Evolution of global inequality in well-being: A copula-based approach

Koen Decancq, Vanesa Jordá

[1]Herman Deeleck Centre for Social Policy

[2]Department of Economics, Universidad de Cantabria

III/LIS Comparative Economic Inequality Conference
London School of Economics, February 25, 2023

Contents

1. Introduction
2. The joint distribution of well-being
3. Marginal distributions
4. The copula: A model for the dependence structure
5. Multivariate dependence and inequality

Introduction

RESEARCH AIM

To analyse the evolution of global inequality in well-being from 1990 to 2010.

PRIOR RESEARCH

Centred on purely monetary outcomes: Milanovic (2005), Lakner and Milanovic (2017), Anand and Segal (2015; 2017); Niño Zarazúa et al. (2017); Jorda and Nio Zarazúa (2017).

Multidimensional well-being: Becker et al. (2005), Decancq (2011), Decancq et al. (2009), McGillivray and Pillarisetti (2004), McGillivray and Markova (2010), MartínezÂ (2012)

Introduction

RESEARCH AIM

To analyse the evolution of global inequality in well-being from 1990 to 2010.

PRIOR RESEARCH

Centred on purely monetary outcomes: Milanovic (2005), Lakner and Milanovic (2017), Anand and Segal (2015; 2017); Niño Zarazúa et al. (2017); Jorda and Nio Zarazúa (2017).

Multidimensional well-being: Becker et al. (2005), Decancq (2011), Decancq et al. (2009), McGillivray and Pillarisetti (2004), McGillivray and Markova (2010), MartínezÂ (2012)

Introduction

RESEARCH AIM

To analyse the evolution of global inequality in well-being from 1990 to 2010.

PRIOR RESEARCH

Centred on purely monetary outcomes: Milanovic (2005), Lakner and Milanovic (2017), Anand and Segal (2015; 2017); Niño Zarazúa et al. (2017); Jorda and Nio Zarazúa (2017).

Multidimensional well-being: Becker et al. (2005), Decancq (2011), Decancq et al. (2009), McGillivray and Pillarisetti (2004), McGillivray and Markova (2010), MartínezÂ (2012)

Introduction

RESEARCH AIM

To analyse the evolution of global inequality in well-being from 1990 to 2010.

PRIOR RESEARCH

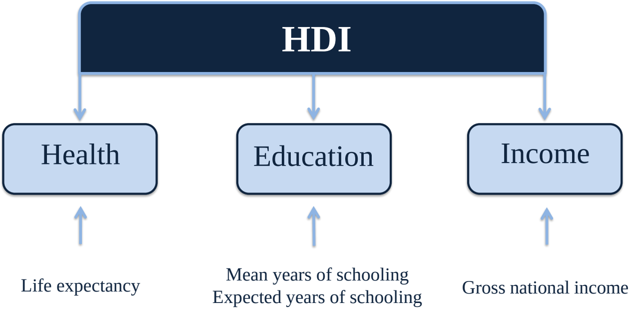
Centred on purely monetary outcomes: Milanovic (2005), Lakner and Milanovic (2017), Anand and Segal (2015; 2017); Niño Zarazúa et al. (2017); Jorda and Niño Zarazúa (2017).

Multidimensional well-being: Becker et al. (2005), Decancq (2011), Decancq et al. (2009), McGillivray and Pillarisetti (2004), McGillivray and Markova (2010), Martínez (2012)

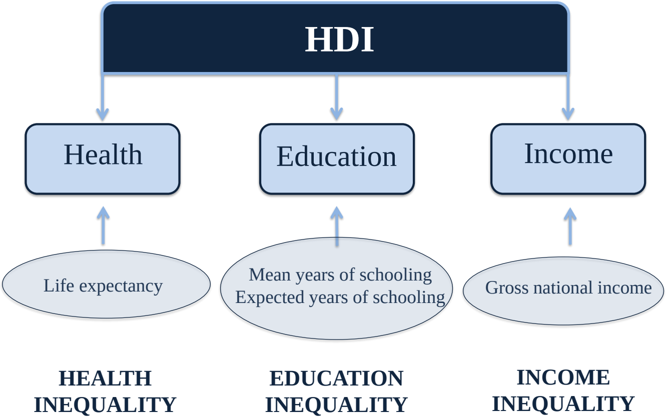
Introduction

- The notion of **well-being** as a **multidimensional** construct that involves aspects beyond economic growth, is well-established in the economic literature.
- Academics became increasingly concerned with exploring the ways in which different aspects of well-being could be summarised in a **composite index**.
- The **Human Development Index** (HDI) is by far the most prominent indicator of well-being, which evaluates not only income capacity, but also education and health opportunities.

Introduction



Introduction



What is the contribution?

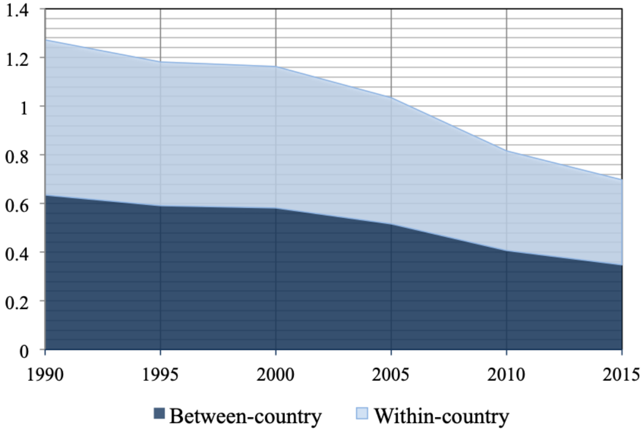


Figure: Evolution of global income inequality (MLD). Source: Jorda and Niño-Zarazúa (2019)

What is the contribution?

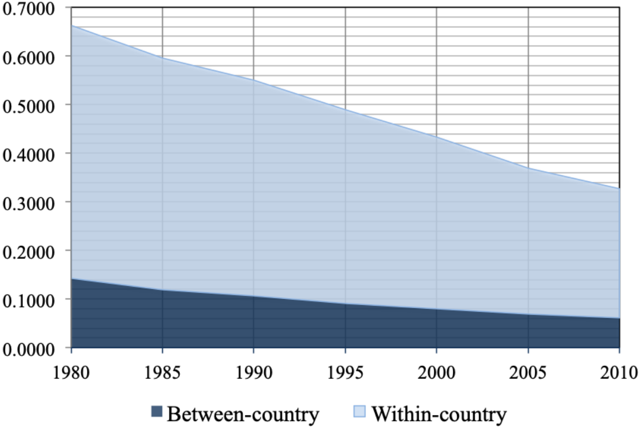


Figure: Evolution of global inequality in education (MLD). Source: Jorda and Alonso (2017)

What is the contribution?

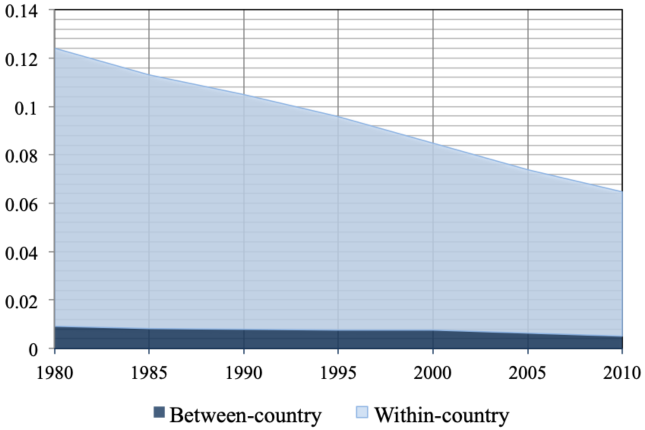


Figure: Evolution of global inequality in lifespans (MLD). Source: Jorda and Niño-Zarazúa (2017)

Why multidimensional inequality?

Scenario 1	Scenario 2
$\begin{bmatrix} \text{inc} & \text{educ} & \text{health} \\ 10000 & 18 & 95 \\ 5000 & 12 & 70 \end{bmatrix}$	$\begin{bmatrix} \text{inc} & \text{educ} & \text{health} \\ 5000 & 18 & 95 \\ 10000 & 12 & 70 \end{bmatrix}$

The joint distribution of well-being

The multivariate distribution of well-being

Let $W = (Y, X, H)$ be a random vector that consists of three random variables that reflect the three dimensions of well-being: income (Y), educational outcomes (X) and length of life (H). A realization of the random vector (y, x, h) reflects the well-being outcomes of an individual. The joint cumulative density function of Y, X, H can be defined as follows (Sklar, 1959),

$$F(y, h, x) = C\left(F_Y(y), F_X(x), F_H(h)\right) \quad (1)$$

where $F_Y(y), F_X(x), F_H(h)$ are the cumulative density functions (cdf) of the global distribution of income, education and health respectively and $C(.)$ is the copula function, a multivariate distribution with uniform marginals with support in $[0, 1]^n$.

Contents

1. Introduction
2. The joint distribution of well-being
3. Marginal distributions
4. The copula: A model for the dependence structure
5. Multivariate dependence and inequality

The global distribution of income

Individual data is not generally available, so we rely on data from the World Income Inequality Database (**WIID**)

Example: USA, 2013 (WIID v3.3)

Population shares	20%	40%	60%	80%	100%
Income shares	3.2%	11.6%	25.9%	48.9%	100%

We approximate the mean of income distribution by per capita **Gross National Income (GNI)** adjusted by purchasing power parities (PPP) at constant prices of 2011, retrieved from the World Bank's World Development Indicators.

The global distribution of income

The Generalized Beta distribution of the Second Kind (GB2) provides an excellent approximation the income distribution, which yields reliable estimates of relative inequality measures (Jorda et al., 2022).

The national distribution of income

Let Y be a non-negative random variable representing the individual income, which follows a GB2 distribution with probability density function (pdf) given by,

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

where $B(p, q) = \int_0^1 t^{p-1}(1-t)^{q-1} dt$ is the beta function.

The parameters a, p and q are shape parameters and b is a scale parameter.

The global distribution of income

We construct the **global income distribution** in each year as a **mixture** of national distributions weighted by population:

$$F(y;\mathbf{a},\mathbf{b},\mathbf{p},\mathbf{q},\boldsymbol{\eta}) = \sum_{k=1}^{K} \eta_k B\left(\frac{(y/b_k)^{a_k}}{1+(y/b_k)^{a_k}}; p_k, q_k\right), \qquad (2)$$

where η_k stands for the population weight of the country $k, k = 1, \ldots, K$ and $B(.,.)$ is the beta function.

$$F(y,h,x) = C\left(F_Y(y), F_X(x), F_H(h)\right)$$

The global distribution of educational outcomes

Example: Japan 1990

Educational level	Attainment rate($h^{(j)}$)	Official duration($D^{(j)}$)
No schooling	0.0024	0
Primary incomplete	0.2413	?
Primary complete	0.0800	6
Secondary incomplete	0.2973	?
Secondary complete	0.1636	12
Tertiary incomplete	0.1248	?
Tertiary complete	0.0905	16

The global distribution of educational outcomes

We deploy the parametric approach developed by Jorda and Alonso (2017) avoids relying on discretionary assumptions and yields an accurate representation of the distribution of education.

The national distribution of education

Let X be a non-negative random variable representing the time of schooling until either completing the maximum level of education or dropping out school, which follows a generalised gamma distribution with pdf given by (Stacy, 1962),

$$f(x; a, b, p) = \frac{a x^{ap-1} e^{-(x/b)^a}}{b^{ap}\Gamma(a)}, \quad x \geq 0,$$

where $\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx$ is the gamma function, $b > 0$ is a scale parameter, and $a > 0$ and $p > 0$ are shape parameters.

The global distribution of educational outcomes

The **global distribution** of education in each year is constructed from national estimates, as a **mixture** of distributions,

$$F(x; \mathbf{a}; \mathbf{b}; \mathbf{p}; \boldsymbol{\eta}) = \sum_{k=1}^{K} \eta_k \mathrm{IG}\left[x^{a_k}; p_k, (b_k)^{a_k}\right], \tag{3}$$

where η_k stands for the population weight of the country $k, k = 1, \ldots, K$ and $\mathrm{IG}(.)$ denotes the incomplete gamma function.

$$F(y, h, x) = C\left(F_Y(y), F_X(x), F_H(h)\right)$$

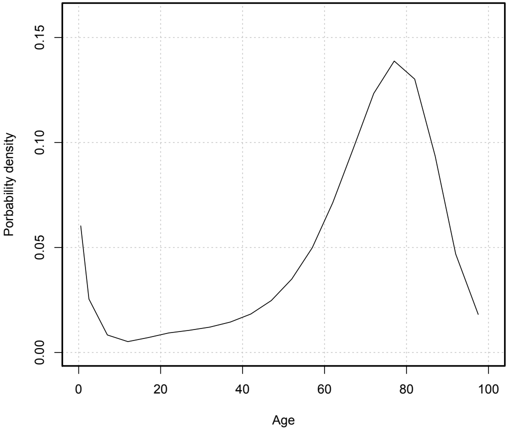
The global distribution of lifespans

Period life tables contain data on the number of **survivors** for every five-year age group up to 100 years for a synthetic cohort of 100,000 individuals (UN Population Division, 2017).

0	1	5	10	...	95	100+
100 000	86 043	78 719	76199	...	308	37

For a certain year, period life tables are constructed using data on the number of deaths in that particular year.

The global distribution of lifespans



The global distribution of lifespans

To obtain a reliable approximation of the global distribution of the variable length of life, we use a mixture of two Weibull distributions.

The national distribution of lifespans

Let H be a non-negative random variable representing the lifespan of an individual with a pdf of the following form:

$$f(h; \alpha_1, \lambda_1, \alpha_2, \lambda_2, w) = w \frac{\lambda_1}{\alpha_1} \left(\frac{h}{\alpha_1} \right)^{\lambda_1 - 1} e^{-(h/\alpha_1)^{\lambda_1}} + (1-w) \frac{\lambda_2}{\alpha_2} \left(\frac{h}{\alpha_2} \right)^{\lambda_1 - 1} e^{-(h/\alpha_2)^{\lambda_2}},$$

where λ_1, λ_2 are shape parameters, α_1, α_2 are scale parameters and w is the weight of each component (i.e. infant mortality and adult mortality).

The global distribution of lifespans

We use the national estimates to construct the **global distribution** of lifespans at the year t as a **mixture** of the national distributions weighted by population:

$$F(h; \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \mathbf{w}, \boldsymbol{\eta}) = \sum_{k=1}^{K} \eta_k \left(w_k \frac{\lambda_{1,k}}{\beta_{1,k}} \left(\frac{h}{\beta_{1,k}} \right)^{\lambda_{1,k}-1} e^{-(h/\beta_{1,k})^{\lambda_{1,k}}} - (1 - w_k) \frac{\lambda_{2,k}}{\beta_{2,k}} \left(\frac{h}{\beta_{2,k}} \right)^{\lambda_{2,k}-1} e^{-(h/\beta_{2,k})^{\lambda_{2,k}}} \right),$$

where η_k stands for the population weight of the country k, k = 1, ..., K.

$$F(y, h, x) = C\left(F_Y(y), F_X(x), F_H(h)\right)$$

Contents

1. Introduction
2. The joint distribution of well-being
3. Marginal distributions
4. The copula: A model for the dependence structure
5. Multivariate dependence and inequality

The copula: A model for the dependence structures

The structure of **dependence** between income, health and education is fully captured by the **copula**.

$$F(y, h, x) = \textcolor{red}{C}\left(F_Y(y), F_X(x), F_H(h)\right)$$

While there is a growing body of studies exploring the use of copulas to measure **multidimensional poverty and inequality** (Decancq, 2014; Tkach and Gigliarano, 2018; Pérez and Prieto-Alaiz, 2016), there are virtually no published studies about the parametric models that provide a **reliable representation** of the multivariate distribution of well-being.

The copula: A model for the dependence structure

We consider the class of elliptical copulas to be particularly appealing to characterise the dependence structure.

The dependence structure

Let X, Y, H be non-negative random variables representing the level of income, education and the length of life of individuals respectively. The pdf of the Gaussian copula with is of the form:

$$c(u_y, u_h, u_x; \mathbf{\Sigma}) = |\mathbf{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}\xi'(\mathbf{\Sigma}^{-1} - \mathbf{I})\xi\right),$$

where $\boldsymbol{\xi} = (\Phi^{-1}(u_Y), \Phi^{-1}(u_H), \Phi^{-1}(u_E))$ are the quantiles of the standard normal distribution and $\mathbf{\Sigma}$ represents the correlation matrix.

The copula: A model for the dependence structure

LIMITATION: There is no global data (either individual or grouped) on the relationship between income, education and health.

We compute the distribution of well-being for a large number correlation matrices that are definite positive using a grid for $\rho_{i,j} = (0, 0.01, \ldots, 1)$.

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & \rho_{Y,H} & \rho_{Y,X} \\ \rho_{Y,H} & 1 & \rho_{H,X} \\ \rho_{Y,X} & \rho_{H,X} & 1 \end{pmatrix}$$

where $\rho_{i,j} = \frac{\sigma i, j}{\sigma i, \sigma j}, i \neq j$. $\sigma i, j$ is the covariance between variables i and j and σi is the standard deviation.

The copula: A model for the dependence structure

The use of the Gaussian copula might simplify this exercise because, using information on the marginal distributions, we are able to define a **lower bound on the correlation** coefficients for each pair of variables.

To illustrate this, let N be a discrete variable with K categories that identifies the country of residence. The covariance between X_1 and X_2 can be decomposed as follows:

$$\mathrm{Cov}(X_1, X_2) = \mathrm{E}[\mathrm{Cov}(X_1, X_2|N)] + \mathrm{Cov}[\mathrm{E}(X_1|N), \mathrm{E}(X_2|N)]. \qquad (4)$$

The copula: A model for the dependence structure

Because N is a discrete variable with K categories, the covariance between X_1 and X_2 can be expressed as:

$$\mathrm{Cov}(X_1, X_2) = \sum_{k=1}^{K} \eta_k \mathrm{Cov}_k(X_1, X_2) + \sum_{k=1}^{K} \eta_k (\mu_{1,k} - \mu_1)(\mu_{2,k} - \mu_2), \quad (5)$$

where $\eta_k, k = 1, \ldots, K$ is the population weight and $\mathrm{Cov}_k, k = 1, \ldots, K$ is the covariance between X_1 and X_2 in the country k. The mean of variables X_1 and X_2 in the k-th country are $\mu_{1,k}$ and $\mu_{2,k}$ respectively and μ_1 and μ_2 are the global means, which can be defined as $\mu = \sum_{k=1}^{K} \eta_k \mu^{(k)}$.

Contents

1 Introduction

2 The joint distribution of well-being

3 Marginal distributions

4 The copula: A model for the dependence structure

5 Multivariate dependence and inequality

Multivariate inequality

Two-step procedure:

The multivariate Atkinson index (Decancq et al., 2009)

$$I(F_W) = 1 - \left[\iiint \left(\frac{U(y,h,x)}{U(\mu_Y, \mu_H, \mu_X)} \right)^{1-\varepsilon} dF_W(y,h,x) \right]^{1/1-\varepsilon}, \qquad (6)$$

where

$$U(y,h,e) = \left[\omega_Y \times g_Y(y)^{1-\beta} + \omega_H \times g_H(h)^{1-\beta} + \omega_X \times g_X(x)^{1-\beta} \right]^{1/1-\beta}. \qquad (7)$$

Multivariate inequality

The multivariate Atkinson index (Decancq et al., 2009)

$$I(F_W) = 1 - \left[\iiint \left(\frac{U(y,h,x)}{U(\mu_Y, \mu_H, \mu_X)} \right)^{1-\varepsilon} dF_W(y,h,x) \right]^{1/1-\varepsilon},$$

where

$$U(y,h,e) = \left[\omega_Y \times g_Y\,(y)^{\,1-\beta} + \omega_H \times g_H\,(h)^{\,1-\beta} + \omega_X \times g_X\,(x)^{\,1-\beta} \right]^{1/1-\beta}.$$

Multivariate inequality

The multivariate Atkinson index (Decancq et al., 2009)

$$I(F_W) = 1 - \left[\iiint \left(\frac{U(y,h,x)}{U(\mu_Y, \mu_H, \mu_X)} \right)^{1-\varepsilon} dF_W(y,h,x) \right]^{1/1-\varepsilon},$$

where

$$U(y,h,e) = \left[\omega_Y \times g_Y(y)^{1-\beta} + \omega_H \times g_H(h)^{1-\beta} + \omega_X \times g_X(x)^{1-\beta} \right]^{1/1-\beta}.$$

Multivariate inequality

The multivariate Atkinson index (Decancq et al., 2009)

$$I(F_W) = 1 - \left[\iiint \left(\frac{U(y,h,x)}{U(\mu_Y, \mu_H, \mu_X)} \right)^{1-\varepsilon} dF_W(y,h,x) \right]^{1/1-\varepsilon},$$

where

$$U(y,h,e) = \left[\omega_Y \times g_Y(y)^{1-\beta} + \omega_H \times g_H(h)^{1-\beta} + \omega_X \times g_X(x)^{1-\beta} \right]^{1/1-\beta}.$$

Multivariate inequality

The multivariate Atkinson index (Decancq et al., 2009)

$$I(F_W) = 1 - \left[\iiint \left(\frac{U(y,h,x)}{U(\mu_Y, \mu_H, \mu_X)} \right)^{1-\varepsilon} dF_W(y,h,x) \right]^{1/1-\varepsilon},$$

where

$$U(y,h,e) = \left[\omega_Y \times g_Y(y)^{1-\beta} + \omega_H \times g_H(h)^{1-\beta} + \omega_X \times g_X(x)^{1-\beta} \right]^{1/1-\beta}.$$

Multivariate inequality

An increase in the dependence between the dimensions only increases multidimensional inequality if (Bourguignon, 1999)

$$\varepsilon > \beta$$

The higher the selected degree of complementarity between the dimensions, the higher the inequality aversion has to be, for an increase in dependence to lead to an increase of well-being inequality.

When $\varepsilon = \beta$, the multidimensional inequality index becomes insensitive to changes in the dependence between the dimensions and multidimensional inequality can be computed based on information about the marginal distributions alone. (Foster et al., 2005; Alkire and Foster, 2010)

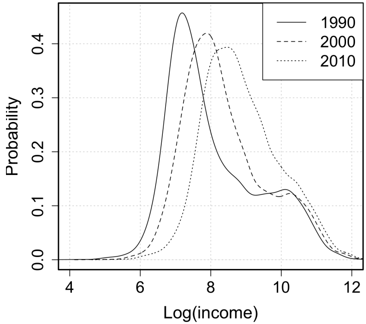
Multivariate dependence

The multivariate Spearman index (Decancq, 2014)

$$\rho\left(C_W\right)=\frac{(m+1)\left[2^{m-1}\int\left[C_{\perp}+\overline{C}_{\perp}\right]dC_W-1\right]}{2^m-(m+1)}, \tag{8}$$

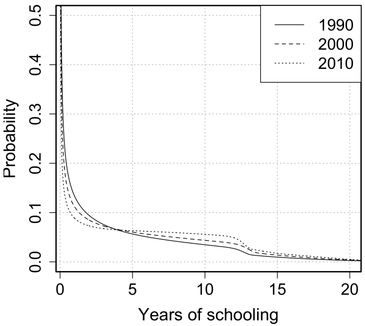
where $C_{\perp}=\Pr\left(Y\leq y,X\leq x,H\leq h\right)=F_Y(y)*F_X(x)*F_H(h)$ is the cdf of the independence copula and $\overline{C}_{\perp}=\Pr\left(Y>y,X>x,H>h\right)$ is the survival independence copula.

The global distribution of income



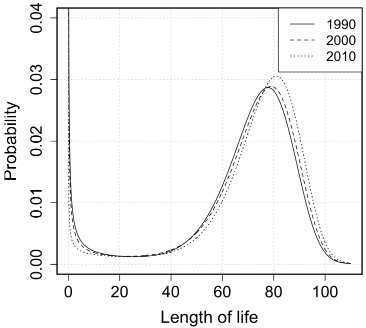
Variable	Year	$\epsilon = 0.5$	$\epsilon = 0.75$	$\epsilon = 1$
Income	1990	0.3891	0.5230	0.6198
	2000	0.3478	0.4676	0.5573
	2010	0.2738	0.3777	0.4622

The global distribution of educational outcomes



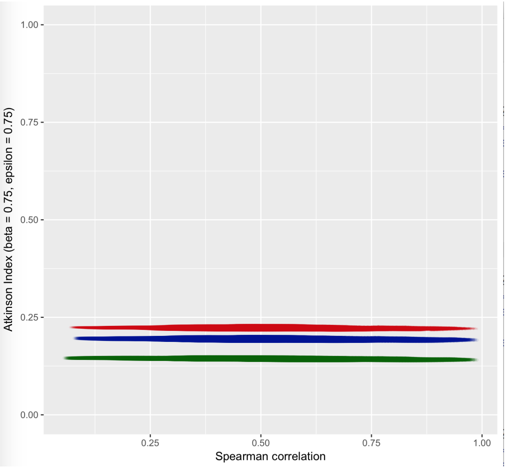
Variable	Year	$\epsilon = 0.5$	$\epsilon = 0.75$	$\epsilon = 1$
Education	1990	0.2962	0.4824	0.7628
	2000	0.2478	0.4116	0.6723
	2010	0.1917	0.3205	0.5201

The global distribution of lifespans

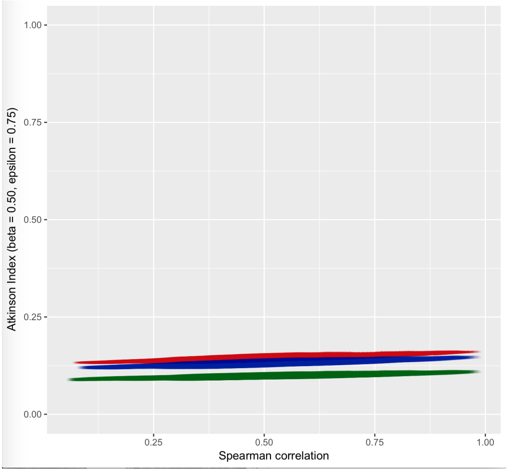


Variable	Year	$\epsilon = 0.5$	$\epsilon = 0.75$	$\epsilon = 1$
Health	1990	0.0650	0.1268	0.2505
	2000	0.0546	0.1053	0.2081
	2010	0.0333	0.0610	0.1118

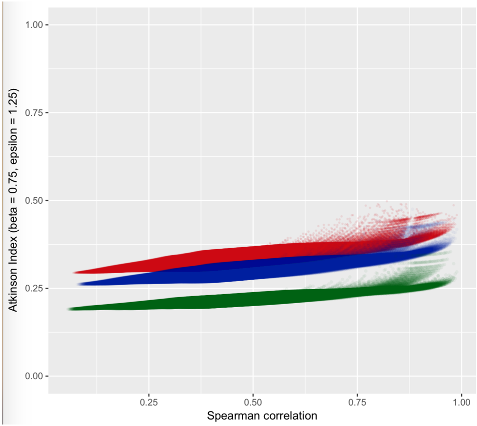
Multivariate Atkinson index ($\beta = \epsilon = 0.75$). 1990 – Red, 2000 – Blue, 2010 – Green



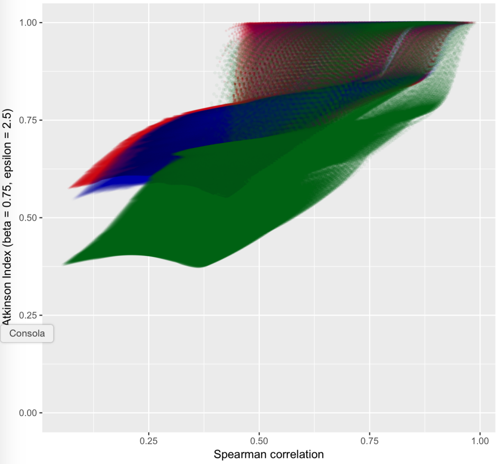
Multivariate Atkinson index ($\beta = 0.5; \epsilon = 0.75$). 1990 – Red, 2000 – Blue, 2010 – Green



Multivariate Atkinson index ($\beta = 0.75; \epsilon = 1.25$). 1990 – Red, 2000 – Blue, 2010 – Green



Multivariate Atkinson index ($\beta = 0.75; \epsilon = 2.5$). 1990 – Red, 2000 – Blue, 2010 – Green



Evolution of global inequality in well-being: A copula-based approach

Koen Decancq, Vanesa Jordá

[1]Herman Deeleck Centre for Social Policy

[2]Department of Economics, Universidad de Cantabria

III/LIS Comparative Economic Inequality Conference
London School of Economics, February 25, 2023