 ca00 ca10 ca18
.25
.3
.35
.4
.45
-4
-2
0
2
4
x
ca1971
ca1981
ca2018
ca1991
ca1971
ca1991
ca1981

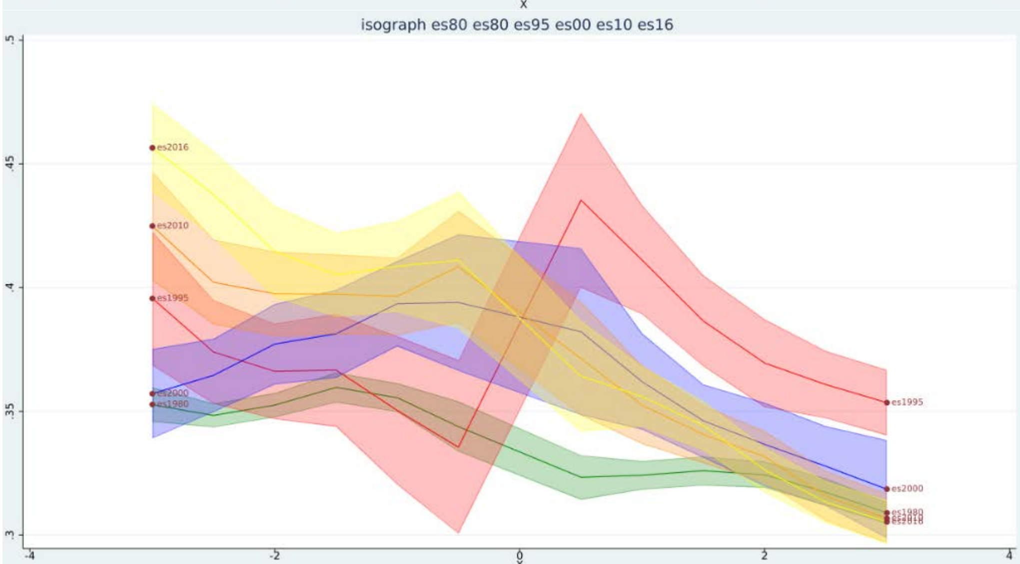

isograph ch82 ch92 ch02 ch10 ch18
.24
.26
.28
.3
.32
.34
-4
-2
0
2
4
x
ch2018
ch2010
ch1982
ch1992
ch1982
ch2010
ch2002

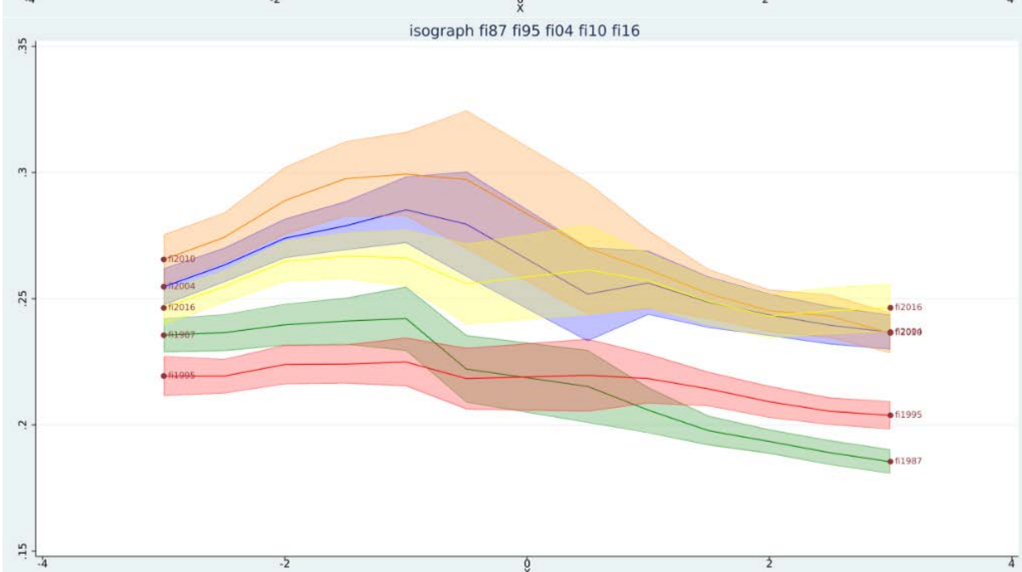

isograph cl90 cl00 cl09 cl17
cl2000
cl1990
cl2009
cl2017
x

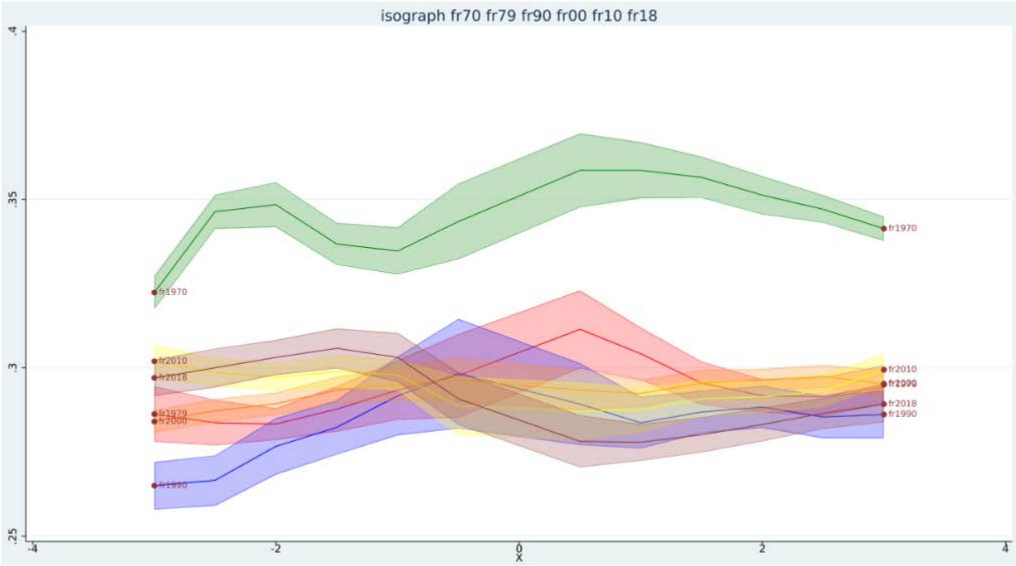

isograph cz92 cz04 cz10 cz16
cz2010
cz2004
cz2016
cz1992
x

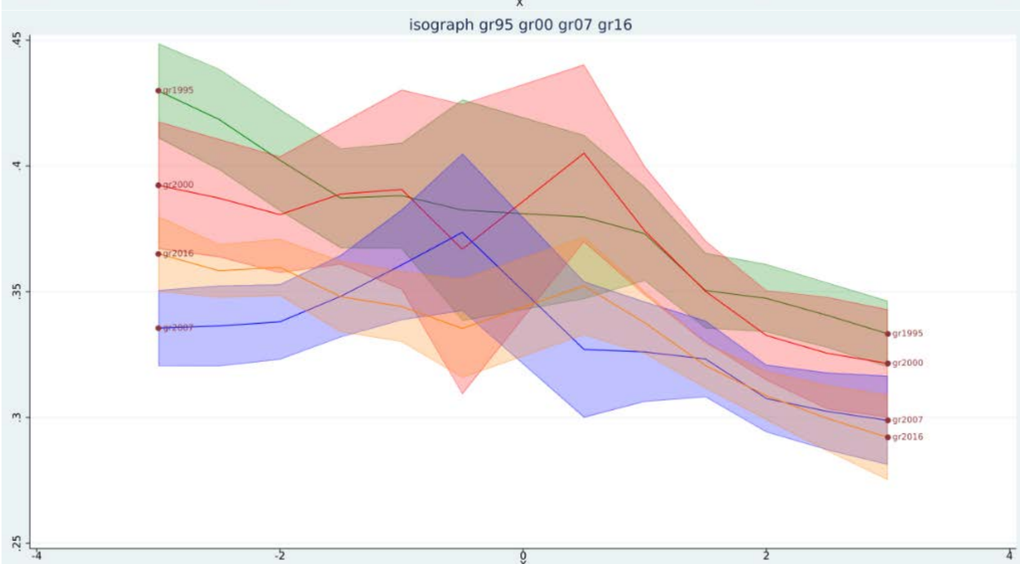

isograph de73 de81 de90 de00 de10 de19
de2019
de2010
de2000
de1973
de1990
de1981
x

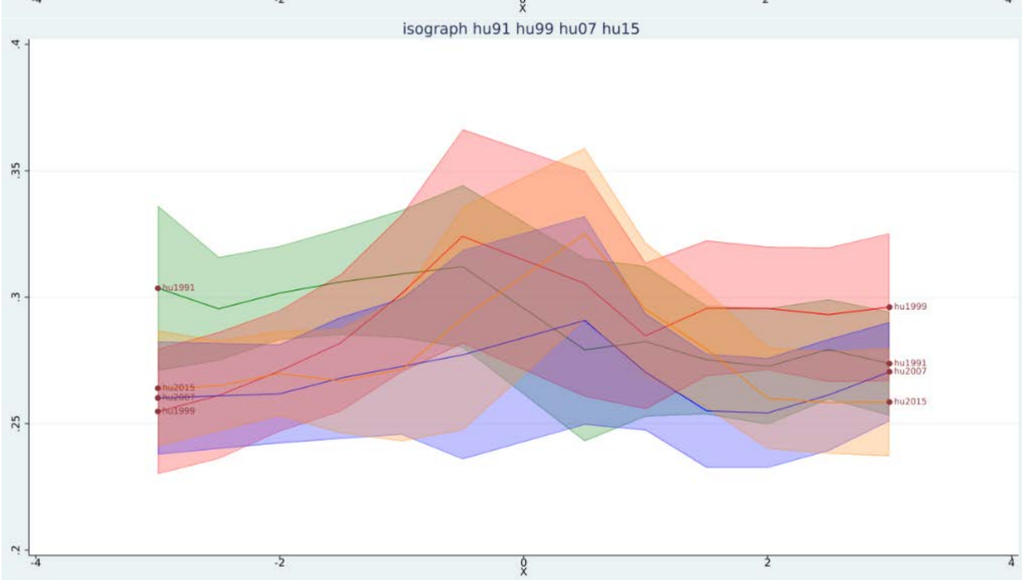

isograph dk87 dk95 dk04 dk10 dk16
.35
.3
.25
.2
-4
-2
0
2
4
x
dk1987
dk2016
dk2010
dk2004
dk1995
dk2016
dk1987
dk2010
dk2004
dk1995

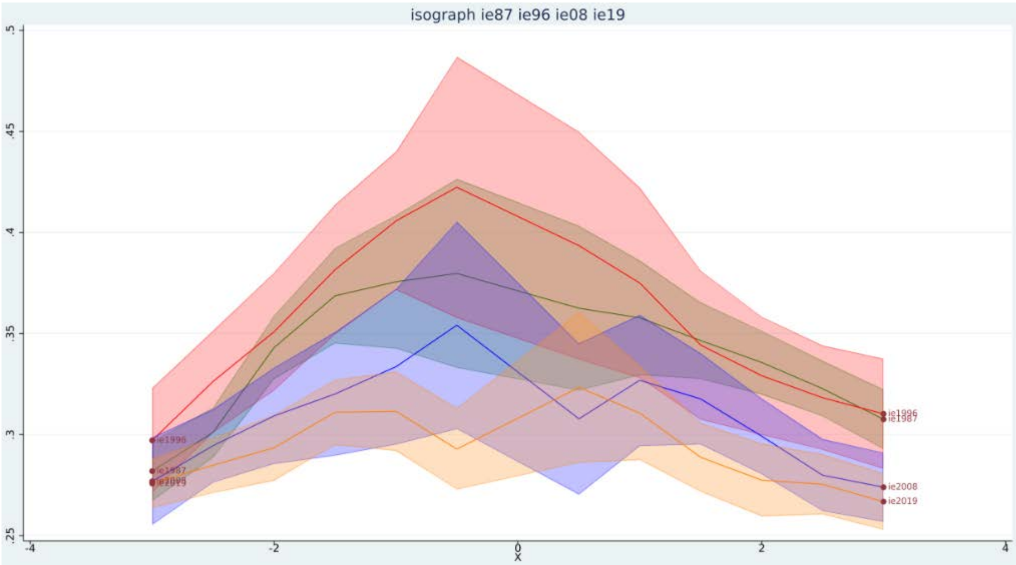

isograph es80 es80 es95 es00 es10 es16
.5
.45
.4
.35
.3
-4
-2
0
2
4
x
es2016
es2010
es1995
es2000
es1980
es1995
es2000

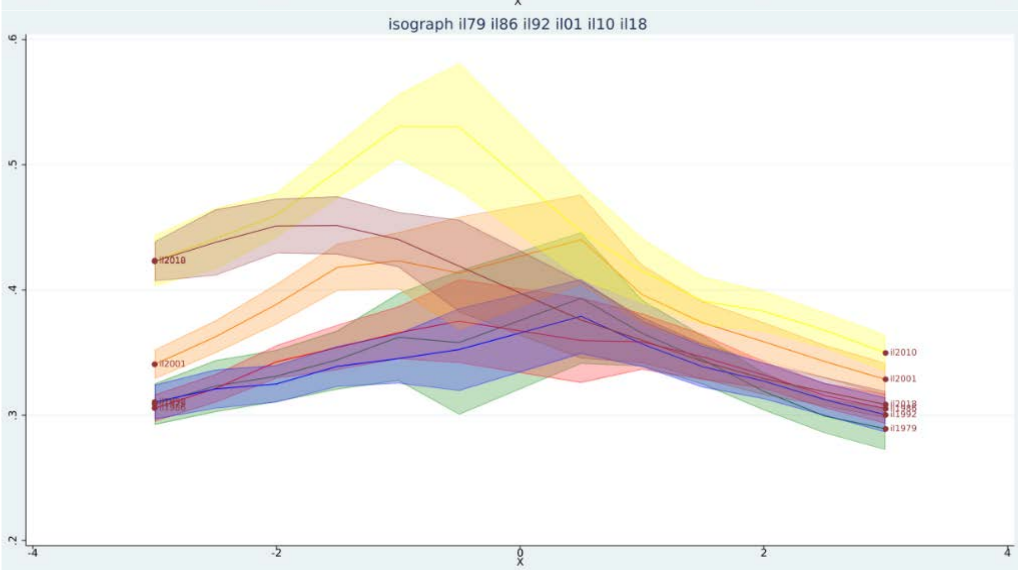

isograph fi87 fi95 fi04 fi10 fi16
.35
.3
.25
.2
.15
-4
-2
0
2
4
x
fi2010
fi2004
fi2016
fi1987
fi1995
fi2016
fi1995
fi1987

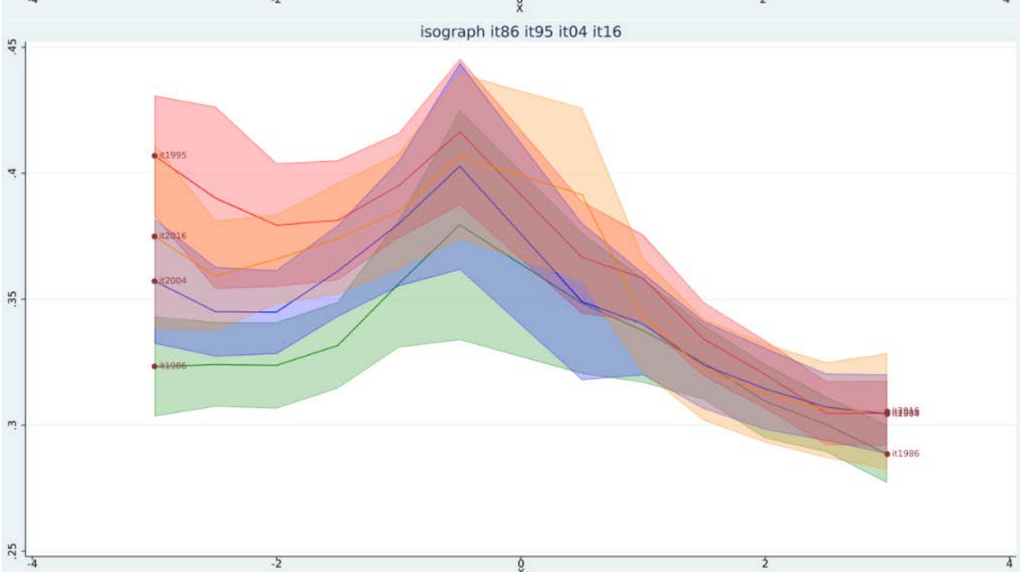

isograph fr70 fr79 fr90 fr00 fr10 fr18
fr1970
fr2010
fr2018
fr1990
x

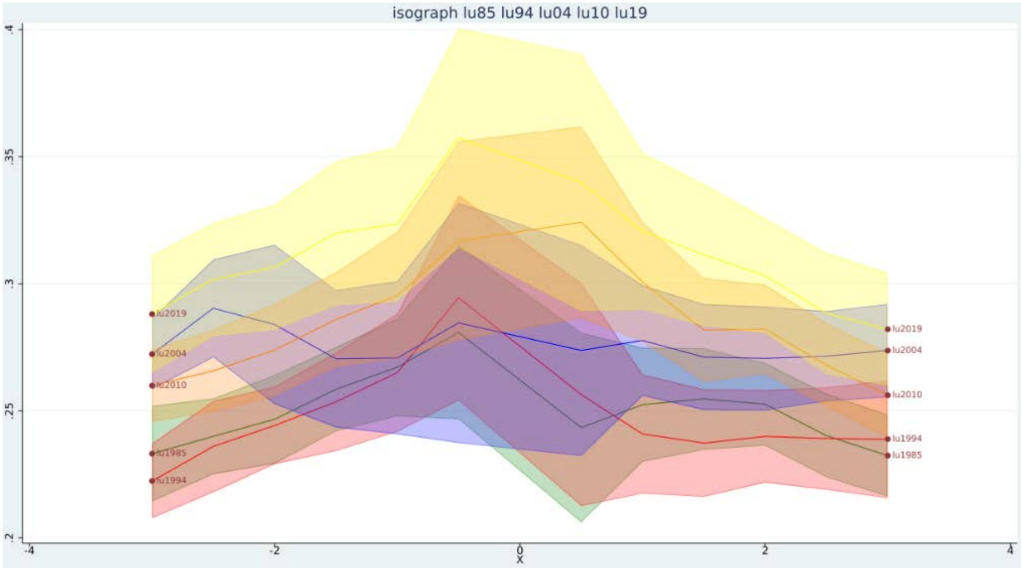

isograph gr95 gr00 gr07 gr16
gr1995
gr2000
gr2016
gr2007
x

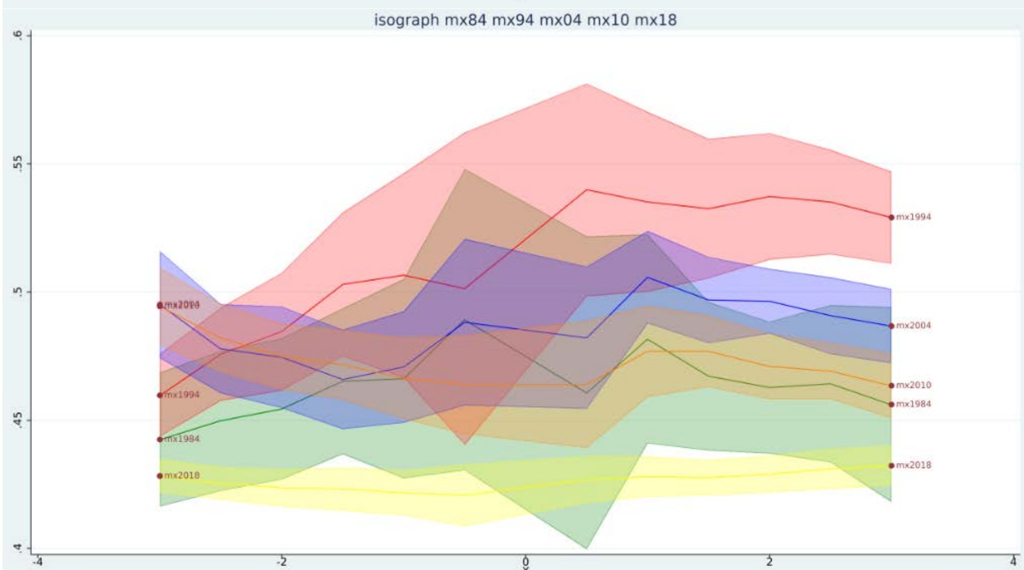

isograph hu91 hu99 hu07 hu15
hu1991
hu1999
hu2007
hu2015
x

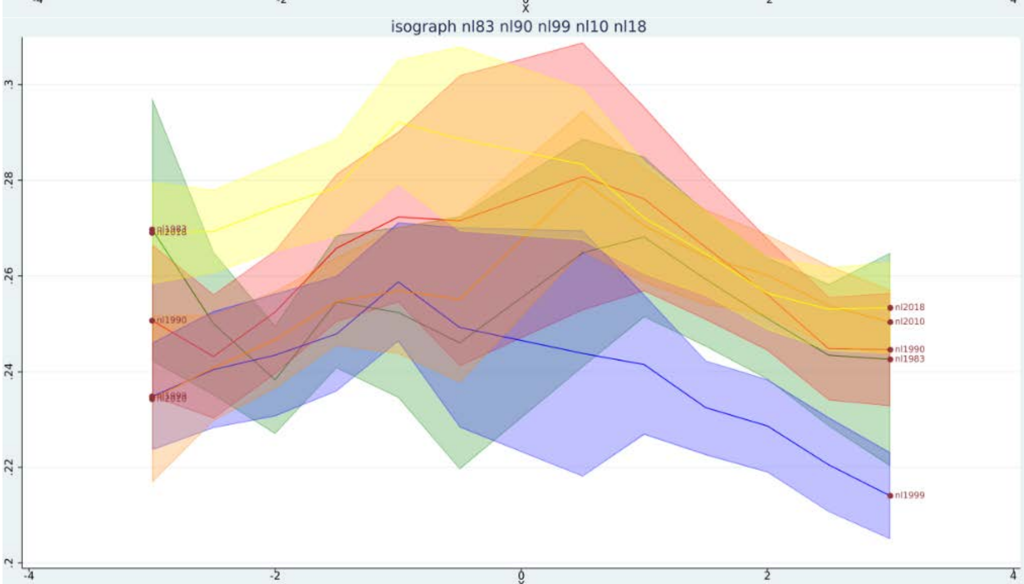

isograph ie87 ie96 ie08 ie19
ie1996
ie1987
ie2008
ie2019
x

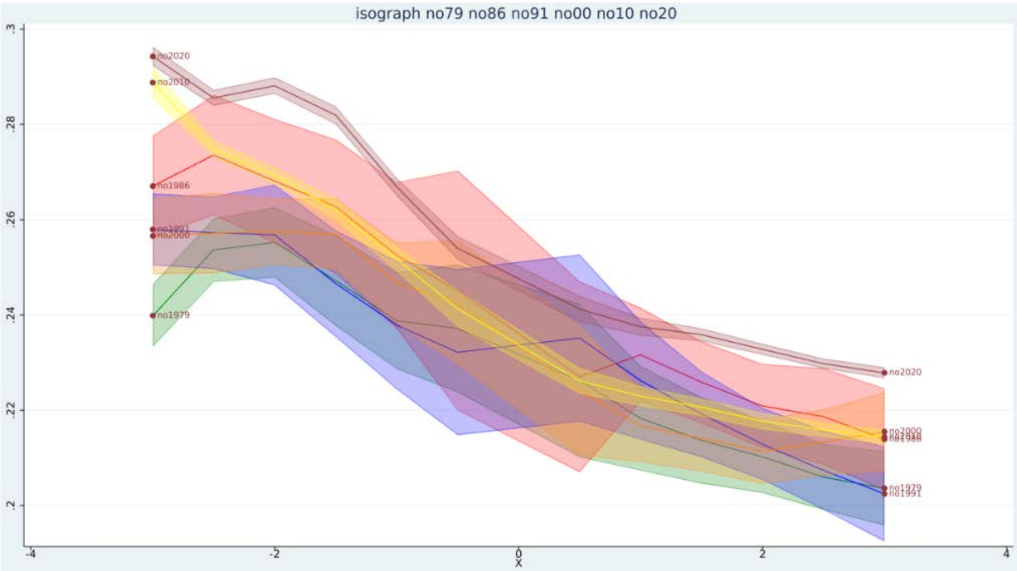

isograph il79 il86 il92 il01 il10 il18
il2018
il2001
il2010
il1979
x

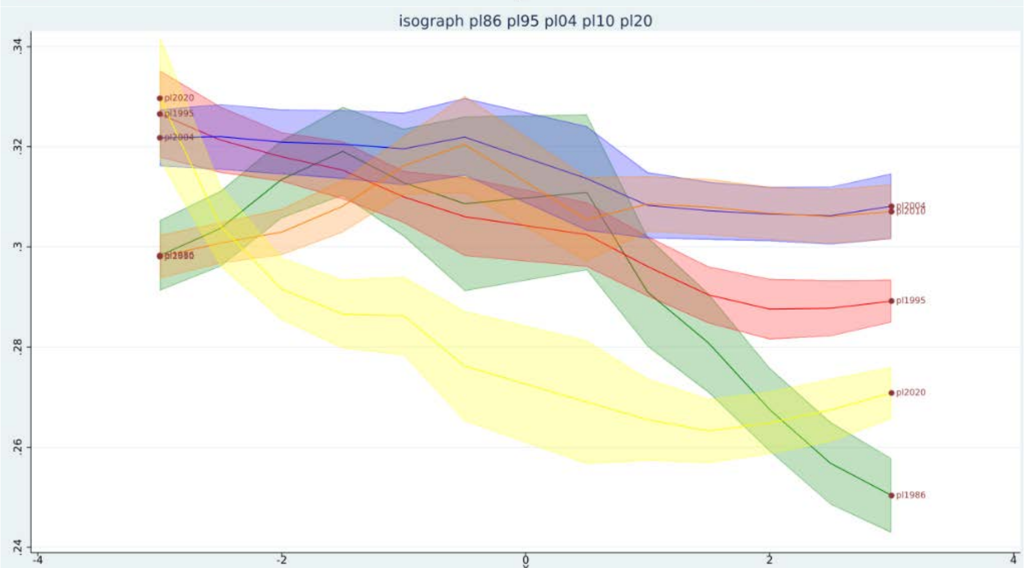

isograph it86 it95 it04 it16
it1995
it2016
it2004
it1986
x

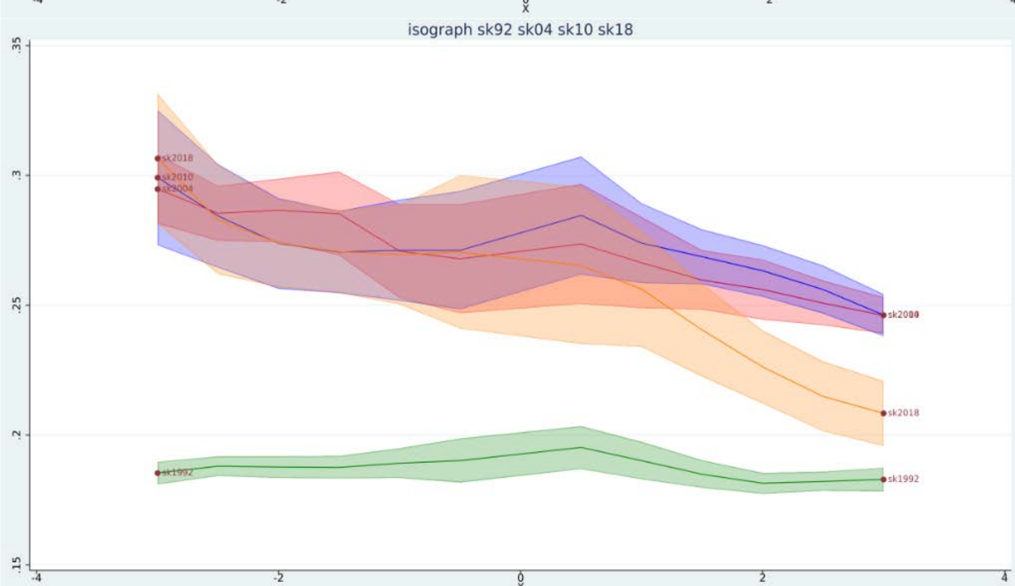

isograph lu85 lu94 lu04 lu10 lu19
lu2019
lu2004
lu2010
lu1985
lu1994
lu2019
lu2004
lu2010
lu1994
lu1985
x

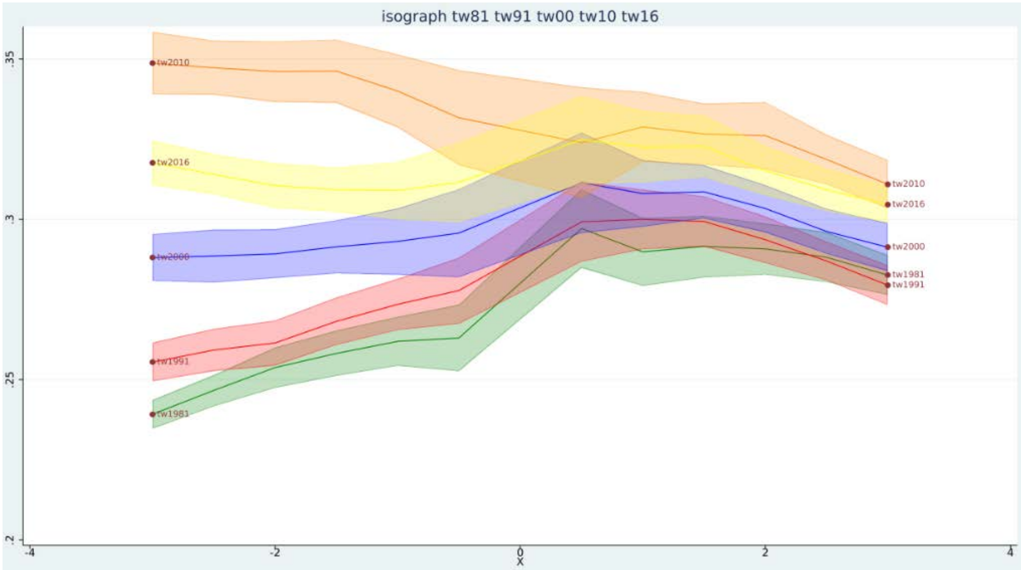

isograph mx84 mx94 mx04 mx10 mx18
mx1994
mx1984
mx2018
mx1994
mx2004
mx2010
mx1984
mx2018
x

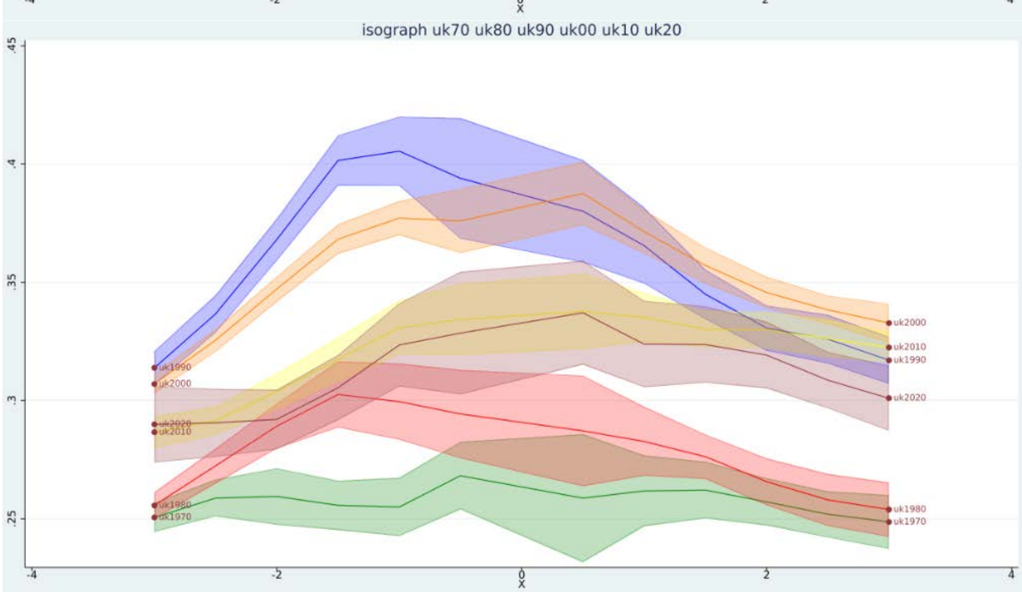

isograph nl83 nl90 nl99 nl10 nl18
nl1990
nl2018
nl2010
nl1990
nl1983
nl1999
x

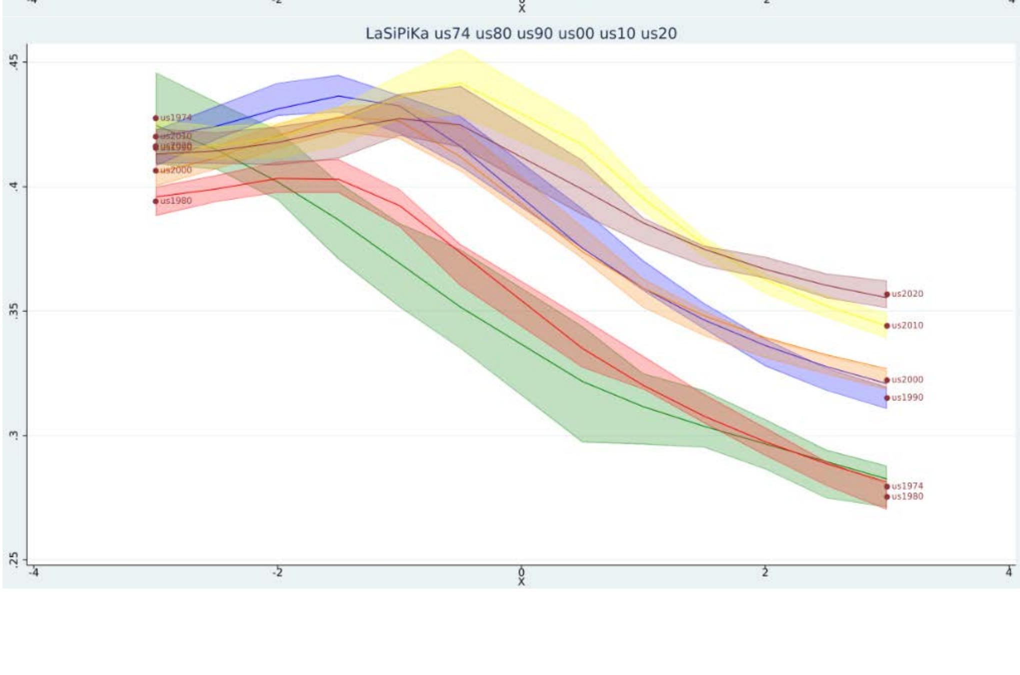

isograph no79 no86 no91 no00 no10 no20
no2020
no2010
no1986
no1991
no2000
no1979
no2020
no1979
no1991
.3
.28
.26
.24
.22
.2
-4
-2
0
2
4
x


isograph pl86 pl95 pl04 pl10 pl20
pl2020
pl1995
pl2004
pl1995
pl2020
pl1986
.34
.32
.3
.28
.26
.24
-4
-2
0
2
4
x


isograph sk92 sk04 sk10 sk18
sk2018
sk2010
sk2004
sk1992
sk2018
sk1992
.35
.3
.25
.2
.15
-4
-2
0
2
4
x


isograph tw81 tw91 tw00 tw10 tw16
tw2010
tw2016
tw2000
tw1991
tw1981
tw2010
tw2016
tw2000
tw1981
tw1991
.35
.3
.25
.2
-4
-2
0
2
4
x


isograph uk70 uk80 uk90 uk00 uk10 uk20
uk1990
uk2000
uk2020
uk2010
uk1980
uk1970
uk2000
uk2010
uk1990
uk2020
uk1980
uk1970
.45
.4
.35
.3
.25
-4
-2
0
2
4
x


LaSiPiKa us74 us80 us90 us00 us10 us20
us1974
us2010
us2000
us1980
us2020
us2010
us2000
us1990
us1974
us1980
.45
.4
.35
.3
.25
-4
-2
0
2
4
x