Journal of Multivariate Analysis 170 (2019) 10–24

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Describing the concentration of income populations by functional principal component analysis on Lorenz curves

Enea G. Bongiorno *, Aldo Goia

Dipartimento di Studi per l'Economia e l'Impresa, Università del Piemonte Orientale, via E. Perrone, 18 - 28100 Novara, Italy

ARTICLE INFO

Article history:
Received 12 August 2017
Available online 21 September 2018

AMS subject classification:
62H25
62F12
62P20

Keywords:
Consistency
Hanging cable problem
Hilbert embedding approach
Modes of variation

ABSTRACT

Lorenz curves are widely used in economic studies (inequality, poverty, differentiation, etc.). From a model point of view, such curves can be seen as constrained functional data for which functional principal component analysis (FPCA) could be defined. Although statistically consistent, performing FPCA using the original data can lead to a suboptimal analysis from a mathematical and interpretation point of view. In fact, the family of Lorenz curves lacks very basic (e.g., vectorial) structures and, hence, must be treated with ad hoc methods. This work aims to provide a rigorous mathematical framework via an embedding approach to define a coherent FPCA for Lorenz curves. This approach is used to explore a functional dataset from the Bank of Italy income survey.

1. Introduction

In order to study the probability law of a random variable X, one can consider different real functions (provided their well-posedness), each one highlighting different aspects: the cumulative distribution function F and the corresponding density f or the quantile function, the quantile-density function, and the density-quantile function, respectively defined, for all $p \in (0, 1)$, by $Q(p) = \inf\{x : F(x) \geq p\}$, $q(p) = Q'(p) = 1/f\{Q(p)\}$ and $f\{Q(p)\}$.

An important aspect connected with the study of a distribution and that plays a key role in applied sciences (economics, biology, chemistry, etc.), is the notion of "concentration". Roughly speaking, it is the propension of a non-negative random variable X to redistribute over the individuals within the population. One of the goals in studying concentration is to characterize different settings ranging from the maximal concentration (one individual owns the total mass) and the equidistribution (the mass is distributed equally among all individuals).

In such a framework, a very useful tool is the so-called Lorenz Curve (LC) that was introduced in [27] to represent the concentration of wealth. Formally, given a non-negative random variable X with finite mean μ, its LC $L(p)$ is defined, see [19], by

$$L : [0, 1] \to \mathbb{R} : p \mapsto L(p) = \frac{1}{\mu} \int_0^p Q(t)dt.$$

It is easy to see that $L(0) = 0$, whereas $L(1) = 1$ as $\int Q(t)dt = \mu$. Moreover, since $Q(p)$ is non-negative and increasing, $L(p)$ is increasing and convex.

Because $X \geq 0$, the quantity $\int_0^p Q(t)dt$ may be interpreted as the mass of X held by the first $100 \times p\%$ of the individuals of a population ordered by increasing values of X, i.e., $\mathrm{E}[X\, \mathbf{1}_{\{X \leq Q(p)\}}]$, where $\mathbf{1}_A$ is the indicator function of the set A, whereas

* Corresponding author.
E-mail addresses: enea.bongiorno@uniupo.it (E.G. Bongiorno), aldo.goia@uniupo.it (A. Goia).

https://doi.org/10.1016/j.jmva.2018.09.005
0047-259X/

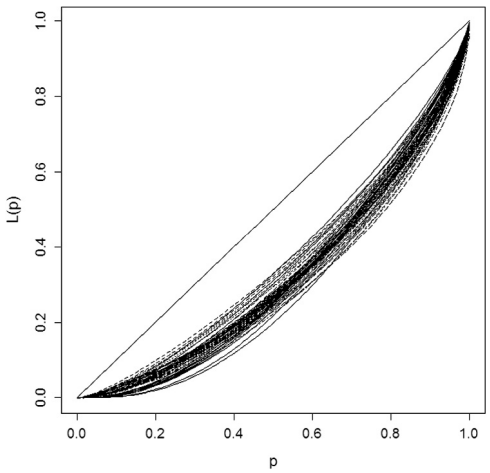

Fig. 1. Lorenz Curves of groups of individuals for class of age during the period 1987–2014.

μ represents the total mass of X. Therefore, $L(p)$ describes and measures how a positive quantity X concentrates within the population. From a mathematical point of view, the LC determines the probability distribution of X up to a scale factor transformation and uniquely characterizes the distribution whenever the latter has a known compact support; see, e.g., [24,35]. Among all LCs, the egalitarian line $L(p) = p$ plays an important role: it corresponds to a perfectly equal distribution in which each individual owns the same quantity or when X is a degenerate random variable that equals μ. Any other LC lies below the equidistribution line that, hence, is used as a basis for comparison which leads to define concentration indexes measuring inequality within the population.

Estimators of $L(p)$ from samples drawn from X and their theoretical properties have been widely studied; see, e.g., [9–11,21,25,39].

Consider now the problem of comparing different distributions in terms of concentration: for instance to study the concentration of incomes over years, countries, regions, groups defined from social–economical stratification criteria, and so on. In all these situations, one deals with a set of estimated LCs, each one referred to a specific group. By way of example, the LCs computed from data gathered with the Survey on Household Income and Wealth of the Bank of Italy (see Section 4 for details) are represented in Fig. 1. Each curve graphs the concentration of personal income for individuals grouped for age range (up to 30 years old, 31–40, 41–50, 51–65 and over 65) and year of survey (from 1987 to 2014, biennial).

To manage that comparison problem, researchers focused on the construction of some hierarchies based on LCs. There exists a wide literature defining different kinds of orderings ranging from the use of synthetic indexes, like the Gini index, to stochastic orderings; see, e.g., [1,34] and references therein. The scientific debate on this topic, especially in the economic and econometric community, is still open, as testified by recent publications; see, e.g., the monographs [4,8] and references therein.

In this paper, the above mentioned problem is tackled for the first time, to our knowledge, using techniques from functional data analysis, which are designed for the analysis of data that are curves or more complicated objects such as images, surfaces, etc. To have an idea, although incomplete, of the wide variety of mathematical models, statistical techniques and feasible applications in such framework, see the collection of papers [3,5], the Special Issue [20], the survey [12] and the monographs [6,18,22,33].

To proceed in this direction, one needs a rigorous formal framework to fully exploit the functional nature of the data and to interpret the results meaningfully. A first difficulty stems from the fact that the considered functional data are not curves directly observed over a suitable grid, as it usually occurs in the classical functional literature, but a set of LCs, each one estimated from a sample of real random variables related to a specific level (a country, a region, and so on). This induces the necessity to manage a double stochasticity in the definition of the functional objects, a first one related to the sampling among the levels and a second one within each level.

Once the functional framework is rigorously defined, suitable statistical functional techniques can be used to identify structural properties and highlight differences among LCs. In particular, we focus on functional principal component analysis (FPCA), a technique which generalizes the well-known principal component analysis from finite- to infinite-dimensional spaces, and allows to reduce the dimensionality and to visualize the most important modes of variation of the data; see, e.g., [33]. This methodology requires that the data belong to a vector space, but the family of LCs cannot be straightforwardly endowed with a vector space structure since curves are non-negative, bounded, increasing and convex functions mapping $[0, 1]$ to itself. Hence, as we show in this paper, a naive direct application of FPCA is suboptimal because it could produce incoherent interpretations.

In order to overcome this drawback, we propose to embed bijectively the family of LCs in a Hilbert space where FPCA results can be coherently interpreted. The bijection is derived in two steps. In the first step, each LC is seen as the unique solution of a boundary value problem whose physical interpretation allows to read the second derivative of each LC as a local inequality weight. In the second step, that second derivative is mapped to a Hilbert space through the negative centered log-ratio transformation. Although the latter map is often used in the compositional data literature, here we show that we deal with data that cannot be classified as compositional and, hence, the negative centered log-ratio transformation plays only a technical role. Besides the theoretical and algorithm aspects, we also study the consistency of the mode of variation of the FPCA obtained both in the naive and the embedding approaches.

To complete the analysis, the proposed method is applied to the study of the evolution of concentration of income of Italians from 1987 to 2014 using micro-data from the Survey on Household Income and Wealth of the Bank of Italy. After introducing three different stratification criteria (geographical, generational and sectorial), the method allows to analyze the positioning and the dynamic of each group, introduced by stratification, over time in the factorial plane.

Given the range of topics covered, this work can be thought as a new step forward in the study of samples of densities (see, e.g., [14,15,23,26,30,32]) and/or their derived objects such as level sets [16], quantile synchronized density [38] and hazard functions [32].

The outline of the paper is as follows. In Section 2, we introduce the mathematical setting. In Section 3, the naive FPCA approach on samples of LCs and the embedding one are introduced; their theoretical and algorithm aspects are then discussed. Finally, in Section 4 the methodology is applied to a real dataset. Proofs of the theoretical results are collected in the Appendix.

2. Lorenz curves as a functional data

This section introduces the mathematical aspects related to LCs as functional data. Consider a population $\mathcal{X}$ formed by real and positive random variables X having probability density functions sharing the same compact support $[0, 1]$ and define the family of LCs

$$\mathcal{L}or = \{L : [0, 1] \to [0, 1] \ : \ L(0) = 0, L(1) = 1, L \in C^2[0, 1], L'(p) > 0 \text{ and } L''(p) > 0 \text{ for } p \in (0, 1)\}.$$

Because $\mathcal{L}or$ is a subset of $\mathcal{L}^2_{[0,1]}$, i.e., the Hilbert space of square integrable real functions on $[0, 1]$ with the inner product $\langle g, h\rangle = \int g(t)h(t)dt$ and the induced norm $\|g\|^2 = \langle g, g\rangle$, $\mathcal{L}or$ can be endowed with $\mathcal{B}_{\mathcal{L}or}$, the σ-algebra induced by $\|\cdot\|$ on $\mathcal{L}^2_{[0,1]}$. On the population $\mathcal{X}$ we define the random LC as the map

$$\mathbb{L} : (\mathcal{X}, \mathcal{B}_{\mathcal{X}}) \to (\mathcal{L}or, \mathcal{B}_{\mathcal{L}or}) : X \mapsto \mathbb{L}(X) = L(\cdot),$$

where $\mathcal{B}_{\mathcal{X}} = \{A \subseteq \mathcal{X} : \mathbb{L}(A) \in \mathcal{B}_{\mathcal{L}or}\}$ is the σ-algebra induced by $\mathbb{L}$, so that it is measurable. From now on and for the sake of simplicity, we will use L instead of $\mathbb{L}$.

By considering L as a random element in $\mathcal{L}^2_{[0,1]}$, it is possible to define its mean curve and the covariance operator as follows. For all $p \in [0, 1]$ and $v \in \mathcal{L}^2_{[0,1]}$,

$$\ell(p) = \mathrm{E}\{L(p)\}, \quad \Sigma(v) = \mathrm{E}\{\langle L - \ell, v\rangle(L - \ell)\}.$$

Consider now the empirical counterpart, and suppose we deal with a sample $X_1, \ldots, X_n$ of elements drawn from $\mathcal{X}$. To each X_i is associated an LC L_i which, in practice, is estimated from a sample drawn from X_i of size n_i, denoted $X_i^1, \ldots, X_i^{n_i}$. This can be done by introducing the empirical LC for the ith sample, defined for all $p \in [0, 1]$, by

$$\widehat{L}_i(p) = \frac{1}{\overline{X}_i} \int_0^p \widehat{Q}_i(t)dt, \tag{1}$$

where $\overline{X}_i$ is the empirical mean and $\widehat{Q}_i$ denotes the ith empirical quantile function associated to $X_i^1, \ldots, X_i^{n_i}$. Consistency results for each empirical LC are available whenever X_i is absolutely continuous; see, e.g., [11]. Finally one has a sample of n functional data $\widehat{L}_1(p), \ldots, \widehat{L}_n(p)$. For computational purposes, each empirical LC can be evaluated over a common selected grid of points over $[0, 1]$.

From estimator (1) one can define the empirical versions of the Lorenz mean curve ℓ and the covariance operator Σ by letting, for each $p \in [0, 1]$ and $v \in \mathcal{L}^2_{[0,1]}$,

$$\widehat{\ell}_n(p) = \frac{1}{n}\sum_{i=1}^n \widehat{L}_i(p), \quad \widehat{\Sigma}_n(v) = \frac{1}{n}\sum_{i=1}^n \langle \widehat{L}_i - \widehat{\ell}_n, v\rangle(\widehat{L}_i - \widehat{\ell}_n).$$

Summarizing, this setting allows to model those situations in which a measurement (such as the income) concentrates in different regions or levels. In these cases, researchers deal with different empirical LCs, each one related to a different level. From a theoretical point of view, this means that one has to handle a double stochasticity: one related to the randomness of the distribution (between the levels) and the other linked to the sample variability (within the levels). In other words, a random unobserved distribution is associated to each individual (the level or the region) and, from such distribution, a sample is drawn to get the corresponding empirical LC.

3. Functional principal component analysis for Lorenz curves

In the previous section, LCs and their empirical counterparts are defined to fit the "classical" functional data analysis framework in which random functions are observed over a grid of deterministic points. We are now ready to tackle the problem of reasonably applying the FPCA to empirical LCs. In Section 3.1 we discuss how a naive use of FPCA could produce some incoherences. In Section 3.2 we present the embedding that is the starting point to perform FPCA whose results are coherent, interpretable and statistically consistent as shown in Section 3.3.

3.1. Problems using a naive approach

Since $\mathcal{L}or \subset \mathcal{L}^2_{[0,1]}$, the latter seems a good candidate to implement FPCA in a naive way. Consider the eigenvalues $\lambda_1, \lambda_2, \ldots$ and eigenfunctions $\xi_1, \xi_2, \ldots$ of the covariance operator Σ. From them it is possible to approximate L by means of a truncated version of the Karhunen–Loève decomposition of integer order $q \geq 1$; see Theorem 1.5 in [6]. For all $p \in [0, 1]$,

$$L^q(p) = \ell(p) + \sum_{j=1}^{q} \theta_j \xi_j(p), \tag{2}$$

where $\theta_j = \langle L - \ell, \xi_j \rangle$ is the so-called jth principal component of L, satisfying $\mathrm{E}(\theta_j) = 0$, $\mathrm{E}(\theta_j^2) = \lambda_j$ and $\mathrm{E}(\theta_j \theta_h) = 0$ if $j \neq h$. Because the eigenfunctions $\xi_1, \xi_2, \ldots$ can also be seen as the result of a variance maximization iterative procedure, they are commonly used to visualize the most important modes of variation of the random curve as perturbations of the mean. In practice, the jth modes of variation, defined, for all $p \in [0, 1]$ and real $k \geq 0$, by

$$\omega_j(p, k) = \ell \pm k(\lambda_j)^{1/2} \xi_j(p),$$

are interpreted as the effects of adding and subtracting a suitable multiple of each non standardized eigenfunction $\sqrt{\lambda_j}\,\xi_j$. The constant k is usually chosen subjectively to better appreciate how the different ξ_js affect the mean.

From the estimator (1) one can define the empirical versions of objects introduced above. In particular, once estimates $(\widehat{\lambda}_{1,n}, \widehat{\xi}_{1,n}), \ldots, (\widehat{\lambda}_{n,n}, \widehat{\xi}_{n,n})$ of the eigenelements $(\lambda_1, \xi_1), (\lambda_2, \xi_2), \ldots$ are derived from the eigendecomposition of the empirical covariance operator $\widehat{\Sigma}_n$, one can obtain the empirical version $\widehat{L}^q_{i,n}$ of (2), the empirical PCs $\widehat{\theta}_{i,j,n} = \langle \widehat{L}_i - \widehat{\ell}_n, \widehat{\xi}_{j,n} \rangle$, and the empirical jth modes of variation, given, for all $p \in [0, 1]$ and $k \geq 0$, by

$$\widehat{\omega}_{j,n}(p, k) = \widehat{\ell}_n(p) \pm k\widehat{\xi}_{j,n}(p)(\widehat{\lambda}_{j,n})^{1/2}.$$

For practical purposes, as empirical LCs are computed over a grid of finite points over $[0, 1]$, all the calculations are made replacing integrals by summations.

As a matter of completeness, it is useful to analyze the behavior of the estimated jth modes of variation when the sizes n and n_i diverge. To do this and to manage the double stochasticity, we assume that a family $\{F(x, \gamma)\}$ of random cumulative distribution function is associated to the population $\mathcal{X}$. The randomness depends on the real random vector γ and the following conditions are assumed:

(A1) $F(\gamma, \cdot)$ and $F^{-1}(\gamma, \cdot)$ are a.s. continuous on $[0, \infty)$ and $(0, 1)$ respectively.
(A2) There exists a positive constant Λ independent on γ, such that $\int_0^\infty x^2 dF(\cdot, x) \leq \Lambda < \infty$ a.s.
(A3) There exist $\delta > 0$ and two positive constants c_1 and c_2 such that $c_1 n^{2\delta} \leq n_i \leq c_2 n^{2\delta}$ as $n \to \infty$.

In this framework we derive the following consistency result whose proof can be found in Appendix A.3.

Proposition 1. *Under Assumptions (A1)–(A3), for a fixed integer $j \in \mathbb{N} = \{1, 2, \ldots\}$ and real $k \geq 0$, one has $\widehat{\omega}_{j,n}(p, k) \to \omega_j(p, k)$ in probability, as $n \to \infty$.*

At this stage, it is worth noticing that the Karhunen–Loève decomposition leads to approximations L^q and $\widehat{L}^q_{i,n}$, and modes of variations $\omega_j(p, k)$ and $\widehat{\omega}_{j,n}(p, k)$ that are functions in $\mathcal{L}^2_{[0,1]}$ but not necessarily in $\mathcal{L}or$, as illustrated in the following example.

Example 1. Consider $L(p) = Up^a + (1 - U)p^b$, $p \in [0, 1]$, with $1 \leq a < b$, and U uniformly distributed on $[0, 1]$. The covariance operator has a unique eigenvalue $\lambda = 1/12$ with associated eigenfunction $\xi = |p^a - p^b|$. Direct calculations lead to

$$\omega(p, k) = (p^a + p^b)/2 \pm k(p^a - p^b)/\sqrt{12},$$

which is a $\mathcal{L}^2_{[0,1]}$ function belonging to $\mathcal{L}or$ only for $0 \leq k < \sqrt{3}$. This is visualized in Fig. 2 where we have graphed a set of curves from L with $a = 1$, $b = 6$ and selected values of $U \in \{0, 0.1, \ldots, 1\}$ (left panel), and the corresponding modes of variation when $k = 1$ and $k = 2$ (right panel). It is evident that the $\omega(p, 1)$s are elements of $\mathcal{L}or$ whereas the $\omega(p, 2)$s are not.

Hence, the idea of using FPCA naively in $\mathcal{L}or$, though it produces consistent estimates, suffers from drawbacks that practically restrict its applicability and its interpretability. This is a consequence of the fact that $\mathcal{L}or$ is not a vector subspace

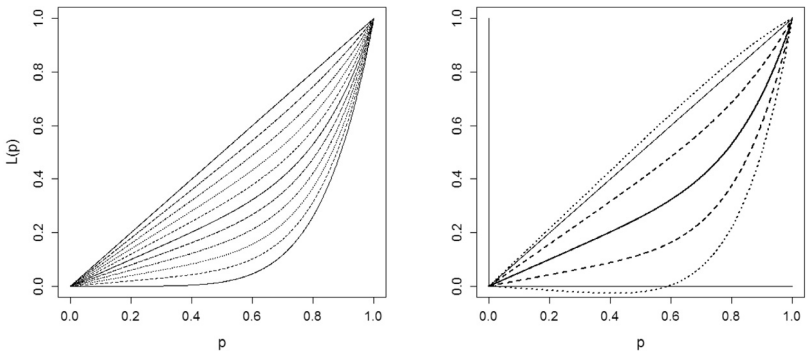

Fig. 2. Left panel: a set of LCs with $a = 1$, $b = 6$ and $U \in \{0, 0.1, \ldots, 1\}$. Right panel: the mean curve and the egalitarian line (solid lines), the modes of variation when $k = 1$ (dashed lines) and $k = 2$ (dotted lines).

of $\mathcal{L}^2_{[0,1]}$, as can be verified straightforwardly. Thus, in what follows we consider a bijective continuous embedding for LCs that allows to represent them in a structured space and to apply FPCA to the transformed functions. It turns out that such embedding spontaneously induces a Hilbert space structure on $\mathcal{L}or$.

3.2. An embedding approach

When one deals with a dataset of constrained functions (e.g., positive, and/or monotone, and so on), a typical approach to provide structured spaces, is to consider some transformations of the original data. In particular, it can be useful to express them as solutions of differential equations which often are related to physical interpretation; see [33].

The features of LCs suggest to draw inspiration from the so-called "hanging cable". Interpreting an LC L as a cable whose extremes are fixed at positions $(0, 0)$ and $(1, 1)$ with linear density mass given by the second derivative L'', the aim is to find the shape which minimizes its potential energy. In this perspective, L is the unique solution of the boundary value problem (BVP)

$$\begin{cases} u''(p) = L''(p) & \text{if } p \in (0, 1), \\ u(0) = 0, \qquad u(1) = 1. \end{cases}$$

As shown in Appendix A.1, such a solution can be expressed, for all $p \in [0, 1]$, as

$$L(p) = p + (p - 1)\int_0^p zL''(z)dz + p\int_p^1 (z - 1)L''(z)dz. \tag{3}$$

The above representation provides a characterization of the LCs by means of their second derivative. This is summarized in the following commutative diagram:

$$\mathcal{L}or \underset{BVP}{\overset{D^2}{\rightleftarrows}} D^2\mathcal{L}or,$$

where $D^2\mathcal{L}or = \{L'' : L \in \mathcal{L}or\}$, D^2 denotes the second derivative operator and BVP is the operator which, applied to an element in $D^2\mathcal{L}or$, associates its BVP solution according to (3). It is worth noting that L'' can be seen as the non-linear warped version of the probability density function $f(x)$ given, for all $p \in (0, 1)$, by

$$L''(p) = 1/\mu f\{Q(p)\} = s(p)/\mu, \tag{4}$$

where $f\{Q(p)\}$ is called density-quantile (see [31]), and its reciprocal $s(p)$ is known as the sparsity function; see [36]. Because $s(p)$, and consequently $L''(p)$, measures the extent of sparseness of the data around the p-quantile, from the concentration point of view, $L''(p)$ can be also interpreted as a local measure of inequality of individuals close to the p-quantile. Summarizing, the physical interpretation of the hanging cable problem, together with (4), provides a new economic interpretation of $L''(p)$ as the local inequality weight at the p-quantile.

The properties and the interpretation above move the attention to $D^2\mathcal{L}or$, which is still not a vector space, as it contains only non-negative functions. Nevertheless it is an important tool in developing further analysis. In fact, consider $\mathcal{L}^2_c = \{g : g \in \mathcal{L}^2_{[0,1]}, \int g = 0\}$, the Hilbert space of centered $\mathcal{L}^2_{[0,1]}$-functions, and the negative centered log-ratio transformation

$$nclr : D^2\mathcal{L}or \to \mathcal{L}^2_c : h \mapsto -\ln(h) + \int_0^1 \ln\{h(t)\}dt. \tag{5}$$

The latter embeds $D^2\mathcal{L}or$ in $\mathcal{L}_c^2$ and its inverse function is

$$nclr^{-1} : \mathcal{L}_c^2 \to D^2\mathcal{L}or : g \mapsto \exp(-g)/\kappa_g,$$

where $\kappa_g = \int_0^1 \int_0^p \exp\{-g(z)\}dzdp$. Hence, as shown in Appendix A.2, the following commutative diagram holds:

$$D^2\mathcal{L}or \underset{nclr^{-1}}{\overset{nclr}{\rightleftarrows}} \mathcal{L}_c^2. \tag{6}$$

Combining the above maps, we get the bijective transformation $\psi(L) = nclr(L'')$ that associates each LC (having a square integrable log-second derivative) to an element of the Hilbert space $\mathcal{L}_c^2$. Its inverse is given by

$$\psi^{-1}(g) = BVP\{\exp(-g)/\kappa_g\} \tag{7}$$

for any centered square integrable function g. For the sake of readability, technical aspects concerning the invertibility of ψ are discussed in Appendix A.2.

It is worth noting that the *nclr* transformation is often employed in the literature which generalizes compositional data (see [2]) to the continuous case in order to work in a proper Hilbert space. For instance, in [23] the authors consider the family of probability density functions; each one is a representative of the equivalence class containing all its positive multiples. One may thus wonder whether L'' can be considered as a continuous compositional data, i.e., if any positive multiple of L'' leads to the same LC by means of (3). In general it is not true: for instance, take the LC $L(p) = p(p+1)/2$ whose second derivative is $L''(p) = 1$ for all $p \in (0, 1)$ and consider cL'', with $c > 2$; Eq. (3) gives a function which is not even an LC because it is negative for $p \in (0, 1 - 2/c)$. The main difference between the compositional framework and our setting is related to the inversion of the *nclr*. Indeed, in the compositional setting, the inverse function is $C\exp(-g)$, where the positive constant C can be chosen arbitrarily, whereas in our setting, the constant must be κ_g to guarantee the invertibility of ψ; see Appendix A.2.

To conclude this section, we provide some comments on the assumption that the considered random variables X have pdfs with common support $[0, 1]$. Since each LC identifies the underlying pdf up to a scale factor, the proposed FPCA approach still holds if one takes $[0, b]$, with $0 < b < \infty$, instead of $[0, 1]$. In contrast, if one takes $[a, b]$, with $0 < a < b < \infty$, then ψ is no longer bijective. One possible solution is to work on the family of LCs obtained from $X - a$ keeping in mind that, although the LCs of X and $X - a$ are different, they are related by an affine transformation whose known coefficients depend on a and the mean of X.

3.3. Computational aspects and consistency results

Consider the problem of estimating FPCA by using the embedding approach discussed above. Given a sample of empirical LCs $\widehat{L}_1(p), \ldots, \widehat{L}_n(p)$, as in Section 2, one has to evaluate the second derivatives. Since each $\widehat{L}_i(p)$ is linear piecewise, to work directly on it does not make sense: one possible solution is to obtain a smoothed version $\widetilde{L}_i(p)$ from which to derive $\widetilde{L}_i''(p)$. Various approaches are feasible for algorithmic purposes: for instance, one can use B-splines [13], constrained penalized splines [29], a suitable kernel smoothing approach [37], or local polynomial smoothing [17].

Once a sample of smooth curves $\widetilde{L}_1''(p), \ldots, \widetilde{L}_n''(p)$ is available, each curve can be transformed by using the *nclr* map (5) to obtain, for each $i \in \{1, \ldots, n\}$,

$$\widehat{\Psi}_i(p) = -\ln\{\widetilde{L}_i''(p)\} + \int_0^1 \ln\{\widetilde{L}_i''(p)\}dp,$$

where the integral is evaluated numerically.

Given $\widehat{\Psi}_1(p), \ldots, \widehat{\Psi}_n(p)$, the empirical mean $\widehat{\psi}_n$, the covariance operator $\widehat{\Sigma}_{\Psi,n}$ and its eigenelements $(\widehat{\alpha}_{j,n}, \widehat{v}_{j,n})$ are computed. This allows to obtain a truncated reconstruction of order q, viz.

$$\widehat{\Psi}_{i,q}(p) = \widehat{\psi}_n(p) + \sum_{j=1}^{q} \langle \widehat{v}_{j,n}, \widehat{\Psi}_i \rangle \widehat{v}_{j,n},$$

and the jth modes of variation,

$$\widehat{m}_{j,n}(k, p) = \widehat{\psi}_n(p) \pm k\widehat{v}_{j,n}(p)(\widehat{\alpha}_{j,n})^{1/2}.$$

As usual in FPCA, the fraction of explained variance is computed from the eigenvalues $\widehat{\alpha}_{j,n}$.

To obtain the reconstructions of order q of L_i and the jth modes of variation in $\mathcal{L}or_{[0,1]}$, we apply the inverse transformation (7) to $\widehat{\Psi}_{i,q}(p)$ and $\widehat{m}_{j,n}(k, p)$, respectively:

$$\widehat{L}_{i,q}(p) = p + \frac{p-1}{\kappa_\Psi} \int_0^p z e^{-\widehat{\Psi}_{i,q}(z)} dz + \frac{p}{\kappa_\Psi} \int_p^1 (z-1) e^{-\widehat{\Psi}_{i,q}(z)} dz,$$

$$\widehat{M}_j(k, p) = p + \frac{p-1}{\kappa_m} \int_0^p z e^{-\widehat{m}_{j,n}(k,z)} dz + \frac{p}{\kappa_m} \int_p^1 (z-1) e^{-\widehat{m}_{j,n}(k,z)} dz,$$

Table 1
Some synthesis indicators for each age group during the time: Means and standard deviations of personal income (divided by 1000), and sample sizes.

Survey	<30			31–40			41–50			51–65			>65		
Year	Mean	Std.	Size	Mean	Std.	Size	Mean	Std.	Size	Mean	Std.	Size	Mean	Std.	Size
1987	16.1	9.2	2004	22.9	13.6	2269	25.8	18.9	2248	22.5	19.3	3272	14.2	11.6	2468
1989	17.3	8.9	2525	24.5	14.7	2515	28.1	20.6	2561	24.0	20.0	3742	15.3	13.1	2465
1991	16.2	10.8	2159	23.0	13.8	2494	26.8	17.0	2596	23.3	18.5	3838	15.5	11.8	2777
1993	13.5	10.1	2070	21.5	15.7	2464	25.3	20.7	2500	23.2	21.9	3934	15.6	14.0	3305
1995	12.1	8.4	2135	20.0	15.2	2481	23.5	18.4	2650	23.0	25.2	3891	16.0	14.7	3341
1998	12.4	10.0	1785	20.8	17.0	2275	25.2	18.6	2404	24.4	25.1	3470	18.2	24.4	2682
2000	12.1	8.3	1896	20.7	18.0	2442	23.7	18.3	2655	24.2	22.4	3976	17.3	18.0	3333
2002	12.9	11.5	1592	20.0	14.1	2180	23.7	19.3	2559	23.5	21.1	3963	17.4	14.5	3724
2004	13.0	11.9	1524	20.7	25.5	2119	24.1	30.6	2495	24.0	22.9	3941	17.9	14.8	3827
2006	12.8	9.3	1355	20.6	28.2	1930	24.4	21.6	2494	24.4	23.4	3784	18.8	15.1	3851
2008	11.7	8.3	1344	18.0	10.9	1838	22.7	17.6	2545	24.3	21.2	3872	19.4	16.6	4074
2010	11.0	7.7	1230	17.6	11.3	1611	22.5	17.5	2622	24.5	20.3	4069	19.8	18.4	4152
2012	9.9	6.6	1042	16.0	9.7	1516	19.8	14.5	2462	22.5	18.4	4212	19.3	17.2	4377
2014	9.9	8.2	949	15.2	9.8	1271	19.5	15.2	2156	21.8	17.2	4190	18.9	13.9	4907

with

$$\kappa_\Psi = \int_0^1 \int_0^p \exp\{-\widehat{\Psi}_{i,q}(z)\}dzdp, \quad \kappa_m = \int_0^1 \int_0^p \exp\{-\widehat{m}_{j,n}(k,z)\}dzdp.$$

The integrals are computed as above.

To conclude this section, we present the main consistency results on the jth modes of variation for a given positive integer j when the smoothed curves $\widetilde{L}_i$ are obtained by means of B-splines with τ_i equispaced knots and $\tau_i = o(n_i)$. In addition to assumptions (A1)–(A3) as in Section 3.1, we also consider

(A4) The pdf f belongs to $\mathfrak{F} = \{f : \text{supp} f = [0, 1],\ f > 0,\ \int f = 1,\ \int f \ln^4 f < \infty\}$.

Proposition 2. *Under Assumptions (A1)–(A4), for a fixed integer $j \geq 1$ and real $k \geq 0$, one has, as $n \to \infty$, $\widehat{m}_{j,n}(k, p) \to m_j(k, p)$ and $\widehat{M}_{j,n}(k, p) \to M_j(k, p)$ in probability, where $m_j(k, p)$ and $M_j(k, p)$ are the theoretical jth modes of variation when LCs are observed integrally and not over samples.*

The proof of Proposition 2 can be found in Appendix A.4.

4. Application to real data

Since 1965, the Bank of Italy conducts the Survey on Household Income and Wealth. From 1987 it is biennial and the collected data are comparable over time. The survey supplies information on income, saving, consumption expenditure and real wealth of Italian households, as well as anagraphic and labor aspects. The total sample size, in the most recent surveys, is about 8000 households, corresponding to about 20000 individuals.

Starting from the data of personal income in Italy from 1987 to 2014, appropriately adjusted for inflation, we estimated the LCs for specific groups of individuals. Among the available stratification criteria in the survey (geographical, socio-economic, cultural), we considered a demographic variable, the age of each income earner, which seems interesting in studying the generational gap. For this variable, five age-classes are available: 30 years old or less, 31–40, 41–50, 51–65, and over 65. With this choice, matching age group and survey year, we had a sample of $n = 70$ empirical LCs plotted in Fig. 1. Other details about the groups are collected in Table 1, where we report the mean and the standard deviation of income (for readability, they are divided by 1000) and the sample size.

In order to apply the methodology illustrated in Section 3.2, we need to estimate the second derivatives. Here we used a local polynomial smoothing approach, that, from our experience on data, seems to produce good results. In particular, we fitted a cubic polynomial with bandwidths computed according to the plug-in selector described in [17]. The shape of these functions is depicted in Fig. 3.

By performing FPCA on the set of transformed data $\widehat{\Psi}_i$, we estimated the PCs. The fractions of cumulative explained variance by the first three PCs are 0.553, 0.812, 0.884, respectively. In the left panel of Fig. 4 the first three eigenfunctions of the empirical covariance operator in the transformed space are depicted. To illustrate how the first two eigenfunctions affect the shape of the LCs, we exhibit in the mid and right panels of Fig. 4 the estimated modes of variation (in the original space) $\widehat{M}_{j,n}(k, p)$ with $j \in \{1, 2\}$ and $k = \pm 3$ (the dotted lines) and the theoretical LC obtained when $k = 0$ (the continuous line), that we denote $\widehat{M}_n(0, p)$.

From the graphics, it transpires that the first eigenfunction takes its highest values close to zero, and relatively high values close to 1, whereas the values have opposite signs in the central part of the graphic. This suggests that the first eigenfunction describes the relationship between the weight of inequalities in correspondence to the extreme quantiles (in particular,

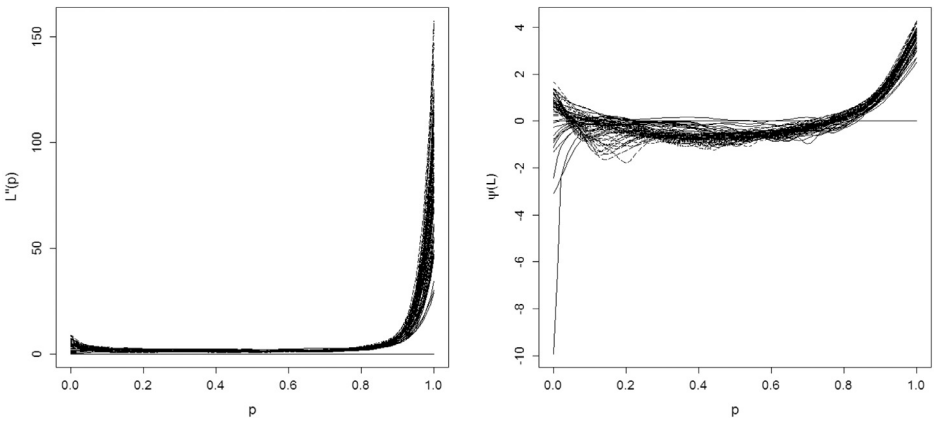

Fig. 3. Inequality weight functions (left) and transformed LC by means of ψ (right), of groups of individuals for class of age during the period 1987–2014.

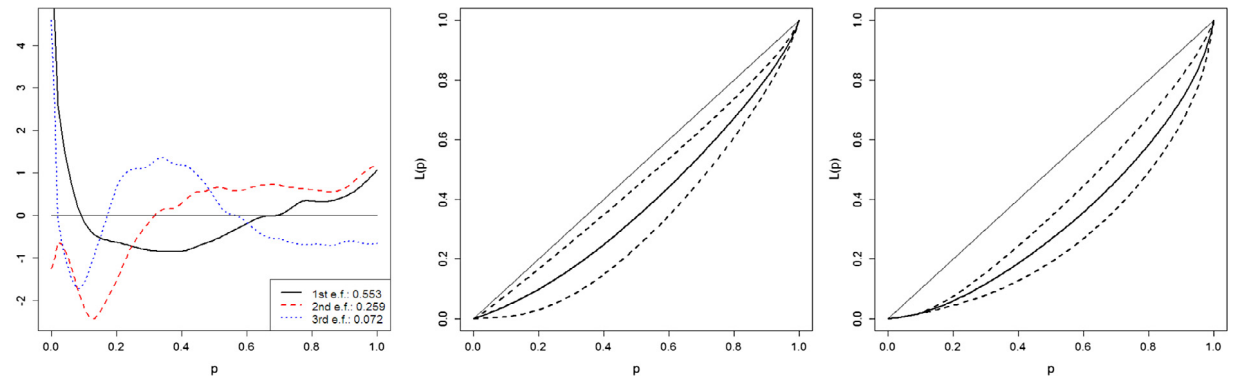

Fig. 4. Left panel: First three eigenfunctions obtained performing FPCA in $\mathcal{H}$. Mid and right panels: First and second modes of variation (the solid lines correspond to $k = 0$, the dashed lines correspond to $k = \pm 3$).

left-quantiles) and the central ones. Thus, the scores of the first PC emphasize how the correspondent LCs behave near zero. In particular, the first PC opposes the LCs having almost horizontal tangent in a right-neighborhood of zero to those with a positive one. In other words, it seems that the first PC allows to distinguish groups where the first 10% of the individuals are very poor from the others, and this can be better appreciated by observing the shapes of the first modes of variation. As for the second eigenfunction, it seems to describe the curvature of LCs mainly due to a change of sign around $p = 20\%$, and the level of the global concentration, i.e., the distance from the egalitarian line, in particular in correspondence of the central quantiles.

To complete the analysis, we present the factorial plane based on the first two PCs. In Fig. 5, we depict the track-plots associated to each age group. With respect to the first PC, the most significant result is a contrast between the group of under 30s and the other ones: the first ones exhibit a high inequality weight in the poorest part of the population and this aspect appears to have gotten worse over time. If we match this fact with the dynamic of the mean of income from 1987, and also consider the trend with respect to the second PC, it appears that the under 30s are becoming poorer (on average) with an ever stronger concentration at the expense of the poorest part of the population. A similar behavior, but with a moderate trend, can be found for people aged 41–50; also in these cases, one witnesses an (albeit less dramatic) impoverishment of the individuals and an increment of concentration. In fact, the tracks tend to converge toward the origin for the first PC, and this is a signal that the tangent of LCs close to zero becomes horizontal over time.

For what concerns the oldest part of the population, the most relevant movement is the vertical one (i.e., with respect the second PC). In these cases, the lines tend to converge toward the origin of the second PC: if the average of income appears rather steady over time (even tendentiously growing for the over 65), the LCs, which in the past denoted a better situation than the one described by $\widehat{M}_n(0, p)$, tends more and more to look like the latter, denoting a worsening of the status of individuals in these groups.

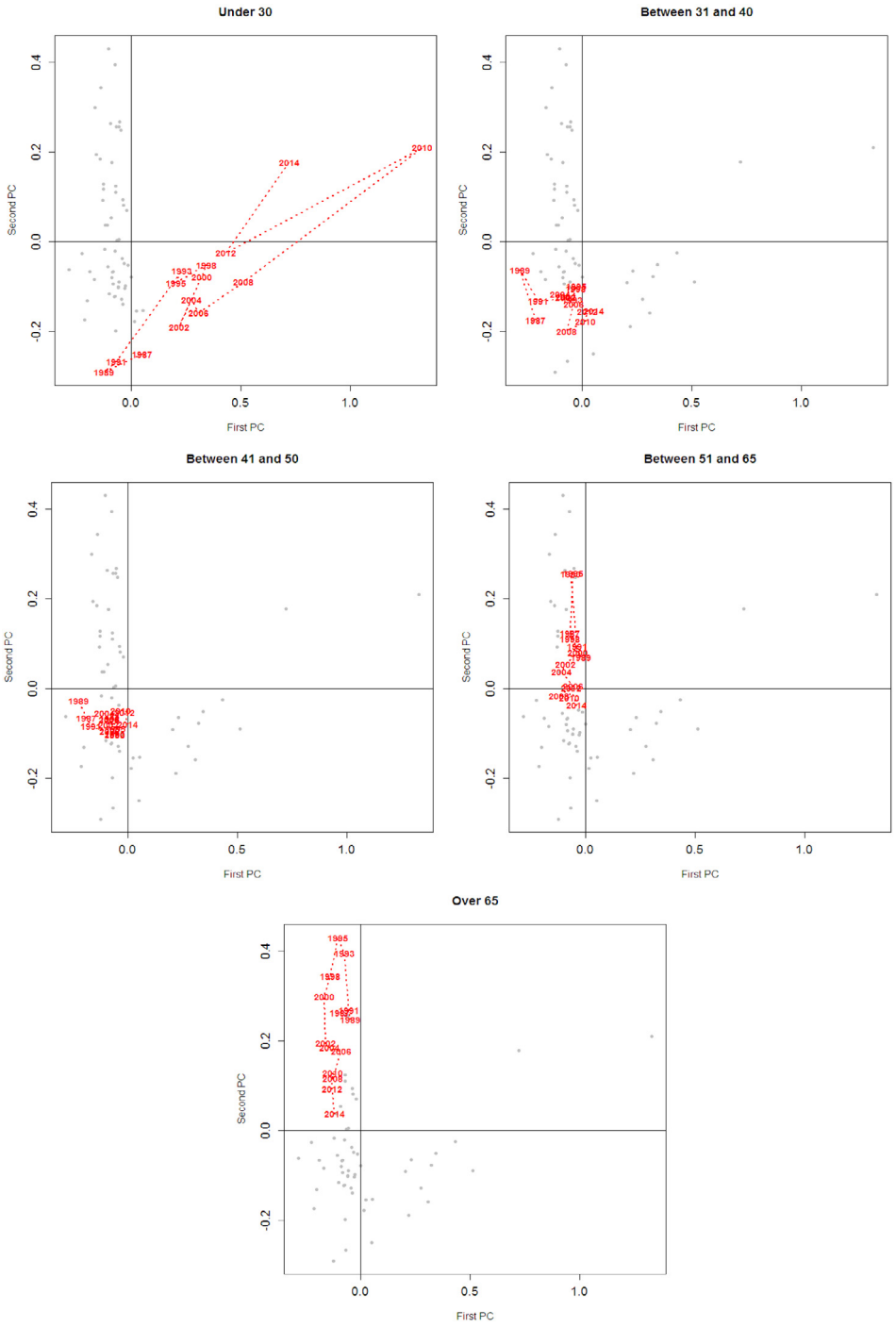

Fig. 5. Track-plots in the factorial plane of the first two PCs for different groups age.

Acknowledgments

The authors thank three anonymous referees, an Associate Editor, and the Editor-in-Chief for their valuable comments that allowed to improve the content and presentation of this paper. The authors are members of the Gruppo Nazionale per l'Analisi Matematica, la Probabilità e le loro Applicazioni (GNAMPA) of the Istituto Nazionale di Alta Matematica (INdAM).

The financial support of CRoNoS–COST Action IC1408 is acknowledged by the first author. The financial support of Università del Piemonte Orientale is acknowledged by the authors.

Appendix. Proofs

A.1. BVP solution

Consider the following BVP

$$\begin{cases} u''(p) = f(p) & \text{if } p \in (0,1), \\ u(0) = 0 & \text{if } u(1) = 1, \end{cases}$$

and its general solution

$$u(p) = c_1 + c_2 p + \int_0^p \int_0^z f(t)dtdz.$$

The boundary conditions lead to $c_1 = 0$ and $c_2 = 1 - \int_0^1 \int_0^z f(t)dtdz$. By integration by parts, the latter can be rewritten as $c_2 = 1 + \int_0^1 (z-1)f(z)dz$. Hence one has

$$u(p) = p\left\{1 + \int_0^1 (z-1)f(z)dz\right\} + \int_0^p \int_0^z f(t)dtdz$$

which, integrated by parts, leads to

$$u(p) = p + p\int_0^1 (z-1)f(z)dz + \int_0^p (p-z)f(z)dz$$

and, by straightforward calculation, to

$$u(p) = p + (p-1)\int_0^p zf(z)dz + p\int_p^1 (z-1)f(z)dz.$$

A.2. About the bijection ψ

We prove that the function ψ is bijective and, as a by product, that the diagram (6) commutes. Thanks to standard results on bijective functions (see Chapter 1 in [28]), it is enough to show that $\psi^{-1}\{\psi(L)\} = L$ and $\psi\{\psi^{-1}(g)\} = g$ for any $g \in \mathcal{L}_c^2$ and $L \in \{L \in \mathcal{L}or,\ \ln L'' \in \mathcal{L}^2_{[0,1]}\}$.

Concerning the first equality, by Eq. (7), one has

$$\psi^{-1}\{\psi(L)\} = BVP\left[\frac{\exp\{-\psi(L)\}}{\kappa_{\psi(L)}}\right],$$

where $\kappa_{\psi(L)} = \int_0^1 \int_0^p \exp[-\psi\{L(z)\}]dzdp$. Expliciting $\psi(L) = \ln L'' - \int \ln L''$ and simplifying, one gets

$$\psi^{-1}\{\psi(L)\} = BVP\left\{\frac{L''}{\int_0^1 \int_0^p L''(z)dzdp}\right\}$$

which equals L because

$$\int_0^1 \int_0^p L''(z)dzdp = L(1) - L(0) - Q(0) = 1,$$

thanks to the definition of L, the fact that $L'(p) = Q(p)/\mu$ and $Q(0) = 0$. Note that all the densities share $[0, b]$ as a common support.

Concerning the second equality, by definition of ψ and ψ^{-1}, one has

$$\begin{aligned} \psi\{\psi^{-1}(g)\} &= -\ln\left[D^2\{\psi^{-1}(g)\}\right] + \int_0^1 \ln\left[D^2[\psi^{-1}\{g(p)\}]\right]dp \\ &= -\ln\left[D^2\left[BVP\left\{\frac{\exp(-g)}{\kappa_g}\right\}\right]\right] + \int_0^1 \ln\left(D^2\left[BVP\left[\frac{\exp\{-g(p)\}}{\kappa_g}\right]\right]\right)dp. \end{aligned}$$

Recalling that $D^2\{BVP(g)\} = g$, the result follows directly. □

A.3. Proof of Proposition 1

To prove that, as $n \to \infty$, $\widehat{\omega}_{j,n}(k,p) \to \omega_j(k,p)$ in probability for a fixed j and given k (we will drop k in the following expression), first observe that

$$\|\widehat{\omega}_{j,n} - \omega_j\| \leq \|\omega_{j,n} - \omega_j\| + \|\widehat{\omega}_{j,n} - \omega_{j,n}\|. \tag{A.1}$$

The result is obtained by showing that the two terms on the right-hand side of (A.1) tend to zero in probability as $n \to \infty$.

First term. For the first summand on the right-hand side of (A.1), definitions and triangular inequalities lead to

$$\begin{aligned}\|\omega_{j,n} - \omega_j\| &= \|\ell_n \pm k\xi_{j,n}(\lambda_{j,n})^{1/2} - \{\ell \pm k\xi_j(\lambda_j)^{1/2}\}\| \\ &\leq \|\ell_n - \ell\| + k\{(\lambda_{j,n})^{1/2}\|\xi_{j,n} - \xi_j\| + \|\xi_j\| \times |(\lambda_{j,n})^{1/2} - (\lambda_j)^{1/2}|\},\end{aligned} \tag{A.2}$$

where, to avoid identification problems, we have supposed that $\langle \xi_{j,n}, \xi \rangle$ is positive, i.e., $\xi_{j,n}$ and ξ_j point in the same direction.

For the first summand in (A.2), the Strong Law of Large Numbers guarantees that for any $p \in (0,1)$ and $n \to \infty$, $\ell_n(p) \to \ell(p)$ in probability, and, since ℓ_n and ℓ are bounded, then $\|\ell_n - \ell\|^2 \to 0$ in probability. Now consider the remaining terms in (A.2), viz.

$$k\{(\lambda_{j,n})^{1/2}\,\|\xi_{j,n} - \xi_j\| + \|\xi_j\| \times |(\lambda_{j,n})^{1/2} - (\lambda_j)^{1/2}|\}.$$

From now on, denote by C a universal positive constant. Using the fact that $\|\xi_j\| = 1$, $\lambda_{1,n} \geq \lambda_{j,n}$, $\mathrm{E}(\|\ell\|^4) < \infty$ (given by the boundedness of LCs), standard consistency results for the eigenelements $(\lambda_{j,n}, \lambda_{j,n})$ of the empirical covariance operator (see, e.g., Chapter 4 in [6]) apply and

$$|(\lambda_{j,n})^{1/2} - (\lambda_j)^{1/2}| \leq C\,\|\Sigma_n - \Sigma_L\|_\infty\,, \quad \|\xi_{j,n} - \xi_j\| \leq \frac{2\sqrt{2}}{\min(\lambda_{j-1} - \lambda_j,\, \lambda_j - \lambda_{j+1})}\,\|\Sigma_n - \Sigma_L\|_\infty\,.$$

Hence,

$$k\{(\lambda_{j,n})^{1/2}\,\|\xi_{j,n} - \xi_j\| + \|\xi_j\| \times |(\lambda_{j,n})^{1/2} - (\lambda_j)^{1/2}|\} \leq kC\,\|\Sigma_n - \Sigma_L\|_\infty$$

and the right-hand side tends to 0 in probability as $n \to \infty$.

Second term. Consider now the second summand in (A.1), i.e., $\|\widehat{\omega}_{j,n} - \omega_{j,n}\|$. As before, for fixed k and j, one has

$$\|\widehat{\omega}_{j,n} - \omega_{j,n}\| \leq \|\widehat{\ell}_n - \ell_n\| + k\{(\widehat{\lambda}_{j,n})^{1/2}\|\widehat{\xi}_{j,n} - \xi_{j,n}\| + \|\widehat{\xi}_{j,n}\| \times |(\widehat{\lambda}_{j,n})^{1/2} - (\lambda_{j,n})^{1/2}|\} \tag{A.3}$$

and

$$\|\widehat{\ell}_n - \ell_n\|^2 = \int_0^1 \left\{\frac{1}{n}\sum_{i=1}^n \widehat{L}_i(p) - \frac{1}{n}\sum_{i=1}^n L_i(p)\right\}^2 dp \leq \frac{1}{n}\sum_{i=1}^n \|\widehat{L}_i - L_i\|^2.$$

In order to study the asymptotic behavior of $\|\widehat{L}_i - L_i\|$, we fix $\gamma = \gamma_0$ (the dependence on γ_0 will appear when necessary) and assume (A1)–(A3). For each i, when $n_i \to \infty$, the theorem on p. 114 in [11] states that there exists a positive and finite constant depending on γ_0, namely $c(\gamma_0)$, such that

$$\sup_{p\in[0,1]} |(\widehat{L}_i - L_i)(\alpha_0, p)| \leq c(\gamma_0)\sqrt{(\ln\ln n_i)/n_i} \ \text{ a.s.}$$

Using the same arguments as in the proof of Lemma 2.3 in [11], and thanks to Assumption (A2), we have $c(\gamma_0) \leq \int_0^\infty x^2 dF(\gamma_0, x) \leq \Lambda$ and

$$\|\widehat{L}_i - L_i\| \leq \Lambda\sqrt{(\ln\ln n_i)/n_i} \ \text{ a.s.}$$

The latter, together with (A3), gives

$$\|\widehat{L}_i - L_i\| \leq c\sqrt{(\ln\ln n_i)/n^{2\delta}} \ \text{ a.s.} \tag{A.4}$$

and

$$\frac{1}{n}\sum_{i=1}^n \|\widehat{L}_i - L_i\| \to 0 \ \text{ a.s.}, \tag{A.5}$$

which guarantees that the first term in (A.3) vanishes as $n \to \infty$.

Consider now the remaining terms in (A.3), i.e.,

$$(\widehat{\lambda}_{j,n})^{1/2}\|\widehat{\xi}_{j,n} - \xi_{j,n}\| + \|\widehat{\xi}_{j,n}\| \times |(\widehat{\lambda}_{j,n})^{1/2} - (\lambda_{j,n})^{1/2}|.$$

Given that $(\widehat{\lambda}_{j,n})^{1/2} \leq (\widehat{\lambda}_{1,n})^{1/2}$ a.s. and $\|\widehat{\xi}_{j,n}\| = 1$,

$$(\widehat{\lambda}_{j,n})^{1/2}\|\widehat{\xi}_{j,n} - \xi_{j,n}\| + \|\widehat{\xi}_{j,n}\| \times |(\widehat{\lambda}_{j,n})^{1/2} - (\lambda_{j,n})^{1/2}| \leq C\,\|\widehat{\Sigma}_n - \Sigma_n\|_\infty,$$

where

$$\|\widehat{\Sigma}_n - \Sigma_n\|_\infty = \sup_{\|v\|=1} \left\| \frac{1}{n}\sum_{i=1}^{n} \{\langle \widehat{L}_i - \widehat{\ell}_n, v\rangle(\widehat{L}_i - \widehat{\ell}_n) - \langle L_i - \ell_n, v\rangle(L_i - \ell_n)\} \right\|.$$

Note that

$$\begin{aligned}\langle \widehat{L}_i - \widehat{\ell}_n, v\rangle(\widehat{L}_i - \widehat{\ell}_n) - \langle L_i - \ell_n, v\rangle(L_i - \ell_n) = {} & \langle \widehat{L}_i - \widehat{\ell}_n, v\rangle(\widehat{L}_i - L_i) + \langle \widehat{L}_i - \widehat{\ell}_n, v\rangle(\ell_n - \widehat{\ell}_n) + \\ & + \langle \widehat{L}_i - L_i, v\rangle(L_i - \ell_n) + \langle \ell_n - \widehat{\ell}_n, v\rangle(L_i - \ell_n).\end{aligned}$$

The second and the fourth term sum to zero (due to the fact that the sum of the deviations from the mean is zero). Applying the triangular and Cauchy–Schwarz inequalities, and using the fact that $\|v\| = 1$, we get

$$\|\widehat{\Sigma}_n - \Sigma_n\|_\infty \leq \frac{C}{n}\sum_{i=1}^{n}(\|\widehat{L}_i - \widehat{\ell}_n\| \times \|\widehat{L}_i - L_i\| + \|L_i - \ell_n\| \times \|\widehat{L}_i - L_i\|) \leq \frac{C}{n}\sum_{i=1}^{n}\|\widehat{L}_i - L_i\| \ \text{ a.s.}, \tag{A.6}$$

where the last inequality holds because $\|\widehat{L}_i - \widehat{\ell}_n\|$ and $\|L_i - \ell_n\|$ are almost surely bounded. Thank to (A.5), we get the desired conclusion. □

A.4. Proof of Proposition 2

We have to prove that for a fixed integer $j \in \mathbb{N}$ and $0 \leq k < \infty$, when $n \to \infty$, one has

$$\widehat{m}_{j,n}(k,p) \to m_j(k,p) \ \text{ in probability} \tag{A.7}$$

and

$$\widehat{M}_{j,n}(k,p) \to M_j(k,p) \ \text{ in probability}, \tag{A.8}$$

which is equivalent to proving that, as $n \to \infty$,

$$\psi^{-1}(\widehat{m}_{j,n}) = \widehat{M}_{j,n}(k,p) \to \psi^{-1}(m_{j,n}) = M_j(k,p) \ \text{ in probability.}$$

Given that $\psi^{-1}(g) = BVP\{\exp(-g)/\kappa_g\}$ is continuous with respect to g, (A.8) is a consequence of (A.7) and thus we only prove (A.7).

For a given k, by the triangular inequality, we have (dropping the dependence on k)

$$\|\widehat{m}_{j,n} - m_j\| \leq \|m_{j,n} - m_j\| + \|\widehat{m}_{j,n} - m_{j,n}\|. \tag{A.9}$$

The result is obtained by showing that the two terms on the right-hand side of (A.9) converge to zero in probability, as $n \to \infty$.

First term. For the first summand in (A.9), one has

$$\begin{aligned}\|m_{j,n} - m_j\| &= \|\psi_n \pm k\nu_{j,n}(\alpha_{j,n})^{1/2} - \{\psi \pm k\nu_j(\alpha_j)^{1/2}\}\| \\ &\leq \|\psi_n - \psi\| + k\{(\alpha_{j,n})^{1/2}\|\nu_{j,n} - \nu_j\| + \|\nu_j\| \times |(\alpha_{j,n})^{1/2} - \sqrt{\alpha_j}|\},\end{aligned} \tag{A.10}$$

where, to avoid identification problems, we take $\nu_{j,n}$ such that $\langle \nu_{j,n}, \nu_j\rangle$ is positive.

Concerning the first summand in (A.10), the Strong Law of Large Numbers guarantees that for any $p \in (0,1)$, $\psi_n(p) \to \psi(p)$ in probability as $n \to \infty$. Thanks to Assumption (A4), $\psi_n \in \mathcal{L}^2_{[0,1]}$ (see technical Lemma 3, presented at the end of this proof to improve readability) and thus, when $n \to \infty$, $\|\psi_n - \psi\|^2 \to 0$ in probability.

Consider now the remaining terms of (A.10), viz.

$$k\{(\alpha_{j,n})^{1/2}\,\|\nu_{j,n} - \nu_j\| + \|\nu_j\| \times |(\alpha_{j,n})^{1/2} - (\alpha_j)^{1/2}|\}.$$

From now on, denote by C a universal positive constant. Using the fact that $\|\nu_j\| = 1$, and $\alpha_{1,n} \geq \alpha_{j,n}$ together with standard consistency results for the eigenelements $(\alpha_{j,n}, \nu_{j,n})$ of the empirical covariance operator (see Chapter 4 in [6]), we have

$$|(\alpha_{j,n})^{1/2} - (\alpha_j)^{1/2}| \leq C\|\Sigma_{\Psi,n} - \Sigma_\Psi\|_\infty, \quad \|\nu_{j,n} - \nu_j\| \leq \frac{2\sqrt{2}}{\min(\alpha_{j-1} - \alpha_j, \alpha_j - \alpha_{j+1})}\,\|\Sigma_{\Psi,n} - \Sigma_\Psi\|_\infty,$$

and, if $\mathrm{E}(\|\Psi\|^4) < \infty$, as $n \to \infty$,

$$|(\alpha_{j,n})^{1/2} - (\alpha_j)^{1/2}| \to 0 \quad \text{and} \quad \|\nu_{j,n} - \nu_j\| \to 0 \ \text{ in probability.}$$

Hence, to prove the boundedness of the fourth moment, note that

$$\begin{aligned}
\mathrm{E}(\|\Psi\|^4) &= \mathrm{E}\left[\left[\int_0^1 \left\{\ln L''(p) - \int_0^1 \ln L''(t)\,dt\right\}^2 dp\right]^2\right] \\
&\le 4\,\mathrm{E}\left[\left[\int_0^1 \ln^2 L''(p)\,dp + \left\{\int_0^1 \ln L''(t)\,dt\right\}^2\right]^2\right] \\
&\le 16\,\mathrm{E}\left\{\int_0^1 \ln^4 L''(p)\,dp\right\} = 16\,\mathrm{E}\left[\int_0^1 \left[\ln\frac{1}{\mu f\{Q(p)\}}\right]^4 dp\right] \\
&\le 64\ln^4\mu + 64\,\mathrm{E}\left[\int_0^1 \ln^4 f\{Q(p)\}\,dp\right] \\
&= C + \mathrm{E}\left[\int_0^1 \ln^4\{f(x)\} f(x)\,dx\right] < \infty,
\end{aligned}$$

where we have combined the definition, the Cauchy–Schwarz and Jensen inequalities, the substitution $x = F^{-1}(p)$ and Assumption (A4).

Second term. For the second summand of (A.9), one has

$$\begin{aligned}
\|\widehat{m}_{j,n} - m_{j,n}\| &= \|\widehat{\psi}_n + k\widehat{\nu}_{j,n}(\widehat{\alpha}_{j,n})^{1/2} - \{\psi_n + k\nu_{j,n}(\alpha_{j,n})^{1/2}\}\| \\
&\le \|\widehat{\psi}_n - \psi_n\| + k\{(\widehat{\alpha}_{j,n})^{1/2}\,\|\widehat{\nu}_{j,n} - \nu_{j,n}\| + \|\nu_{j,n}\| \times |(\widehat{\alpha}_{j,n})^{1/2} - (\alpha_{j,n})^{1/2}|\}.
\end{aligned} \tag{A.11}$$

Consider $\|\widehat{\psi}_n - \psi_n\|$. By the Cauchy–Schwarz inequality, the boundedness of the second derivatives and Lipschitz arguments, one can write

$$\|\widehat{\psi}_n - \psi_n\|^2 \le \frac{4}{n}\sum_{i=1}^n \|\ln\widetilde{L}_i'' - \ln L_i''\|^2 \le \frac{C}{n}\sum_{i=1}^n \|\widetilde{L}_i'' - L_i''\|^2.$$

Because $\widetilde{L}_i$ is a smoothed version of $\widehat{L}_i$, then proving that $\|\widetilde{L}_i'' - L_i''\| \to 0$ in probability, as $n_i(n) \to \infty$, guarantees that $\|\widehat{\psi}_n - \psi_n\| \to 0$ as $n \to \infty$.

Let $B_{i1}, \ldots, B_{ir}$ be a B-spline basis of degree $\nu \ge 2$ and τ_i equispaced knots, $r_i = \nu + \tau_i$. Then

$$\widetilde{L}_i''(p) = \sum_{j=1}^{r_i} \widetilde{s}_{ij} B_{ij}''(p),$$

where

$$\widetilde{\mathbf{s}}_i = \widetilde{\mathbf{b}}_i^\top \widetilde{\mathbf{C}}_i^{-1}, \quad \widetilde{\mathbf{C}}_i^{-1} = \frac{1}{n_i}\sum_{j=1}^{n_i} \mathbf{B}_i(p_j)\mathbf{B}_i^\top(p_j), \quad \widetilde{\mathbf{b}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \mathbf{B}_i(p_j)\widehat{L}_i(p_j)$$

and $\mathbf{B}_i(p)$ being the B-spline vector. Moreover, denote by $\overline{L}_i''$ the smoothed version of L_i'', i.e.,

$$\overline{L}_i''(p) = \sum_{j=1}^{r_i} \overline{s}_{ij} B_{ij}''(p),$$

where $\overline{\mathbf{s}}_i = \overline{\mathbf{b}}_i^\top \widetilde{\mathbf{C}}_i^{-1}$ and $\overline{\mathbf{b}}_i = \sum_{j=1}^{n_i} \mathbf{B}_i(p_j)L_i(p_j)/n_i$ and consider the bound

$$\|\widetilde{L}_i'' - L_i''\| \le \|\widetilde{L}_i'' - \overline{L}_i''\| + \|\overline{L}_i'' - L_i''\|. \tag{A.12}$$

Concerning the first summand on the right-hand side of (A.12), note that

$$\|\widetilde{L}_i'' - \overline{L}_i''\|^2 \le \|\widetilde{\mathbf{b}}_i - \overline{\mathbf{b}}_i\|^2 \times \|\widetilde{\mathbf{C}}_i^{-1}\|^2 \times \|\mathbf{G}_i\|^2,$$

where $[\mathbf{G}_i]_{\ell m} = \int B_{i\ell}''(p)B_{im}''(p)dp$. In view of the definitions of $\widetilde{\mathbf{b}}_i$ and $\overline{\mathbf{b}}_i$, and given (A.4), one has

$$\|\widetilde{\mathbf{b}}_i - \overline{\mathbf{b}}_i\|^2 \le C\,\|\widehat{L}_i - L_i\|^2/\tau_i \le (Cn_i^{-2\delta}\ln\ln n_i)/\tau_i.$$

Using standard results on B-splines for functions observed on a regular discretization grid (see, e.g., Lemma 6.2 in [7]), $\|\widetilde{\mathbf{C}}_i^{-1}\|^2 = O(\tau_i^{-2})$ and $\|\mathbf{G}\|^2 = O(\tau_i^3)$. Hence, when $n_i \to \infty$,

$$\|\widetilde{L}_i'' - \overline{L}_i''\|^2 = O(n_i^{-2\delta}\ln\ln n_i) \ \text{ a.s.} \tag{A.13}$$

For the second summand on the right-hand side of (A.12), choosing $\tau_i = o(n_i)$ and due to Lemma 6.2 in [7], one has

$$\|\bar{L}_i'' - L_i''\|^2 = O(\tau_i^{-4}). \tag{A.14}$$

Thus the bounds (A.13) and (A.14) guarantee that as $n \to \infty$, $\|\widehat{\psi}_n - \psi_n\| \to 0$ in probability.

Consider now the remaining terms in (A.11):

$$(\widehat{\alpha}_{j,n})^{1/2}\,\|\widehat{\nu}_{j,n} - \nu_{j,n}\| \le (\widehat{\alpha}_{1,n})^{1/2}\|\widehat{\nu}_{j,n} - \nu_{j,n}\| \le C\,\|\widehat{\Sigma}_{\widehat{\Psi},n} - \Sigma_{\Psi,n}\|_\infty,$$
$$\|\nu_{j,n}\| \times |(\widehat{\alpha}_{j,n})^{1/2} - (\alpha_{j,n})^{1/2}| \le C\,\|\widehat{\Sigma}_{\Psi,n} - \Sigma_{\Psi,n}\|_\infty,$$

where

$$\widehat{\Sigma}_{\Psi,n}[\cdot] = \frac{1}{n}\sum_{i=1}^{n}\langle\widehat{\Psi}_i - \widetilde{\psi}_n, \cdot\rangle(\widehat{\Psi}_i - \widetilde{\psi}_n), \quad \Sigma_{\Psi,n}[\cdot] = \frac{1}{n}\sum_{i=1}^{n}\langle\Psi_i - \widehat{\psi}_n, \cdot\rangle(\Psi_i - \widehat{\psi}_n).$$

Analogously to what was done to derive (A.6), one gets

$$\|\widehat{\Sigma}_{\Psi,n} - \Sigma_{\Psi,n}\|_\infty \le \frac{C}{n}\sum_{i=1}^{n}(\|\widehat{\Psi}_i - \widehat{\psi}_n\| \times \|\widehat{\Psi}_i - \Psi_i\| + \|\Psi_i - \psi_n\| \times \|\widehat{\Psi}_i - \Psi_i\|) \le \frac{C}{n}\sum_{i=1}^{n}\|\widehat{\Psi}_i - \Psi_i\| \quad \text{a.s.}$$

Since the fourth moment of Ψ is bounded, by using B-spline and similar arguments as above, one has $\|\widehat{\Psi}_i - \Psi_i\|$ tends to zero in probability as $n \to \infty$ and thus

$$\|\widehat{\Sigma}_{\Psi,n} - \Sigma_{\Psi,n}\|_\infty \to 0 \quad \text{in probability.}$$

Lemma 3. *If (A4) holds, then* $\psi_n \in \mathcal{L}^2_{[0,1]}$.

Proof. By definition of $\psi_n(p)$ and applying the Cauchy–Schwarz inequality, we get

$$\|\psi_n\|^2 = \int_0^1\left\{\frac{1}{n}\sum_{i=1}^{n}\Psi_i(p)\right\}^2 dp \le \frac{1}{n^2}\sum_{i=1}^{n} n\int_0^1 \Psi_i(p)^2 dp = \frac{1}{n}\sum_{i=1}^{n}\left\|\ln L_i''(p) - \int_0^1 \ln L_i''(t)\,dt\right\|^2$$
$$\le \frac{2}{n}\sum_{i=1}^{n}\left[\|\ln L_i''(p)\|^2 + \left\{\int_0^1 \ln L_i''(t)\,dt\right\}^2\right].$$

Thanks to Jensen's inequality,

$$\left\{\int_0^1 \ln L_i''(t)\,dt\right\}^2 \le \int_0^1 \left\{\ln L_i''(t)\right\}^2 dt$$

and, thanks to (4),

$$\|\psi_n\|^2 \le \frac{4}{n}\sum_{i=1}^{n}\int_0^1\left\{\ln L_i''(t)\right\}^2 dt = \frac{4}{n}\sum_{i=1}^{n}\int_0^1\left[\ln\frac{1}{\mu_i f_i\{Q_i(p)\}}\right]^2 dp$$
$$= \frac{4}{n}\sum_{i=1}^{n}\ln^2\mu_i + \frac{4}{n}\sum_{i=1}^{n}\int_0^1 \ln^2 f_i\{Q_i(p)\}\,dp.$$

For each $i \in \{1, \ldots, n\}$, substitute $x = F_i^{-1}(p)$ to get

$$\|\psi_n\|^2 \le C + \frac{4}{n}\sum_{i=1}^{n}\int_0^1 f_i(x)\ln^2 f_i(x)dx$$

which is finite, thanks to (A4). This concludes the proof. □

References

[1] R. Aaberge, Ranking intersecting Lorenz curves, Soc. Choice Welf. 33 (2009) 235–259.
[2] J. Aitchison, The Statistical Analysis of Compositional Data, Chapman & Hall, London, 1986.
[3] G. Aneiros, E.G. Bongiorno, R. Cao, P. Vieu, Functional Statistics and Related Fields, Springer, Berlin, 2017.
[4] G. Betti, A. Lemmi, Advances on Income Inequality and Concentration Measures, Routledge, London, 2008.
[5] E.G. Bongiorno, A. Goia, E. Salinelli, P. Vieu, Contributions in Infinite-Dimensional Statistics and Related Topics, Esculapio, Bologna, 2014.
[6] D. Bosq, Linear Processes in Function Spaces, Springer, New York, 2000.
[7] H. Cardot, Nonparametric estimation of smoothed principal components analysis of sampled noisy functions, J. Nonparametr. Stat. 12 (2000) 503–538.
[8] D. Chotikapanich, Modeling Income Distributions and Lorenz Curves, Springer-Verlag, New York, 2008.

[9] M. Csörgő, H. Yu, Weak approximations for empirical Lorenz curves and their Goldie inverses of stationary observations, Adv. Appl. Probab. 31 (1999) 698–719.
[10] M. Csörgő, R. Zitikis, Strassen's LIL for the Lorenz curve, J. Multivariate Anal. 59 (1996) 1–12.
[11] M. Csörgő, R. Zitikis, On the rate of strong consistency of Lorenz curves, Statist. Probab. Lett. 34 (1997) 113–121.
[12] A. Cuevas, A partial overview of the theory of statistics with functional data, J. Statist. Plann. Inference 147 (2014) 1–23.
[13] C. de Boor, A Practical Guide to Splines, revised ed., Springer, New York, 2001.
[14] P. Delicado, Functional k-sample problem when data are density functions, Comput. Statist. 22 (2007) 391–410.
[15] P. Delicado, Dimensionality reduction when data are density functions, Comput. Statist. Data Anal. 55 (2011) 401–420.
[16] P. Delicado, P. Vieu, Optimal level sets for bivariate density representation, J. Multivariate Anal. 140 (2015) 1–18.
[17] J. Fan, I. Gijbels, Local Polynomial Modelling and its Applications, Chapman & Hall, London, 1996.
[18] F. Ferraty, P. Vieu, Nonparametric Functional Data Analysis, Springer, New York, 2006.
[19] J.L. Gastwirth, A general definition of the Lorenz curve, Econometrica (1971) 1037–1039.
[20] A. Goia, P. Vieu, An introduction to recent advances in high/infinite dimensional statistics, J. Multivariate Anal. 146 (2016) 1–6.
[21] C.M. Goldie, Convergence theorems for empirical Lorenz curves and their inverses, Adv. Appl. Probab. 9 (1977) 765–791.
[22] L. Horváth, P. Kokoszka, Inference for Functional Data with Applications, Springer, New York, 2012.
[23] K. Hron, A. Menafoglio, M. Templ, K. Hruzová, P. Filzmoser, Simplicial principal component analysis for density functions in bayes spaces, Comput. Statist. Data Anal. 94 (2016) 330–350.
[24] J. Iritani, K. Kuga, Duality between the Lorenz curves and the income distribution functions, Econ. Stud. Q. (Tokyo. 1950) 34 (1983) 9–21.
[25] N.C. Kakwani, N. Podder, Efficient estimation of the lorenz curve and associated inequality measures from grouped observations, Econometrica (1976) 137–148.
[26] A. Kneip, K.J. Utikal, Inference for density families using functional principal component analysis, J. Amer. Statist. Assoc. 96 (2001) 519–542.
[27] M.O. Lorenz, Methods of measuring the concentration of wealth, Publ. Am. Stat. Assoc. 9 (1905) 209–219.
[28] G. McCarty, Topology, second ed., Dover Publications, New York, 1988.
[29] M.C. Meyer, Constrained penalized splines, Canad. J. Statist. 40 (2012) 190–206.
[30] D. Nerini, B. Ghattas, Classifying densities using functional regression trees: applications in oceanology, Comput. Statist. Data Anal. 51 (2007) 4984–4993.
[31] E. Parzen, Density quantile estimation approach to statistical data modelling, in: Smoothing Techniques for Curve Estimation, Heidelberg, 1979, in: Proc. Workshop, Springer, Berlin, 1979, pp. 155–180.
[32] A. Petersen, H.-G. Müller, Functional data analysis for density functions by transformation to a Hilbert space, Ann. Statist. 44 (2016) 183–218.
[33] J.O. Ramsay, B.W. Silverman, Functional Data Analysis, second ed., Springer, New York, 2005.
[34] A.F. Shorrocks, Ranking income distributions, Economica 50 (1983) 3–17.
[35] P.D. Thistle, Duality between generalized Lorenz curves and distribution functions, Econ. Stud. Q. 40 (1989) 183–187.
[36] J.W. Tukey, Which part of the sample contains the information?, Proc. Nat. Acad. Sci. USA 53 (1965) 127–134.
[37] M.P. Wand, M.C. Jones, Kernel Smoothing, Chapman & Hall, London, 1995.
[38] Z. Zhang, H.-G. Müller, Functional density synchronization, Comput. Statist. Data Anal. 55 (2011) 2234–2249.
[39] Y.Y. Zhang, X. Wu, Q. Li, A simple consistent nonparametric estimator of the Lorenz curve, in: Essays in Honor of Aman Ullah, Emerald Group Publishing Limited, 2016, pp. 635–653.