Non parametric estimation of the structural expectation of a stochastic increasing function

Jean-François Dupuy, Jean-Michel Loubes, Elie Maza

DOI 10.1007/s11222-009-9152-9

Non parametric estimation of the structural expectation of a stochastic increasing function

J.-F. Dupuy · J.-M. Loubes · E. Maza

Abstract This article introduces a non parametric warping model for functional data. When the outcome of an experiment is a sample of curves, data can be seen as realizations of a stochastic process, which takes into account the variations between the different observed curves. The aim of this work is to define a mean pattern which represents the main behaviour of the set of all the realizations. So, we define the structural expectation of the underlying stochastic function. Then, we provide empirical estimators of this structural expectation and of each individual warping function. Consistency and asymptotic normality for such estimators are proved.

Keywords Functional data analysis · Non parametric warping model · Structural expectation · Curve registration

1 Introduction and main concept

When dealing with functional data, one main issue is how to extract information from a sample of curves. Indeed, curves usually not only present amplitude variability (a variation in the y-axis) but also time or phase variability (a variation in the x-axis). In this case, the classical cross-sectional mean does not make any sense, and defining a mean curve is even not obvious. Giving a sense to the common behaviour of a sample of curves, and finding a meaningful mean curve in this context thus constitute important issues, also encountered in curve registration, or time warping problems (see for example the engineering literature, Sakoe and Chiba 1978).

In this framework, we observe $i = 1, \dots, m$ curves $f_i : [a, b] \to \mathbb{R}$ at discrete times $t_{ij} \in [a, b]$, $j = 1, \dots, n$, taken, for sake of simplicity, equally-spaced for all the individuals, and for this, denoted t_j. The data can thus be written as

$$Y_{ij} = f_i(t_j), \quad i = 1, \dots, m, \ j = 1, \dots, n. \tag{1}$$

The registration problem aims at finding a mean pattern f and warping functions h_i which align the observed curves to the template f, i.e such that $f = f_i \circ h_i$. Hence, each curve f_i is obtained by warping the original curve f using the warping functions h_i^{-1}.

The registration operator can be modeled using a random warping procedure, which takes account of the variability of the deformation as a random effect. For example, in Rønn (2001), the author considers a model for randomly shifted curves, where the shifts are assumed to be distributed according to a known parametric distribution with mean 0 and unknown variance. In Gervini and Gasser (2005), the authors propose a random warping model where the warping process is modeled through a known increasing function and a parametric random effect. In our work, we aim at providing a purely non parametric, and therefore more flexible model

J.-F. Dupuy (✉)
Laboratoire Mathématiques, Image et Applications,
Université de La Rochelle, Avenue Michel Crépeau,
17000 La Rochelle, France
e-mail: jean-francois.dupuy@univ-lr.fr

J.-M. Loubes
Université Toulouse 3, Institut de Mathématiques (UMR 5219),
Route de Narbonne, 31062 Toulouse, France
e-mail: jean-michel.loubes@math.univ-toulouse.fr

E. Maza
Laboratoire Génomique et Biotechnologie des Fruits,
INP-ENSAT (UMR 990), 31326 Castanet-Tolosan, France
e-mail: Elie.Maza@ensat.fr

for randomly warped curves. Let H be a warping stochastic process defined as

$$\begin{aligned} H : \Omega &\to \mathcal{C}([a,b]), \\ w &\mapsto H(w,\cdot), \end{aligned} \tag{2}$$

where $(\Omega, \mathcal{A}, \mathbf{P})$ is an unknown probability space, and $(\mathcal{C}([a,b]), \|\cdot\|_\infty, \mathcal{B})$ is the set of all real-valued continuous functions defined on the interval $[a,b]$, equipped with the uniform norm and with its Borel algebra. Consider $h_1, \ldots, h_m$ as i.i.d. realizations of the process H. Then h_i warps a mean pattern f onto the i-th observation curve f_i. Hence, model (1) can be written as

$$Y_{ij} = f_i(t_j) = f \circ h_i^{-1}(t_j). \tag{3}$$

We point out that h_i^{-1} is well defined since the warping processes are assumed to be continuous increasing functions.

Without any constraint on model (3), the problem of recovering f and H from the Y_{ij} has many solutions, since model (3) is not identifiable. More precisely, the unknown function f and the unknown warping process H cannot be uniquely estimated. Indeed, if $\tilde{h} : [a,b] \to [a,b]$ is an increasing continuous function, then for all $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, n\}$, we have that $Y_{ij} = f \circ \tilde{h}^{-1} \circ \tilde{h} \circ h_i^{-1}(t_j)$. Hence, the function $f \circ \tilde{h}^{-1}$, associated with the warping process $H \circ \tilde{h}^{-1}$, is also a solution of model (3).

To solve this issue, one popular method consists of the following two-stage procedure: (i) first align the curves to a given template (the first curve or the mean of the observed curves are usually chosen as the template) by warping the time axis, (ii) take the mean of the sample of dewarped curves. Such methods have become increasingly common in statistics, and we refer to Ramsay and Silverman (2002) for a review. A landmark registration approach is proposed by Kneip and Gasser (1992) and further developed by Bigot (2005), while a non parametric method based on local regressions is investigated in Kneip et al. (2000) and Ramsay and Li (1998). A dynamic time warping methodology is developed by Wang and Gasser (1999). An alternative approach is provided by Gamboa et al. (2007) and Loubes et al. (2006), where semi-parametric estimation of shifted curves is investigated. But these methods all imply choosing a starting curve as a fixed point for the alignment process. This initial choice may either bias the estimation procedure, or imply strong and restrictive identifiability conditions.

In this work, an alternative point of view is considered to overcome the identifiability problem stated above. Precisely we define, directly from the data, an archetype representing the common behaviour of the sample curves, without stressing any particular curve but taking into account the information conveyed by the warping process itself. One advantage of this method is that we do not assume any technical restriction on the data, which enables us to handle a large variety of cases. Letting ϕ be the expectation of the warping process H, we define the *structural expectation* f_{ES} as

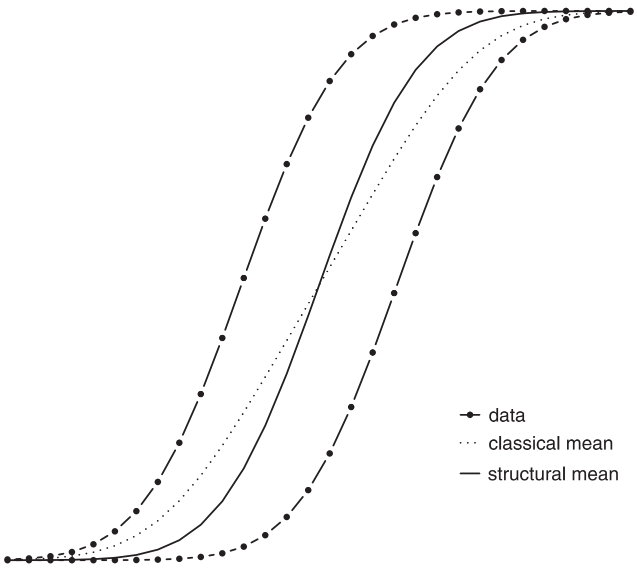

Fig. 1 Classical and structural mean of a two-curves sample

$$f_{ES} := f \circ \phi^{-1}.$$

Note that each curve can be warped to this feature by taking, for all $i = 1, \ldots, m$,

$$f_i = f_{ES} \circ \phi \circ h_i^{-1} = f_{ES} \circ h_{ES,i}^{-1}.$$

Figure 1 illustrates the intuitive meaning of this definition, and relates it to alternative points of view. Consider the case of a two-curves sample. In the classical setting of curve registration, the curves would be aligned either onto the classical mean, or onto one of the two observed curves. Choosing the usual mean curve as the representative feature leads to a wrong pattern (the dotted line), which is estimated at a good location. On the contrary, alignment onto the first or second curve yields a good pattern, but does not take account of the deformation. In our approach, the structural mean is obviously representative of the information in the curve sample. Moreover, this approach also conveys information on the warping process.

The structural expectation is obviously not the function f, but the function f composed with the inverse ϕ^{-1} of the expectation of H. Thus, it can be understood as the *mean* warping of the function f by the stochastic process H. Note that since ϕ is monotone, f_{ES} inherits the same pattern as f, and thus as the observed curves f_i. The structural expectation therefore provides a sensible representation for the behaviour of the sample curves f_i. The aim of this paper is to properly define this new archetype, to study its properties, and to propose and investigate estimators for both the structural expectation and warping paths $h_{ES,i}^{-1} = \phi \circ h_i^{-1}$.

The article is organized as follows. In Sect. 2, we define empirical estimators of the structural expectation and of the individual warping functions. Asymptotic properties of these estimators are investigated. Proofs are postponed to Appendix. Section 3 investigates some extensions of the proposed methodology, in particular to the case of noisy non increasing functions. The results of a simulation study are reported in Sect. 4. There, we also apply the proposed estimators to a real dataset.

2 Theoretical study of a particular case: warping of strictly increasing functions

First, consider the case where f is strictly increasing. Hence, the inverse function f^{-1} exists and is also strictly increasing. Moreover, a phase warping of the function f (i.e. on the x-axis) corresponds to an amplitude warping of the function f^{-1} (i.e. on the y-axis). We propose estimators for the inverse of the structural expectation $f_{ES}^{-1} = \phi \circ f^{-1}$, for the individual warping functions $h_{ES,i}^{-1} = \phi \circ h_i^{-1}$, $i \in \{1, \ldots, m\}$, and for the structural expectation $f_{ES} = f \circ \phi^{-1}$.

Note that all the asymptotic results are taken with respect to m and n, so we recall that $u(m,n) \underset{m,n\to\infty}{\longrightarrow} c$ if, and only if, we have

$$\forall \epsilon > 0, \exists (m_0, n_0) \in \mathbb{N}^2, \forall (m,n) \in \mathbb{N}^2, m > m_0 \text{ and}$$
$$n > n_0 \Rightarrow |u(m,n) - c| < \epsilon.$$

In order to define a good registration procedure, we assume that the warping process does not change the time line (time inversion is not allowed), and leaves fixed the two extreme points. Precisely, for almost all $w \in \Omega$, we assume that

(i) $H(w, \cdot)$ is an increasing function,
(ii) $H(w, a) = a$ and $H(w, b) = b$.

The following proposition introduces respectively the expectation, second order moment and covariance functions of H.

Proposition 2.1 *Under (2), the expectation $\phi(\cdot)$, the second order moment $\gamma(\cdot)$ and the covariance function* $\mathrm{r}(\cdot,\cdot)$ *of the stochastic process H are well defined. ϕ and γ are also continuous increasing functions. Moreover, we have that $\phi(a) = a$, $\phi(b) = b$, $\gamma(a) = a^2$ and $\gamma(b) = b^2$. As a consequence, we have that* $\operatorname{var} H(\cdot, a) = \operatorname{var} H(\cdot, b) = 0$.

Proof The process H is bounded and increasing. As a consequence, ϕ and γ exist. Moreover, H is a continuous increasing process, which leads to continuous and increasing first and second order moments. □

In order to prove asymptotic results, the following technical assumptions on the warping process H and on the function f are needed:

1. There exists a constant $C_1 > 0$ such that for all $(s,t) \in [f(a), f(b)]^2$, we have
$$\mathbf{E}|H(s) - \mathbf{E}H(s) - (H(t) - \mathbf{E}H(t))|^2 \leq C_1 |s-t|^2. \quad (4)$$
2. There exists a constant $C_2 > 0$ such that for all $(s,t) \in [f(a), f(b)]^2$, we have
$$|f^{-1}(s) - f^{-1}(t)|^2 \leq C_2 |s-t|^2. \quad (5)$$
3. There exists a constant $C_3 > 0$ such that for all $\omega \in \Omega$, for all $(s,t) \in [a,b]^2$, we have
$$|H^{-1}(\omega, s) - H^{-1}(\omega, t)|^2 \leq C_3 |s-t|^2. \quad (6)$$

These conditions ensure that the warping process H and the unknown function f are sufficiently regular, so that the observed functions f_i do not behave erratically. Other contributions to random models for warped curves (see Gervini and Gasser 2005; Rønn 2001) also assume some conditions (monotonicity, invertibility of the warping transformation) that ensure a regular behaviour for the observed sample curves. The conditions stated in our work thus provide a reasonable setting for the investigation of randomly warped curves.

2.1 Estimator of the structural expectation f_{ES}

Since $f_i = f \circ h_i^{-1}$, we have $f_i^{-1} = h_i \circ f^{-1}$. Hence $\mathbf{E}(f_i^{-1}) = \mathbf{E}(H) \circ f^{-1}$. Then, it seems natural to consider the mean of the functions f_i^{-1}, $i \in \{1, \ldots, m\}$ in order to estimate the inverse of the structural expectation. For all $y \in [f(a), f(b)]$, and for all $i \in \{1, \ldots, m\}$, define

$$j_i(y) = \arg\min_{j \in \{1,\ldots,n\}} |Y_{ij} - y| \quad \text{and} \quad T_i(y) := t_{j_i(y)}. \quad (7)$$

Then, the empirical estimator of the inverse of the structural expectation is defined by

$$\widehat{f_{ES}^{-1}}(y) = \frac{1}{m} \sum_{i=1}^{m} T_i(y). \quad (8)$$

From (8) and (7), $\widehat{f_{ES}^{-1}}$ is an increasing step function with jumps occurring at $K(m,n)$ points $v_1, \ldots, v_{K(m,n)}$ in $[f(a), f(b)]$, such that $f(a) = v_0 < v_1 < \cdots < v_{K(m,n)} < v_{K(m,n)+1} = f(b)$. Hence, for all $y \in [f(a), f(b)] \backslash (v_k)_{k \in \mathcal{K}}$, $\mathcal{K} = \{0, \ldots, K(m,n)+1\}$, $\widehat{f_{ES}^{-1}}(y)$ can be expressed as

$$\widehat{f_{ES}^{-1}}(y) = \sum_{k=0}^{K(m,n)} u_k \mathbf{1}_{(v_k, v_{k+1})}(y)$$

with $a = u_0 < u_1 < \cdots < u_{K(m,n)-1} < u_{K(m,n)} = b$. A natural estimator of the structural expectation f_{ES} is then obtained by linear interpolation between the points (u_k, v_k). For all $t \in [a, b]$, let define

$$\widehat{f_{ES}}(t) = \sum_{k=0}^{K(m,n)-1} \left(v_k + \frac{v_{k+1} - v_k}{u_{k+1} - u_k}(t - u_k) \right) \mathbf{1}_{[u_k, u_{k+1})}(t) + v_{K(m,n)} \mathbf{1}_{\{b\}}(t).$$

Note that by construction, this estimator is strictly increasing and continuous on $[a, b]$. The following theorem states its consistency.

Theorem 2.2 (Consistency of the estimator of the Structural Expectation) *Under the assumption* (2), *we have*

$$\|\widehat{f_{ES}} - f_{ES}\|_\infty \xrightarrow[m,n\to\infty]{a.s.} 0.$$

Obtaining confidence bands for f_{ES}^{-1} is useful for describing and visualizing the uncertainty in the estimate of f_{ES}^{-1}. This requires finding first the asymptotic behaviour of $\widehat{f_{ES}^{-1}}$, and then providing an estimator of $\operatorname{var}(G(y)) = \gamma \circ f^{-1}(y) - \{f_{ES}^{-1}(y)\}^2$, $y \in [f(a), f(b)]$. The following theorem states the consistency and asymptotic normality of estimator (8).

Lemma 2.3 (Consistency of the inverse of the structural expectation) *Under the assumption* (2),

$$\|\widehat{f_{ES}^{-1}} - f_{ES}^{-1}\|_\infty \xrightarrow[m,n\to\infty]{a.s.} 0.$$

Moreover, let $n = m^{\frac{1}{2}+\alpha}$ *with* $\alpha > 0$, *and assume that conditions* (4) *and* (5) *hold. Then,*

$$\sqrt{m}\left(\widehat{f_{ES}^{-1}} - f_{ES}^{-1}\right) \xrightarrow[m\to\infty]{\mathbf{D}} G,$$

where G is a centered Gaussian process with covariance given by: for all $(s, t) \in [f(a), f(b)]^2$,

$$\operatorname{cov}(G(s), G(t)) = \mathrm{r}\left(f^{-1}(s), f^{-1}(t)\right).$$

Lemma 2.4 *Let* $y \in [f(a), f(b)]$. *Let* $\widehat{\gamma \circ f^{-1}}(y) = \frac{1}{m}\sum_{i=1}^m T_i^2(y)$, *with* $T_i(.)$ *defined as in* (7). *Then*

$$\widehat{\gamma \circ f^{-1}}(y) - \left\{\widehat{f_{ES}^{-1}}(y)\right\}^2 \xrightarrow[m,n\to\infty]{a.s.} \operatorname{var}(G(y)).$$

The proof of this lemma is given in Appendix. Combining this lemma with the asymptotic normality result stated in Lemma 2.3 yields a pointwise asymptotic confidence band for f_{ES}^{-1}.

Corollary 2.5 *An asymptotic* $(1-\alpha)$*-level pointwise confidence band for* f_{ES}^{-1} *is given by*

$$\left[\widehat{f_{ES}^{-1}}(y) - u_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{\operatorname{var}}(G(y))}{m}},\ \widehat{f_{ES}^{-1}}(y) + u_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{\operatorname{var}}(G(y))}{m}}\right],$$

where $\widehat{\operatorname{var}}(G(y)) = \widehat{\gamma \circ f^{-1}}(y) - \{\widehat{f_{ES}^{-1}}(y)\}^2$ *and* $u_{1-\frac{\alpha}{2}}$ *is the quantile of order* $1 - \frac{\alpha}{2}$ *of the standard normal distribution.*

Note that the construction of a simultaneous asymptotic confidence band for f_{ES}^{-1} would require the determination of the distribution of $\sup_{f(a)\le y\le f(b)} |G(y)|$. This, however, falls beyond the scope of this paper.

2.2 Estimator of an individual structural warping function $h_{ES,i}^{-1}$

In a warping framework, it is necessary to estimate the mean pattern, but also the individual warping functions. Indeed, these functions show how far a given curve f_i is from the common mean pattern, and allow us understand the particular time-warping transformation that was applied to f, so as to yield the observed f_i.

Remember that we chose to align the curves to a mean pattern which takes account of the mean warping, hence

$$f_i = f \circ h_i^{-1} = f \circ \phi^{-1} \circ \phi \circ h_i^{-1} = f_{ES} \circ h_{ES,i}^{-1}.$$

Thus, as previously, we cannot directly estimate the functions h_i^{-1}, but only the functions $h_{ES,i}^{-1} = \phi \circ h_i^{-1}$, called structural warping functions.

Let $i_0 \in \{1, \ldots, m\}$. We want to compute $T_i^\star(t) = f_i^{-1} \circ f_{i_0}(t)$, for all $i \neq i_0$. For this, define

$$j_0(t) = \arg\min_{j\in\{1,\ldots,n\}} |t_{i_0 j} - t|. \tag{9}$$

This point is the observation time for the i_0-th curve, which is the closest to t. Note that the index $j_0(t)$ depends on i_0, but for the sake of simplicity, we drop this index in the notations. Then, for all $t \in [a, b]$ and $i \in \{1, \ldots, m\} \setminus i_0$, compute

$$T_i(t) = \arg\min_{t_j\in\{t_{i1},\ldots,t_{in}\}} |Y_{ij} - Y_{i_0 j_0(t)}| \tag{10}$$

as an estimate of $T_i^\star$. Then, for a fixed i_0, noting that $T_i^\star = h_i \circ h_{i_0}^{-1}$, we can see that an empirical estimator of each individual warping function $\phi \circ h_{i_0}^{-1}$ is given by

$$\widehat{h_{ES,i_0}^{-1}}(t) = \widehat{\phi \circ h_{i_0}^{-1}}(t) := \frac{1}{m-1}\sum_{\substack{i=1\\ i\neq i_0}}^m T_i(t). \tag{11}$$

The following theorem asserts the consistency and asymptotic normality of this estimator.

Theorem 2.6 *Under assumption* (2),

$$\|\widehat{h_{ES,i_0}^{-1}} - h_{ES,i_0}^{-1}\|_\infty \xrightarrow[m,n\to\infty]{a.s.} 0.$$

Let $n = m^{\frac{1}{2}+\alpha}$ (*with* $\alpha > 0$) *and assume that* (4) *and* (6) *hold. Then* $\sqrt{m}(\widehat{h_{ES,i_0}^{-1}} - h_{ES,i_0}^{-1})$ *converges weakly to a centered Gaussian process* Z,

$$\sqrt{m}\big(\widehat{h_{ES,i_0}^{-1}} - h_{ES,i_0}^{-1}\big) \xrightarrow[m\to\infty]{\mathbf{D}} Z,$$

with covariance function defined for all $(s,t) \in [a,b]^2$ *by*

$$\mathrm{cov}(Z(s), Z(t)) = \mathrm{r}\big(h_{i_0}^{-1}(s), h_{i_0}^{-1}(t)\big).$$

We may also compute confidence bands for $\phi \circ h_{i_0}^{-1}$, based on a consistent estimator of $\mathrm{var}(Z(t)) = \gamma \circ h_{i_0}^{-1}(t) - \{\phi \circ h_{i_0}^{-1}(t)\}^2$.

Lemma 2.7 *Let* $t \in [a,b]$. *Let* $\widehat{\gamma \circ h_{i_0}^{-1}}(t) = \frac{1}{m}\sum_{i=1}^m T_i^2(t)$, *with* $T_i(.)$ *defined by* (9) *and* (10). *Then*

$$\widehat{\gamma \circ h_{i_0}^{-1}}(t) - \big\{\widehat{\phi \circ h_{i_0}^{-1}}(t)\big\}^2 \xrightarrow[m,n\to\infty]{a.s.} \mathrm{var}(Z(t)).$$

The proof of this lemma relies on the same arguments as the proof of Lemma 2.4, and is outlined in Appendix. A pointwise asymptotic confidence band for $\phi \circ h_{i_0}^{-1}$ is now given by

Corollary 2.8 *An asymptotic* $(1-\alpha)$*-level pointwise confidence band for* $\phi \circ h_{i_0}^{-1}$ *is given by*

$$\Bigg[\widehat{\phi \circ h_{i_0}^{-1}}(t) - u_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{\mathrm{var}}(Z(t))}{m}},\quad \widehat{\phi \circ h_{i_0}^{-1}}(t) + u_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{\mathrm{var}}(Z(t))}{m}}\Bigg],$$

where $\widehat{\mathrm{var}}(Z(t)) = \widehat{\gamma \circ h_{i_0}^{-1}}(t) - \{\widehat{\phi \circ h_{i_0}^{-1}}(t)\}^2$.

3 Extensions to the general case

In the preceding part, we studied the asymptotic behaviour of a new warping methodology. However, drastic restrictions over the class of functions are needed: boundary constraints, monotonicity of the observed functions, and a non noisy model. In this section, we get rid of such assumptions, and we provide a practical way of handling more realistic observations.

3.1 Boundary constraints

First, we note that the assumptions $H(a) \overset{a.s.}{=} a$ and $H(b) \overset{a.s.}{=} b$ imply that the observed measures $f_i(a)$ and $f_i(b)$ are respectively equal for all individuals. These assumptions can be weakened by assuming:

(ii′) $H^{-1}(\cdot, a)$ and $H^{-1}(\cdot, b)$ are compactly supported random variables, with

$$\sup_{w\in\Omega} H^{-1}(w,a) \overset{a.s.}{<} \inf_{w\in\Omega} H^{-1}(w,b).$$

Thus, the observed measures $f_i(a)$ and $f_i(b)$ can vary from individuals to individuals. Obviously in that case, the proposed estimator for the structural expectation f_{ES} is not defined on the whole range $[a,b]$, but on a smaller interval $]a', b'[\subset [a,b]$, with $a' = f_{ES}^{-1}(\sup_i f_i(a))$ and $b' = f_{ES}^{-1}(\inf_i f_i(b))$.

3.2 Breaking monotonicity

The main idea is to build a transformation $\mathcal{G}$ which turns a non monotone function into a monotone one, while preserving the warping functions. For the sake of simplicity, the observation times will be taken equal for all the curves, hence t_{ij} will be denoted t_j. Hence, the observations

$$Y_{ij} = f \circ f_i^{-1}(t_j), \quad i = 1,\ldots,m,\ j = 0,\ldots,n,$$

are transformed into

$$Z_{ij} = \mathcal{G}(f) \circ h_i^{-1}(t_j) := g \circ h_i^{-1}(t_j), \quad i = 1,\ldots,m,\ j = 0,\ldots,n, \tag{12}$$

where g is a monotone function. Thus, estimating the warping process of the monotonized model can be used to estimate the real warping functions, and then align the original observations Y_{ij} to their structural mean.

For this, consider a non monotone function $f : [a,b] \to \{1,\ldots,m\}$, and let $a = s_0 < s_1 < \cdots < s_r < s_{r+1} = b$ be the various variational change points, in the sense that $\forall k \in \{0,\ldots,r-1\}$,

$$\forall (t_1,t_2) \in]s_k, s_{k+1}[\,, \forall (t_3,t_4) \in]s_{k+1}, s_{k+2}[\,, \quad t_1 < t_2 \text{ and } t_3 < t_4 \Rightarrow (f(t_1) - f(t_2))(f(t_3) - f(t_4)) < 0. \tag{13}$$

Thus, consider functional warping over the subset

$$\mathcal{F} = \{f : [a,b] \to \{1,\ldots,m\} \subset \mathbb{R} \text{ such that (13) holds}\}.$$

Let $\pi :]a,b[\setminus \{s_1,\ldots,s_r\} \to \{-1,1\}$ be a tool function, indicating whether around a given point t, the function f is increasing or decreasing, and defined by

$$\pi : s_{l(t)} < t < s_{l(t)+1} \mapsto \pi(t) = \begin{cases} -1 & \text{if } s_{l(t)} - s_{l(t)+1} > 0, \\ 1 & \text{if } s_{l(t)} - s_{l(t)+1} < 0, \end{cases}$$

with $l(t) \in \{0, \ldots, r\}$.

Monotonizing operator For all $f \in \mathcal{F}$, define the operator $\mathcal{G}(., f) : t \in]a, b[\setminus \{s_0, \ldots, s_{r+1}\} \rightarrow \mathcal{G}(t, f)$ by

$$\mathcal{G}(t, f) = f(t)\pi(t) - \sum_{k=0}^{r} \pi(t) f(s_k) \mathbf{1}_{]s_k, s_{k+1}[}(t) + f(s_0) + \sum_{k=1}^{r} |f(s_{k-1}) - f(s_k)| \mathbf{1}_{]s_k, b[}(t),$$

and for all $k \in \{0, \ldots, r+1\}$, by

$$\mathcal{G}(s_k, f) = f(a) + \sum_{l=1}^{k} |f(s_{l-1}) - f(s_l)|,$$

with the notation $\sum_{l=1}^{0} |f(s_{l-1}) - f(s_l)| = 0$.

By construction, it is obvious that $t \rightarrow \mathcal{G}(t, f)$ is strictly increasing. Moreover, the following proposition proves that the warping functions remain unchanged.

Proposition 3.1 *Consider $f \in \mathcal{F}$ and the warping functions h_i, $i = 1, \ldots, m$. For every $i = 1, \ldots, m$, set $\mathcal{G}(., f_i) = g_i(.)$. We have $g_i = g \circ h_i^{-1}$.*

The discretization implies however that the Z_{ij} cannot be directly computed, since the functions f_i $(i = 1, \ldots, m)$ are known on the grid t_j $(j = 0, \ldots, n)$, while the values s_k and s_k^i are unknown. Thus, we consider estimates for the Z_{ij} values, which are defined as follows:

$$\begin{aligned} &\tilde{Z}_{i0} = Y_{i0}, \\ &\forall j \in \{1, \ldots, n\}, \quad \tilde{Z}_{ij} = \tilde{Z}_{ij-1} + |Y_{ij} - Y_{ij-1}|. \end{aligned} \tag{14}$$

The following proposition proves the consistency of this estimation procedure.

Proposition 3.2 *For $f \in \mathcal{F}$, $i \in \{1, \ldots, m\}$ and $t \in [a, b]$, define a sequence $j(n)$ such that $\frac{j(n)}{n} \underset{n \rightarrow +\infty}{\longrightarrow} t$. Then,*

$$\tilde{Z}_{ij(n)} - Z_{ij(n)} \xrightarrow[n \rightarrow +\infty]{a.s.} 0.$$

As a conclusion, we can extend our results to the case of non monotone functions, since we transform the problem into a monotone warping problem with the same warping functions. These functions h_i, $i = 1, \ldots, m$, can be estimated with our methodology, by using the new observations $\tilde{Z}_{ij}$, $i = 1, \ldots, m$, $j = 0, \ldots, n$. The resulting estimator can then be written in the form:

$$\widetilde{\phi \circ h_{i_0}^{-1}}_{mn}(t) = \frac{1}{m-1} \sum_{\substack{i=1 \\ i \neq i_0}}^{m} T_{j_i}, \tag{15}$$

with

$$T_{j_i} = \arg \min_{t_j \in \{t_0, \ldots, t_n\}} \left| \tilde{Z}_{ij} - \tilde{Z}_{i_0 j_0} \right|$$

and

$$t_{j_0} = \arg \min_{t_j \in \{t_0, \ldots, t_n\}} |t_j - t|.$$

Other methods are also possible to break the monotonicity assumption, see for instance Liu and Müller (2004). However, our approach preserves the warping functions (see Proposition 3.1), which is crucial when building the structural expectation f_{ES}.

3.3 Dealing with noisy data

While the theoretical asymptotic results are only given in a non noisy framework, we can still handle the case where the data are observed in the standard regression model

$$Y_{ij} = f \circ h_i^{-1}(t_j) + \varepsilon_{ij}, \quad i = 1, \ldots, m, \; j = 0, \ldots, n, \tag{16}$$

with $\varepsilon_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$. To apply our algorithm, we first denoise the data. For this, we separately estimate each function f_i, $i = 1, \ldots, m$, by a kernel estimator. The resulting estimation procedure (to be used in the simulation study) is described as follows:

1. For every $i \in \{1, \ldots, m\}$, f_i is estimated as:

$$\hat{f}_i(t_0) = \frac{1}{m} \sum_{i=1}^{m} Y_{i0} \xrightarrow[m \rightarrow +\infty]{a.s.} f(t_0),$$

$$\hat{f}_i(t_n) = \frac{1}{m} \sum_{i=1}^{m} Y_{in} \xrightarrow[m \rightarrow +\infty]{a.s.} f(t_n).$$

   Given a Gaussian kernel Φ, for every $j \in \{1, \ldots, n-1\}$, $f_i(t_j)$ is estimated by

$$\hat{f}_i(t_j) = \frac{\sum_{k=0}^{n} Y_{ik} \Phi(\frac{t_k - t_j}{\nu_i})}{\sum_{k=0}^{n} \Phi(\frac{t_k - t_j}{\nu_i})}. \tag{17}$$

   The bandwidths ν_i are to be properly chosen.
2. The estimation procedure can be conducted using the denoised observations $\hat{f}_i(t_j)$, leading to new estimates $\hat{f}(t)$ of the structural expectation $f \circ \phi^{-1}$.

We point out that the efficiency of this procedure heavily relies on a proper choice of the bandwidths ν_i, $i = 1, \ldots, m$. Cross-validation techniques do not provide good results since the aim is not to get a good estimation of the function, but only a good separation of the different functions. Hence, over-smoothing the data is not a drawback in this setting. The smoothing parameters $\nu = \nu_i$, for all $i = 1, \ldots, m$ are thus obtained by minimizing the following matching criterion over a grid L:

$$\hat{\nu} = \arg\min_{\nu \in L} \sum_{i=1}^{m} \sum_{j=0}^{n} \left| \hat{f}_i(t_j) - \hat{f}(t_j) \right|. \tag{18}$$

Practical applications of this algorithm are given in Sect. 4.

4 Numerical study

In this section, we estimate the structural expectation, using both the proposed method and the analytic registration approach developed in Ramsay and Silverman (2002). The basic idea of this latter method is to find, for every function f_i, a parameterized monotonic warping function h_i, such that $f_i \circ h_i$ closely matches a target function g in a penalized least square sense:

$$\arg\min_{h_i} \left\{ \sum_{i=1}^{m} \int_0^1 (f_i(h_i(t)) - g(t))^2 dt + \lambda \int_0^1 (h_i''(t))^2 dt \right\}. \tag{19}$$

In the absence of any other information, the target function g is an estimate of the cross-sectional mean of the functions f_i. Note that an alternative approach to analytic registration is provided by the so-called landmarks registration approach (see Kneip and Gasser 1992), but this approach requires the determination of landmarks (such as local extrema), which can be difficult in our simulations. Therefore, this method is not implemented here.

First, some results on simulated data are given, in order to compare the two methods mentioned above. Then, an application of our methodology is given for a real data set.

4.1 Simulations

Two simulation studies are carried out in this section. The first involves a strictly increasing function, while the second involves a non monotone function.

Warped functions Let f and g (see Figs. 2 and 3) be defined by

$$\forall t \in [0, 1], \quad f(t) = \sin(3\pi t) + 3\pi t \quad \text{and}$$

$$g(t) = \frac{\sin(6\pi t)}{6\pi t}.$$

These two functions will be warped by the following random warping process.

Warping processes The stochastic warping functions H_i $(i = 1, \ldots, m)$ are simulated using the iterative process described below.

Let $\epsilon > 0$. First, for all $i = 1, \ldots, m$, let $H_i^{(0)}$ be the identity function. Then, the warping functions $H_i^{(k+1)}$, $i = 1, \ldots, m$, are successively carried out from functions $H_i^{(k)}$, $i = 1, \ldots, m$, by iterating N times the following process:

1. Let U be a uniformly distributed random variable on $[10\epsilon, 1 - 10\epsilon]$.
2. Let V_i, $i = 1, \ldots, m$, be independent and identically uniformly distributed variables on $[U - \epsilon, U + \epsilon]$.
3. For all $i = 1, \ldots, m$, the warping function $H_i^{(k)}$ is warped as follows:

$$H_i^{(k+1)} = W_i \circ H_i^{(k)}$$

where W_i is defined by

$$W_i(t) = \begin{cases} \frac{V_i}{U} t & \text{if } 0 \le t \le U, \\ \frac{1-V_i}{1-U} t + \frac{V_i - U}{1-U} & \text{if } U < t \le 1. \end{cases}$$

We point out that this iterative procedure generates strictly increasing stochastic functions $H_i : [0, 1] \to [0, 1]$, with the desired property that

$$\forall t \in [0, 1], \quad \mathbf{E}(H_i(t)) = \phi(t) = t.$$

Note also that this warping process is centered, in the sense that the structural expectation $f \circ \phi^{-1}$ is equal to f. From a practical point of view, we will never know if ϕ is the identity function or not. Hence, this assumption seems quite natural. Moreover, this simulation procedure enables us to obtain very general classes of warping processes, and is not restricted to parametric deformations.

Our simulated warping functions H_i, $i = 1, \ldots, m$, have been carried out by using the above iterating process with $m = 30$, $N = 3000$ and $\epsilon = 0.005$. These processes are shown on Figs. 2 and 3. In these figures, we can see the very large phase variations. For example, point 0.2 is warped between approximately 0.05 and 0.35 for the first case, and between approximately 0.00 and 0.40 for the second case.

Simulated data Finally, the simulated data are carried out on an equally spaced grid as follows:

$$Y_{ij} = f\left(H_i^{-1}(t_j)\right)$$

and

$$Y_{ij} = g\left(H_i^{-1}(t_j)\right) + \epsilon_{ij}$$

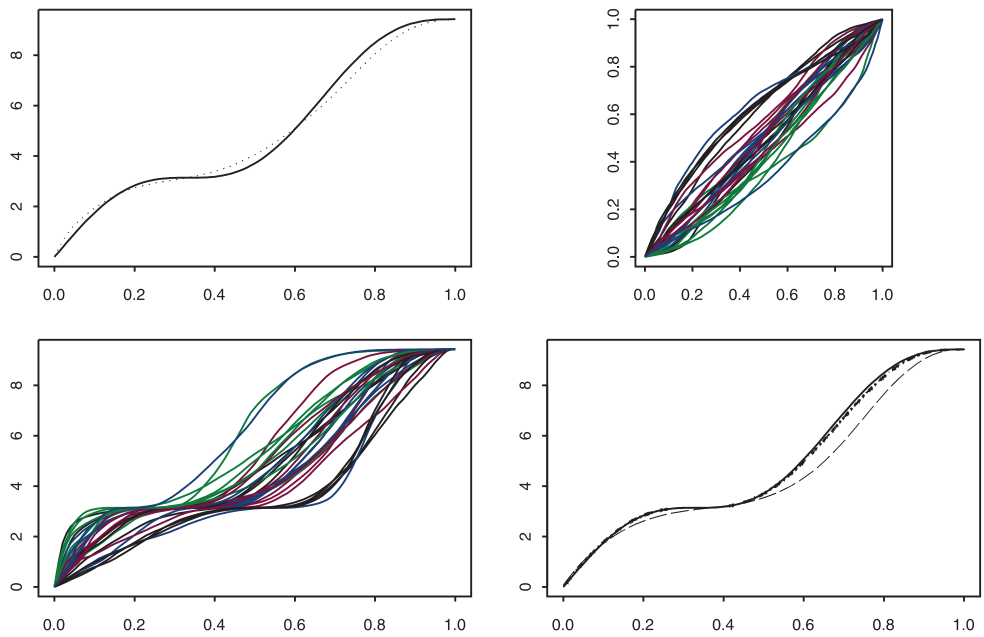

Fig. 2 The function f is shown in the *top left figure* (*solid line*). The simulated warping processes are shown in the *top right figure*. The simulated warped functions are shown in the *bottom left figure*. The classical mean of these functions is drawn in the *top left figure* (*dotted line*). Finally, the estimates of f with the analytic registration procedure (*dashed line*) and our method (*dotted-dashed line*) are shown in the *bottom right figure*

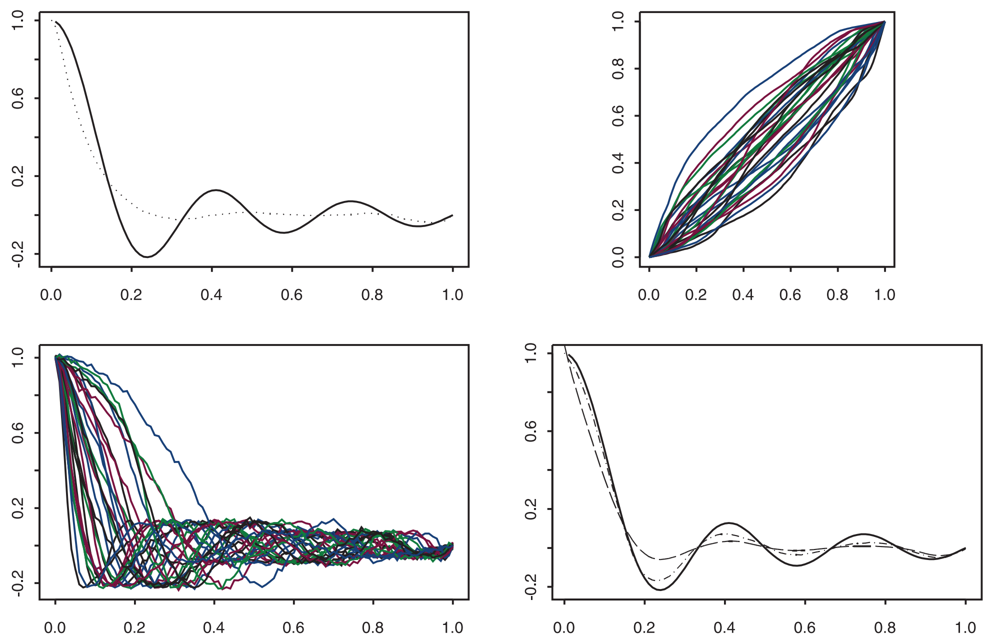

Fig. 3 The function g is shown in the *top left figure* (*solid line*). The simulated warping processes are shown in the *top right figure*. The simulated warped functions are shown in the *bottom left figure*. The classical mean of these functions is drawn in the *top left figure* (*dotted line*). Finally, the estimates of g with the analytic registration procedure (*dashed line*) and our method (*dotted-dashed line*) are shown in the *bottom right figure*

with $t_j = \frac{j}{n}$, $j = 0, 1, \ldots, n$, $n = 100$ points, and ϵ_{ij} i.i.d from a Gaussian law with mean 0 and standard deviation 0.01. The simulated warped functions are respectively shown on Figs. 2 and 3.

Figures 2 and 3 show the functions f and g and the mean functions of the warped curves. We can easily see that the classical mean is not adapted to describe the data. In particular, the mean function of the first simulated data set does not reflect the flat component of the function f (in the range $[0.2, 0.4]$), which yet appears in each individual warped function. In the same way, the mean function of the second simulated data set does not reflect the structure of the individual curves. The classical mean attenuates the curve variations.

Results The estimated structural expectations with both the analytic registration approach and the proposed method are shown on Figs. 2 and 3 (bottom right figures). For noisy data, we use the grid $L = \{0.005, 0.006, \ldots, 0.020\}$ to estimate our bandwidth by minimizing criterion (18). We obtain $\hat{\nu} = 0.011$.

We can easily see that the estimates obtained by using our method are closer to the structural expectations f and g. These results can be explained as follows:

- For the first simulated data set, the analytic registration approach does not directly work on the strictly increasing functions, but on the first derivatives. However, theoretically registering a given function data set is not the same issue as the registration of the first derivatives of these functions.
- The analytic registration approach uses the mean curve to register all the functions. Due to the aforementioned inappropriateness of the mean curve for dealing with large deformations (for instance, in the second data set, where the result is a very flat mean curve), the structural mean approach provides better results.
- For both the simulated data sets, the analytic registration approach works on estimated functions and not directly on the given data, which implies an additional source of error.

4.2 A concrete application: multiple referees and equity

The field of application of the results presented in this paper is large. Here, we consider an example taken from the academic field: how can we ensure equality between the candidates in an exam with several different referees?

Consider an examination with a large number of candidates, such that it is impossible to evaluate the candidates one after another. The students are divided into m groups, and m boards of examiners are charged to grade these m groups: each board grades one group of candidates. The evaluation is performed by assigning a score from 0 to 20.

The m different boards of examiners are supposed to behave the same way, so as to respect the equality among the candidates. Moreover it is assumed that the sampling of the candidates is perfect in the sense that it is done in such a way that each board of examiners evaluates candidates with the same global level. Hence, if all the examiners had the same requirement levels, the distribution of the ranks would be the same for all the boards of examiners. Here, we aim at balancing the effects of the differences between the examiners, gaining equity for the candidates. Our real data set is provided by the French competitive teacher exam *Agrégation de mathématiques*.

The situation can be modeled as follows. For each group i among 13 groups of candidates, let $\mathbf{X}^i = \{X_l^i \in \{1, \ldots, 20\},\ l = 1, \ldots, n_i\}$ denote the scores of the students within this group. Let f_i, $i = 1, \ldots, 13$ be the empirical distribution function of the scores in the i-th group, defined as

$$f_i(t) = \frac{1}{n_i} \sum_{l=1}^{n_i} \mathbf{1}_{X_l^i \leq t}.$$

Figure 4 shows the empirical functions corresponding to the 13 groups.

As a preliminary step in our analysis, we test the null hypothesis H_0 of homogeneity of the distributions of the $\mathbf{X}^i$, $i = 1, \ldots, 13$. Testing for homogeneity of the distributions of any couple $(\mathbf{X}^i, \mathbf{X}^j)$ can be done using the following homogeneity chi-square test (see for instance Bouroche and Saporta 1980). For every score $k = 1, \ldots, 20$, define $n_k^i = \sum_{l=1}^{n_i} \mathbf{1}_{X_l^i = k}$, $n_k^j = \sum_{l=1}^{n_j} \mathbf{1}_{X_l^j = k}$, and $\hat{\mu}_k = (n_k^i + n_k^j)/(n_i + n_j)$. Finally, set $d_k^i = n_i \hat{\mu}_k$, $d_k^j = n_j \hat{\mu}_k$, and

$$D_{n_i}^i = \sum_{k=1}^{20} \frac{(d_k^i - n_k^i)^2}{d_k^i}, \qquad D_{n_j}^j = \sum_{k=1}^{20} \frac{(d_k^j - n_k^j)^2}{d_k^j}.$$

When n_i and n_j both tend to infinity, $D_{n_i}^i + D_{n_j}^j$ converges in distribution to a $\chi^2(19)$ distribution under H_0, and converges almost surely to $+\infty$ if the distributions of $\mathbf{X}^i$ and $\mathbf{X}^j$ are different.

In our example, the total sample size $\sum_{i=1}^{13} n_i$ is equal to $N = 4000$, hence the large sample approximation can be

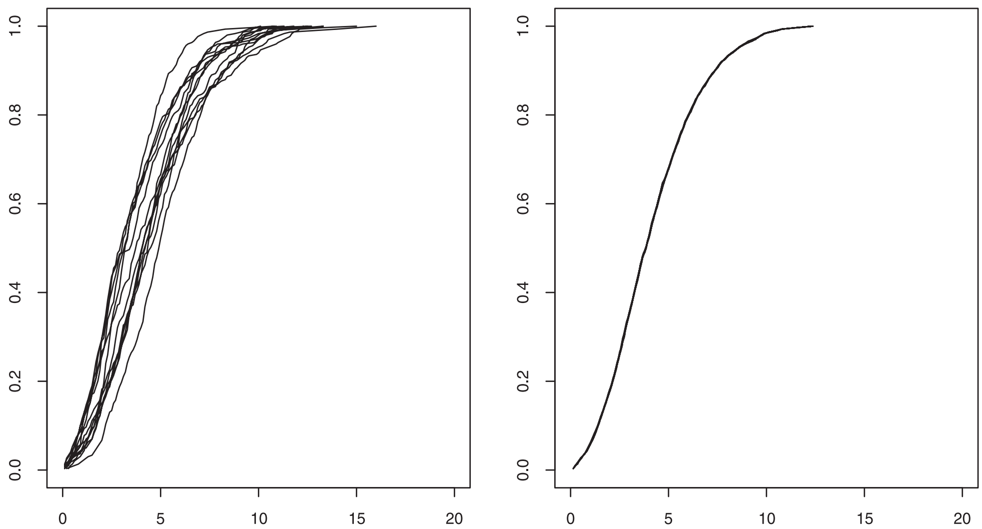

Fig. 4 Empirical distribution functions (*left figure*) and structural expectation estimation (*right figure*)

supposed to hold. In order to test the null hypothesis H_0 of homogeneity over the 13 groups of candidates, we simultaneously perform $13 \times (13-1)/2$ pairwise homogeneity tests, and we control the level (here 5%) of the global test by using the Bonferroni-type correction. This testing procedure leads us to reject H_0.

As a consequence the following procedure is proposed. We aim at finding the average way of ranking, with respect to the ranks that were given within the 13 bunches of candidates. For this, assume that there is such an average empirical distribution function, and that each group-specific empirical distribution function is warped from this reference distribution function by a random process. A good choice is given by the structural expectation: since the functions f_i, $i = 1, \ldots, 13$ are increasing, Lemma 2.3 may apply.

In order to obtain a global common ranking for the N candidates, one can now replace the 13 group-specific rankings by the sole ranking based on the structural expectation (right panel of Fig. 4). To this end, consider a candidate in the i-th group, who obtained a score equal to say, s. We associate to this score the corresponding value $f_i(s)$ of the empirical distribution function in the i-th group. Then, we assign to this candidate the score corresponding to the same probability on the structural expectation graph. As a result, we obtain a new set of scores for the N candidates, which can be interpreted as the scores that would have been obtained, had the candidates been judged by an average board of examiners.

As suggested by a referee, a statistically important question for this dataset is to ask if the 13 group-specific warping processes are identically distributed or not: answering this question may allow one detecting boards of examiners who systematically mark the candidates in a different way from the other boards. This question falls beyond the scope of the present paper but definitely constitutes a stimulating further problem, related to the notion of structural expectation.

In conclusion, structural expectation provides a data-driven pattern, which plays the role of a reference pattern to which all the different curves can be compared. We applied successfully this method to rescale the scores obtained by candidates evaluated by different boards of examiners. This use is not restrictive: the proposed method can be used to provide mean patterns for other types of functional data, in various fields such as econometry, biology or data traffic for instance.

Appendix: Proofs and technical lemmas

In practice, the functions $(f_i)_{i \in \{1,\ldots,m\}}$ are evaluated on a discrete interval of $\mathbb{R}$, as is described in Sect. 1. In order to prove Lemma 2.3 and Theorem 2.6, we first investigate the asymptotic results for the theoretical continuous model, i.e.

$$f_i(t) = f \circ h_i^{-1}(t), \quad i \in \{1, \ldots, m\},\ t \in [a, b], \tag{20}$$

where H is defined in the same way as in model (3). In a second step, we will extend the proofs to the discretized model, and prove the results of Sect. 3.

Thus, consider that all the functions are measured on the entire interval $[a, b]$. After the asymptotic results are proved for this continuous model in Sect. A.1, we use these results to prove Lemma 2.3 and Theorem 2.6 (and subsequently Theorem 2.2) in Sect. A.2.

A.1 Asymptotic results for the continuous model

For the continuous model (20), we provide asymptotic results (analogous to Lemma 2.3 and Theorem 2.6) and their proofs.

A.1.1 Estimator and asymptotic results of the inverse of the structural expectation

Considering the continuous model (20), we define an empirical estimator of the inverse of the structural expectation in the following way. Set

$$\overline{f_{ES}^{-1}} = \frac{1}{m} \sum_{i=1}^{m} f_i^{-1}. \tag{21}$$

The following theorem states the consistency and asymptotic normality of this estimator.

Theorem A.1 *Under assumption* (2), *we have that* $\overline{f_{ES}^{-1}}$ *converges almost surely to* f_{ES}^{-1}

$$\| \overline{f_{ES}^{-1}} - f_{ES}^{-1} \|_\infty \xrightarrow[m\to\infty]{a.s.} 0.$$

Moreover, assume that assumptions (4) *and* (5) *hold. Then, we have that* $\sqrt{m}(\overline{f_{ES}^{-1}} - f_{ES}^{-1})$ *converges weakly to a zero-mean Gaussian process* G:

$$\sqrt{m}\big(\overline{f_{ES}^{-1}} - f_{ES}^{-1}\big) \xrightarrow[m\to\infty]{\mathbf{D}} G,$$

where the covariance function of G is defined, for all $(s, t) \in [f(a), f(b)]^2$, *by*

$$\mathrm{Cov}(G(s), G(t)) = \mathrm{r}\big(f^{-1}(s), f^{-1}(t)\big).$$

Proof Almost sure convergence of the estimator $\overline{f_{ES}^{-1}}$ is directly deduced from Corollary 7.10 (p. 189) in Ledoux and Talagrand (1991). This corollary is an extension of the Strong Law of Large Numbers to Banach spaces.

For all $i \in \{1, \dots, m\}$, the functions $(f_i^{-1})_{i\in\{1,\dots,m\}}$ are obviously strictly increasing, hence $\overline{f_{ES}^{-1}}$ is strictly increasing, and we have

$$\overline{f_{ES}^{-1}} = \frac{1}{m}\sum_{i=1}^{m} f_i^{-1} = \frac{1}{m}\sum_{i=1}^{m}(f \circ h_i^{-1})^{-1}$$
$$= \frac{1}{m}\sum_{i=1}^{m} h_i \circ f^{-1}.$$

For all $i \in \mathbb{N}^*$, let

$$X_i = h_i \circ f^{-1} - f_{ES}^{-1},$$

and, for all $m \in \mathbb{N}^*$, let

$$S_m = \sum_{i=1}^{m} X_i.$$

The $(X_i)_{i\in\{1,\dots,m\}}$ are B-valued random variables, where $B = \mathcal{C}([f(a), f(b)])$ is a separable Banach space. Moreover, the dual space of B is the set of bounded measures on $[f(a), f(b)]$ (Rudin 1987). Hence, our framework corresponds to Chap. 7 of Ledoux and Talagrand (1991), and we can thus apply Corollary 7.10. Indeed, we have

$$\mathbf{E}(\|X_1\|_\infty) < +\infty \quad \text{and} \quad \mathbf{E}(X_1) = 0,$$

then

$$\frac{S_m}{m} \xrightarrow[m\to+\infty]{a.s.} 0,$$

which proves almost sure convergence.

We now turn to the weak convergence. From the multivariate CLT, for any $k \in \mathbb{N}^*$ and fixed $(y_1, \dots, y_k) \in [f(a), f(b)]^k$,

$$\sqrt{m}\left(\begin{pmatrix}\overline{f_{ES}^{-1}}(y_1)\\ \vdots \\ \overline{f_{ES}^{-1}}(y_k)\end{pmatrix} - \begin{pmatrix} f_{ES}^{-1}(y_1)\\ \vdots \\ f_{ES}^{-1}(y_k)\end{pmatrix}\right)$$
$$\xrightarrow[m\to\infty]{\mathcal{D}} \mathcal{N}_k(0, \Gamma),$$

where the covariance matrix $\Gamma = (\Gamma_{ij})_{i,j}$ is given by $\Gamma_{ij} = \mathrm{cov}(F^{-1}(y_i), F^{-1}(y_j)) = \mathrm{cov}(H(f^{-1}(y_i)), H(f^{-1}(y_j)))$, $i, j = 1, \dots, k$. It remains to show that $\{\sqrt{m}(\overline{f_{ES}^{-1}} - f_{ES}^{-1})\}$ is tight. We verify the moment condition stated by Vaart and Wellner (1996, Example 2.2.12).

$$\mathbf{E}\Big[\Big|\sqrt{m}(\overline{f_{ES}^{-1}}(s) - f_{ES}^{-1}(s))$$
$$- \sqrt{m}(\overline{f_{ES}^{-1}}(t) - f_{ES}^{-1}(t))\Big|^2\Big]$$
$$= \mathbf{E}\left[m \left| \frac{1}{m}\sum_{i=1}^{m} f_i^{-1}(s) - \mathbf{E}F^{-1}(s) \right.\right.$$
$$\left.\left. - \left(\frac{1}{m}\sum_{i=1}^{m} f_i^{-1}(t) - \mathbf{E}F^{-1}(t)\right)\right|^2\right]$$
$$= \mathbf{E}\Big[\big|F^{-1}(s) - \mathbf{E}F^{-1}(s) - \big(F^{-1}(t) - \mathbf{E}F^{-1}(t)\big)\big|^2\Big],$$

where the last equality follows from the fact that the h_i's are i.i.d. Then, from (4) and (5), we get that

$$\mathbf{E}\Big[\big|\sqrt{m}(\overline{f_{ES}^{-1}}(s) - f_{ES}^{-1}(s)) - \sqrt{m}(\overline{f_{ES}^{-1}}(t) - f_{ES}^{-1}(t))\big|^2\Big]$$
$$\le C_1 C_2 |s - t|^2,$$

which completes the proof. □

A.1.2 Estimator and asymptotic results of an individual warping function

For the continuous model, we define an empirical estimator of the individual warping function $\phi \circ h_{i_0}^{-1}$ $(i_0 \in \{1, \dots, m\})$ as follows. Conditional on $F_{i_0} = f_{i_0}$, for all $t \in [a, b]$, let

$$\overline{\phi \circ h_{i_0}^{-1}}(t) = \frac{1}{m-1}\sum_{\substack{i=1\\ i\neq i_0}}^{m} f_i^{-1} \circ F_{i_0}(t). \tag{22}$$

The following theorem states the consistency and asymptotic normality of this estimator.

Theorem A.2 *Under assumption* (2), *we have that* $\overline{\phi \circ h_{i_0}^{-1}}$ *converges almost surely to* $\phi \circ h_{i_0}^{-1}$:

$$\|\overline{\phi \circ h_{i_0}^{-1}} - \phi \circ h_{i_0}^{-1}\|_\infty \xrightarrow[m\to\infty]{a.s.} 0.$$

Let $n = m^{\frac{1}{2}+\alpha}$ (*with* $\alpha > 0$) *and assume that* (4) *and* (6) *hold. Then, we have that* $\sqrt{m}(\overline{\phi \circ h_{i_0}^{-1}} - \phi \circ h_{i_0}^{-1})$ *converges weakly to a zero-mean Gaussian process* Z,

$$\sqrt{m}(\overline{\phi \circ h_{i_0}^{-1}} - \phi \circ h_{i_0}^{-1}) \xrightarrow[m\to\infty]{\mathbf{D}} Z,$$

with covariance function defined for all $(s, t) \in [a, b]^2$ *by*

$$\mathbf{Cov}(Z(s), Z(t)) = \mathrm{r}\big(h_{i_0}^{-1}(s), h_{i_0}^{-1}(t)\big).$$

Proof Let $i_0 \in \{1, \dots, m\}$. Given $F_{i_0} = f_{i_0}$,

$$\overline{\phi \circ h_{i_0}^{-1}} = \frac{1}{m-1}\sum_{\substack{i=1\\ i\neq i_0}}^{m} f_i^{-1} \circ f_{i_0}.$$

Noting also that $\phi \circ h_{i_0}^{-1} = f_{ES}^{-1} \circ f_{i_0}$, the consistency of $\overline{\phi \circ h_{i_0}^{-1}}$ follows by the same arguments as in the proof of Theorem A.1.

We now turn to the weak convergence. From the multivariate CLT, for any $k \in \mathbb{N}^*$ and fixed $(t_1, \ldots, t_k) \in [a, b]^k$,

$$\sqrt{m}\left(\begin{pmatrix} \overline{\phi \circ h_{i_0}^{-1}(t_1)} \\ \vdots \\ \overline{\phi \circ h_{i_0}^{-1}(t_k)} \end{pmatrix} - \begin{pmatrix} \phi \circ h_{i_0}^{-1}(t_1) \\ \vdots \\ \phi \circ h_{i_0}^{-1}(t_k) \end{pmatrix}\right) \xrightarrow{\mathcal{D}} \mathcal{N}_k(0, \Gamma_0),$$

where the covariance matrix $\Gamma_0 = (\Gamma_{0,ij})_{i,j}$ is given by $\Gamma_{0,ij} = \text{cov}(H(h_{i_0}^{-1}(t_i)), H(h_{i_0}^{-1}(t_j))) = r(h_{i_0}^{-1}(t_i), h_{i_0}^{-1}(t_j))$, $i, j = 1, \ldots, k$. It remains to show that $\{\sqrt{m}(\overline{\phi \circ h_{i_0}^{-1}} - \phi \circ h_{i_0}^{-1})\}$ is tight. Again, we verify the moment condition stated by Vaart and Wellner (1996, Example 2.2.12).

$$\begin{aligned}
&\mathbf{E}\Big[\big|\sqrt{m}(\overline{\phi \circ h_{i_0}^{-1}}(s) - \phi \circ h_{i_0}^{-1}(s)) \\
&\quad - \sqrt{m}(\overline{\phi \circ h_{i_0}^{-1}}(t) - \phi \circ h_{i_0}^{-1}(t))\big|^2\Big] \\
&= \frac{m}{(m-1)^2}\mathbf{E}\Bigg[\Bigg|\sum_{\substack{i=1 \\ i \neq i_0}}^{m} \big(f_i^{-1}(f_{i_0}(s)) - \mathbf{E} f_i^{-1}(f_{i_0}(s)) \\
&\quad - \big(f_i^{-1}(f_{i_0}(t)) - \mathbf{E} f_i^{-1}(f_{i_0}(t))\big)\big)\Bigg|^2\Bigg] \\
&= \frac{m}{m-1}\mathbf{E}\Big[\big|H(h_{i_0}^{-1}(s)) - \mathbf{E}H(h_{i_0}^{-1}(s)) \\
&\quad - \big(H(h_{i_0}^{-1}(t)) - \mathbf{E}H(h_{i_0}^{-1}(t))\big)\big|^2\Big].
\end{aligned}$$

Now, from the assumptions (4) and (6), we get that

$$\begin{aligned}
&\mathbf{E}\Big[\big|\sqrt{m}(\overline{\phi \circ h_{i_0}^{-1}}(s) - \phi \circ h_{i_0}^{-1}(s)) \\
&\quad - \sqrt{m}(\overline{\phi \circ h_{i_0}^{-1}}(t) - \phi \circ h_{i_0}^{-1}(t))\big|^2\Big] \\
&\leq 2C_1C_3|s-t|^2,
\end{aligned}$$

which completes the proof. □

A.2 Proofs of asymptotic results

We now use Theorem A.1 (given for the continuous model) to prove Lemma 2.3.

Proof of Lemma 2.3 Let $y \in [f(a), f(b)]$. The n observation times are equidistant and for each $i = 1, \ldots, m$, f_i is almost surely increasing, hence $f_i^{-1}(y) - \frac{1}{n} \leq T_i(y) \leq f_i^{-1}(y) + \frac{1}{n}$. This implies that almost surely,

$$\overline{f_{ES}^{-1}}(y) - \frac{1}{n} \leq \widehat{\overline{f_{ES}^{-1}}}(y) \leq \overline{f_{ES}^{-1}}(y) + \frac{1}{n}. \tag{23}$$

Since

$$\left\|\overline{f_{ES}^{-1}} + \frac{1}{n} - f_{ES}^{-1}\right\|_\infty \leq \|\overline{f_{ES}^{-1}} - f_{ES}^{-1}\|_\infty + \frac{1}{n},$$

we get that $\|\overline{f_{ES}^{-1}} + \frac{1}{n} - f_{ES}^{-1}\|_\infty \xrightarrow[m,n\to\infty]{a.s.} 0$, by Theorem A.1. A similar argument holds for LHS (23) and finally, $\|\widehat{\overline{f_{ES}^{-1}}} - f_{ES}^{-1}\|_\infty \xrightarrow[m,n\to\infty]{a.s.} 0$.

Let $n = m^{\frac{1}{2}+\alpha}$ $(\alpha > 0)$. From (23), we get that almost surely,

$$\|\sqrt{m}(\widehat{\overline{f_{ES}^{-1}}} - \overline{f_{ES}^{-1}})\|_\infty \leq \frac{1}{m^\alpha}.$$

Since $\|\sqrt{m}(\widehat{\overline{f_{ES}^{-1}}} - \overline{f_{ES}^{-1}})\|_\infty = \|\sqrt{m}(\widehat{\overline{f_{ES}^{-1}}} - f_{ES}^{-1}) - \sqrt{m}(\overline{f_{ES}^{-1}} - f_{ES}^{-1})\|_\infty$, we get that $\|\sqrt{m}(\widehat{\overline{f_{ES}^{-1}}} - f_{ES}^{-1}) - \sqrt{m}(\overline{f_{ES}^{-1}} - f_{ES}^{-1})\|_\infty$ converges almost surely to 0 as m tends to infinity. Combining Theorems A.1 and 4.1 in Billingsley (1968), it follows that $\sqrt{m}(\widehat{\overline{f_{ES}^{-1}}} - f_{ES}^{-1}) \xrightarrow[m\to\infty]{\mathbf{D}} G$.

We now turn to the proof of Lemma 2.4. □

Proof of Lemma 2.4 Consider first the continuous model (20), and define

$$\overline{\gamma \circ f^{-1}} = \frac{1}{m}\sum_{i=1}^{m}(f_i^{-1})^2.$$

Using similar arguments as in the proof of Theorem A.1, we get that $\|\overline{\gamma \circ f^{-1}} - \gamma \circ f^{-1}\|_\infty \xrightarrow[m\to\infty]{a.s.} 0$.

Now, since $|T_i(y) - f_i^{-1}(y)| \leq \frac{1}{n}$, we obtain by straightforward calculations that almost surely,

$$\begin{aligned}
&-\frac{1}{n} \cdot \frac{1}{m}\sum_{i=1}^{m}|f_i^{-1}(y)| \\
&\leq \frac{1}{m}\sum_{i=1}^{m} T_i . f_i^{-1}(y) - \overline{\gamma \circ f^{-1}}(y) \\
&\leq \frac{1}{n} \cdot \frac{1}{m}\sum_{i=1}^{m}|f_i^{-1}(y)|,
\end{aligned}$$

which implies that $\frac{1}{m}\sum_{i=1}^{m} T_i . f_i^{-1}(y) - \overline{\gamma \circ f^{-1}}(y) \xrightarrow[m,n\to\infty]{a.s.} 0$, from which we deduce that $\frac{1}{m}\sum_{i=1}^{m} T_i . f_i^{-1}(y) \xrightarrow[m,n\to\infty]{a.s.} \gamma \circ f^{-1}(y)$.

From $|T_i(y) - f_i^{-1}(y)| \leq \frac{1}{n}$, we also get that

$$\begin{aligned}
&0 \leq \frac{1}{m}\sum_{i=1}^{m} T_i^2 - \frac{2}{m}\sum_{i=1}^{m} T_i . f_i^{-1}(y) + \frac{1}{m}\sum_{i=1}^{m}(f_i^{-1}(y))^2 \\
&\leq \frac{1}{n^2},
\end{aligned}$$

that is, $0 \leq \widehat{\gamma \circ f^{-1}}(y) - \frac{2}{m}\sum_{i=1}^{m} T_i \cdot f_i^{-1}(y) + \overline{\gamma \circ f^{-1}}(y) \leq \frac{1}{n^2}$, from which we deduce that $\widehat{\gamma \circ f^{-1}}(y) \xrightarrow[m,n\to\infty]{a.s.} \gamma \circ f^{-1}(y)$. Combining this with Lemma 2.3 completes the proof. □

Proof of Theorem 2.2 For F an arbitrary function, let F^{-1} denote its generalized inverse, defined by $F^{-1}(t) = \inf\{y : F(y) \geq t\}$. By Lemma 2.3, for all $y \in [f(a), f(b)]$,

$$\widehat{f_{ES}^{-1}}(y) \xrightarrow[m,n\to\infty]{a.s.} f_{ES}^{-1}(y).$$

By Lemma 21.2 in van der Vaart (1998),

$$(\widehat{f_{ES}^{-1}})^{-1}(t) \xrightarrow[m,n\to\infty]{a.s.} (f_{ES}^{-1})^{-1}(t)$$

at every t where $(f_{ES}^{-1})^{-1}$ is continuous. Since f_{ES}^{-1} is continuous and strictly increasing, $(f_{ES}^{-1})^{-1}$ is a proper inverse and is equal to $f \circ \phi^{-1}$, hence for all $t \in [a, b]$,

$$(\widehat{f_{ES}^{-1}})^{-1}(t) \xrightarrow[m,n\to\infty]{a.s.} f \circ \phi^{-1}(t).$$

Now, for all $t \in [a, b]$, $(\widehat{f_{ES}^{-1}})^{-1}(t)$ can be rewritten as

$$(\widehat{f_{ES}^{-1}})^{-1}(t) = v_0 \mathbf{1}_{\{a\}}(t) + \sum_{k=0}^{K(m,n)-1} v_{k+1} \mathbf{1}_{(u_k, u_{k+1}]}(t),$$

and by construction, letting $t \in (u_{k(m,n)}, u_{k(m,n)+1}]$ ($k(m,n) \in \mathcal{K}$), we have $v_{k(m,n)} \leq \widehat{f_{ES}}(t) \leq v_{k(m,n)+1}$. Combining this and the above equality yields that for all $t \in (u_{k(m,n)}, u_{k(m,n)+1}]$, $|\widehat{f_{ES}}(t) - (\widehat{f_{ES}^{-1}})^{-1}(t)| \leq v_{k(m,n)+1} - v_{k(m,n)}$. Since f_{ES} is continuous, $v_{k(m,n)+1} - v_{k(m,n)} \xrightarrow[m,n\to\infty]{} 0$, and $|\widehat{f_{ES}}(t) - (\widehat{f_{ES}^{-1}})^{-1}(t)| \xrightarrow[m,n\to\infty]{a.s.} 0$. It follows that almost surely, $\widehat{f_{ES}}(t)$ converges to $f_{ES}(t)$. The uniform convergence finally follows from Dini's theorem. □

We now use Theorem A.2 (given for the continuous model) to prove Theorem 2.6. A preliminary definition and a lemma are needed.

Let define the modulus of continuity of H as

$$K_H(\delta) = \sup_{\substack{(s,t)\in[a,b]^2 \\ |s-t|\leq\delta}} |H_1 \circ H_2^{-1}(s) - H_1 \circ H_2^{-1}(t)| \quad (\delta \geq 0),$$

where H_1 and H_2 are two independent copies of H. Since for all $\delta \geq 0$, $0 \leq K_H(\delta) \leq b - a$ almost surely, we can define $L_H^1(H_2, \delta)$ and $L_H^2(H_2, \delta)$ as $L_H^1(H_2, \delta) = \mathbf{E}[K_H(\delta)|H_2]$ and $L_H^2(H_2, \delta) = \mathbf{E}[K_H^2(\delta)|H_2]$ ($\delta \geq 0$), and we have $L_H^1(H_2, 0) = L_H^2(H_2, 0) = 0$. From the continuity of H, it holds that $L_H^1(H_2, \cdot)$ and $L_H^2(H_2, \cdot)$ are right-continuous at 0. Given $H_2 = h$, this implies that for $\epsilon > 0$, there exists $\eta_\epsilon > 0$ such that $L_H^1(h, \eta_\epsilon) < \epsilon$. We shall use this result in proving the following lemma:

Lemma A.3 *Let $t \in [a, b]$. Under assumption (2), the following holds*:

$$\sup_{t\in[a,b]} \left| \overline{\phi \circ h_{i_0}^{-1}}\left(t + \frac{1}{n}\right) - \overline{\phi \circ h_{i_0}^{-1}}(t) \right| \xrightarrow[m,n\to\infty]{a.s.} 0.$$

Proof Letting $t \in [a, b]$, one easily shows that

$$\left| \overline{\phi \circ h_{i_0}^{-1}}\left(t + \frac{1}{n}\right) - \overline{\phi \circ h_{i_0}^{-1}}(t) \right| \leq \frac{1}{m-1} \sum_{\substack{i=1 \\ i\neq i_0}}^{m} \left| h_i \circ h_{i_0}^{-1}\left(t + \frac{1}{n}\right) - h_i \circ h_{i_0}^{-1}(t) \right|.$$

This in turn implies that

$$\sup_{t\in[a,b]} \left| \overline{\phi \circ h_{i_0}^{-1}}\left(t + \frac{1}{n}\right) - \overline{\phi \circ h_{i_0}^{-1}}(t) \right| \leq \frac{1}{m-1} \sum_{\substack{i=1 \\ i\neq i_0}}^{m} \sup_{\substack{(s,t)\in[a,b]^2 \\ |s-t|\leq\frac{1}{n}}} \left| h_i \circ h_{i_0}^{-1}(s) - h_i \circ h_{i_0}^{-1}(t) \right| = \frac{1}{m-1} \sum_{\substack{i=1 \\ i\neq i_0}}^{m} \tilde{K}_H^i\left(\frac{1}{n}\right),$$

where the $\tilde{K}_H^i(\frac{1}{n})$ ($i \in \{1, \ldots, m\}\backslash i_0$) are independent random variables distributed as $K_H(\frac{1}{n})|H_2 = h_{i_0}$. Since $K_H(\cdot)$ is increasing, if $n \in \mathbb{N}$ is sufficiently large so that $\frac{1}{n} < \eta_\epsilon$, we get that

$$\sup_{t\in[a,b]} \left| \overline{\phi \circ h_{i_0}^{-1}}\left(t + \frac{1}{n}\right) - \overline{\phi \circ h_{i_0}^{-1}}(t) \right| \leq \frac{1}{m-1} \sum_{\substack{i=1 \\ i\neq i_0}}^{m} \tilde{K}_H^i(\eta_\epsilon).$$

By the law of large numbers, we get that almost surely,

$$0 \leq \limsup_{m,n} \sup_{t\in[a,b]} \left| \overline{\phi \circ h_{i_0}^{-1}}\left(t + \frac{1}{n}\right) - \overline{\phi \circ h_{i_0}^{-1}}(t) \right| \leq L_H^1(h_{i_0}, \eta_\epsilon) < \epsilon.$$

This holds for any $\epsilon > 0$, hence

$$\sup_{t\in[a,b]} \left| \overline{\phi \circ h_{i_0}^{-1}}\left(t + \frac{1}{n}\right) - \overline{\phi \circ h_{i_0}^{-1}}(t) \right| \xrightarrow[m,n\to\infty]{a.s.} 0.$$

□

Proof of Theorem 2.6 Let $t \in [a, b]$. For each $i \in \{1, \dots, m\}$, f_i is almost surely increasing, hence almost surely $f_i^{-1} \circ f_{i_0}(t_{j_0}) - \frac{1}{n} \le T_i \le f_i^{-1} \circ f_{i_0}(t_{j_0}) + \frac{1}{n}$. Recall that $t_{j_0} = \arg\min_{j \in \{1,\dots,n\}} |t_j - t|$, hence $t - \frac{1}{n} \le t_{j_0} \le t + \frac{1}{n}$. Thus, combining the above two inequalities, we get that almost surely $f_i^{-1} \circ f_{i_0}(t - \frac{1}{n}) - \frac{1}{n} \le T_i \le f_i^{-1} \circ f_{i_0}(t + \frac{1}{n}) + \frac{1}{n}$, from which we deduce:

$$\overline{\phi \circ h_{i_0}^{-1}}\left(t - \frac{1}{n}\right) - \frac{1}{n} \le \widehat{\phi \circ h_{i_0}^{-1}}(t) \le \overline{\phi \circ h_{i_0}^{-1}}\left(t + \frac{1}{n}\right) + \frac{1}{n}. \tag{24}$$

We shall focus on the upper bound in this inequality, since the same kind of argument holds for the lower bound. Straightforward calculation yields

$$\sup_{t \in [a,b]} \left| \overline{\phi \circ h_{i_0}^{-1}}\left(t + \frac{1}{n}\right) + \frac{1}{n} - \phi \circ h_{i_0}^{-1}(t) \right| \le \sup_{t \in [a,b]} \left| \overline{\phi \circ h_{i_0}^{-1}}\left(t + \frac{1}{n}\right) - \overline{\phi \circ h_{i_0}^{-1}}(t) \right| + \frac{1}{n} + \| \overline{\phi \circ h_{i_0}^{-1}} - \phi \circ h_{i_0}^{-1} \|_\infty,$$

where the RHS of this inequality tends to 0 as m and n tend to infinity, by Lemma A.3 and Theorem A.2 in Billingsley (1968). It follows that $\|\widehat{\phi \circ h_{i_0}^{-1}} - \phi \circ h_{i_0}^{-1}\|_\infty \xrightarrow[m,n\to\infty]{a.s.} 0$.

We now turn to the weak convergence. Assume that $n = m^{\frac{1}{2}+\alpha}$ $(\alpha > 0)$. From (24), it holds

$$\begin{aligned}
&\| \sqrt{m}(\widehat{\phi \circ h_{i_0}^{-1}} - \overline{\phi \circ h_{i_0}^{-1}}) \|_\infty \\
&\le \sup_{t \in [a,b]} \left| \sqrt{m}\left(\overline{\phi \circ h_{i_0}^{-1}}\left(t + \frac{1}{n}\right) - \overline{\phi \circ h_{i_0}^{-1}}(t) \right) \right| \\
&\quad + \sup_{t \in [a,b]} \left| \sqrt{m}\left(\overline{\phi \circ h_{i_0}^{-1}}\left(t - \frac{1}{n}\right) - \overline{\phi \circ h_{i_0}^{-1}}(t) \right) \right| + \frac{2}{m^\alpha} \\
&\le 2Z_m + \frac{2}{m^\alpha},
\end{aligned}$$

where $Z_m = \sqrt{m} \frac{1}{m-1} \sum_{\substack{i=1 \\ i \ne i_0}}^m \tilde{K}_H^i(\frac{1}{n})$. Since

$$\text{var}(Z_m) = \frac{m}{m-1}\left(L_H^2\left(h_{i_0}, \frac{1}{n}\right) - \left\{ L_H^1\left(h_{i_0}, \frac{1}{n}\right) \right\}^2 \right) \xrightarrow[m\to\infty]{} 0$$

(recall that $n = m^{\frac{1}{2}+\alpha}$ and that $L_H^1(h_{i_0}, \cdot)$ and $L_H^2(h_{i_0}, \cdot)$ are right-continuous at 0), we get that

$$\| \sqrt{m}(\widehat{\phi \circ h_{i_0}^{-1}} - \overline{\phi \circ h_{i_0}^{-1}}) \|_\infty \xrightarrow[m\to\infty]{P} 0,$$

hence by Theorem A.2 and Theorem 4.1 in Billingsley (1968), $\sqrt{m}(\widehat{\phi \circ h_{i_0}^{-1}} - \phi \circ h_{i_0}^{-1}) \xrightarrow[m\to\infty]{\mathbf{D}} Z$. □

Proof of Lemma 2.7 Consider first the continuous model (20) and define, conditional on $F_{i_0} = f_{i_0}$,

$$\overline{\gamma \circ h_{i_0}^{-1}} = \frac{1}{m-1} \sum_{\substack{i=1 \\ i \ne i_0}}^m (f_i^{-1} \circ f_{i_0})^2.$$

Using similar arguments as in the proof of Theorem A.2, we get that

$$\| \overline{\gamma \circ h_{i_0}^{-1}} - \gamma \circ h_{i_0}^{-1} \|_\infty \xrightarrow[m\to\infty]{a.s.} 0.$$

Now, from $f_i^{-1} \circ f_{i_0}(t - \frac{1}{n}) - \frac{1}{n} \le T_i \le f_i^{-1} \circ f_{i_0}(t + \frac{1}{n}) + \frac{1}{n}$, we get the following inequality:

$$\begin{aligned}
&|T_i - f_i^{-1} \circ f_{i_0}(t)| \\
&\le \left| f_i^{-1} \circ f_{i_0}\left(t - \frac{1}{n}\right) - f_i^{-1} \circ f_{i_0}(t) \right| \\
&\quad + \left| f_i^{-1} \circ f_{i_0}\left(t + \frac{1}{n}\right) - f_i^{-1} \circ f_{i_0}(t) \right| + \frac{2}{n}.
\end{aligned}$$

Acting as in the proof of Lemma 2.4, it is fairly straightforward to show that for $t \in [a, b]$, $\frac{1}{m-1} \sum_{\substack{i=1 \\ i \ne i_0}}^m T_i . f_i^{-1} \circ f_{i_0}(t) \xrightarrow[m,n\to\infty]{a.s.} \gamma \circ h_{i_0}^{-1}(t)$.

From the above inequality, we also obtain that

$$|T_i - f_i^{-1} \circ f_{i_0}(t)| \le 2\tilde{K}_H^i\left(\frac{1}{n}\right) + \frac{2}{n}.$$

Summing over $i \in \{1, \dots, m\} \backslash i_0$, and then using convergence of $\frac{1}{m-1} \sum_{i \ne i_0} T_i . f_i^{-1} \circ f_{i_0}(t)$, of $\overline{\gamma \circ h_{i_0}^{-1}}(t)$, together with right-continuity of $L_H^1(h_{i_0}, \cdot)$ and $L_H^2(h_{i_0}, \cdot)$ at 0, yield that $\widehat{\gamma \circ h_{i_0}^{-1}}(t) \xrightarrow[m,n\to\infty]{a.s.} \gamma \circ h_{i_0}^{-1}(t)$. This and the convergence of $\widehat{\phi \circ h_{i_0}^{-1}}(t)$ complete the proof of Lemma 2.7. □

Proof of Proposition 3.1 For $i \in \{1, \dots, m\}$, let us prove that $f_i \in \mathcal{F}$. Using (13), consider the change points $(s_k)_{k=1,\dots,r}$ and set $s_k^i = h_i(s_k)$. For $k \in \{0, \dots, r-1\}$, consider

$$(t_1, t_2) \in]s_k^i, s_{k+1}^i[, \qquad (t_3, t_4) \in]s_{k+1}^i, s_{k+2}^i[\,/\, t_1 < t_2 \quad \text{and} \quad t_3 < t_4,$$

then,

$$\begin{aligned}
&\big(f_i(t_1) - f_i(t_2)\big)\big(f_i(t_3) - f_i(t_4)\big) \\
&= \big(f \circ h_i^{-1}(t_1) - f \circ h_i^{-1}(t_2)\big) \\
&\quad \times \big(f \circ h_i^{-1}(t_3) - f \circ h_i^{-1}(t_4)\big).
\end{aligned}$$

Since h_i^{-1} is strictly increasing, we get $s_k < h_i^{-1}(t_1) < h_i^{-1}(t_2) < s_{k+1}$ and $s_{k+1} < h_i^{-1}(t_3) < h_i^{-1}(t_3) < s_{k+2}$. Hence

$$\big(f_i(t_1) - f_i(t_2)\big)(f_i(t_3) - f_i(t_4)) < 0,$$

which yields that $f_i \in \mathcal{F}$. Thus, define $g_i = \mathcal{G}(f_i)$. For all $t \in]a, b[\setminus \{s_1^i, \dots, s_r^i\}$, we get

$$g_i(t) = f_i(t)\Pi(t, f_i) - \sum_{k=0}^{r} \Pi(t, f_i) f_i(s_k^i) \mathbf{1}_{]s_k^i, s_{k+1}^i[}(t)$$
$$+ f_i(s_0) + \sum_{k=1}^{r} \big| f_i(s_{k-1}^i) - f_i(s_k^i) \big| \mathbf{1}_{]s_k^i, b[}(t).$$

But

$$f_i(t) = f \circ h_i^{-1}(t),$$
$$\Pi(t, f_i) = \Pi(t, f \circ h_i^{-1}) = \Pi(h_i^{-1}(t), f),$$
$$\mathbf{1}_{]s_k^i, s_{k+1}^i[}(t) = \mathbf{1}_{]h_i(s_k), h_i(s_{k+1})[}(t) = \mathbf{1}_{]s_k, s_{k+1}[}\big(h_i^{-1}(t)\big),$$

which implies that

$$g_i = g \circ h_i^{-1}. \qquad \square$$

Proof of Proposition 3.2 Set $t \in]a, b[\setminus \{s_1^i, \dots, s_r^i\}$, and $l \in \{0, 1, \dots, r\}$ such that $t \in]s_l^i, s_{l+1}^i[$. Consider $j(n)$, $\frac{j(n)}{n} \underset{n \to +\infty}{\longrightarrow} t$, then $\exists n_0 \in \mathbb{N} / \forall n \in \mathbb{N}$, $n \geq n_0 \Rightarrow \frac{j(n)}{n} \in]s_l^i, s_{l+1}^i[$. For all $n \geq n_0$, we have

$$\tilde{Z}_{ij(n)} = \tilde{Z}_{ij(n)-1} + |Y_{ij(n)} - Y_{ij(n)-1}|$$
$$= Y_{i0} + \sum_{k=1}^{j(n)} |Y_{ik} - Y_{ik-1}|.$$

Moreover,

$$Z_{ij(n)} = g \circ h_i^{-1}(t_{j(n)})$$
$$= f_i(t_{j(n)})\Pi(t_{j(n)}, f_i) + \Pi(t_{j(n)}, f_i) f_i(s_l^i)$$
$$+ f_i(s_0) + \sum_{k=1}^{l} \big| f_i\big(s_{k-1}^i\big) - f_i\big(s_k^i\big) \big|,$$

which yields that

$$Z_{ij(n)} = \Pi\big(t_{j(n)}, f_i\big)\big(Y_{ij(n)} - f_i\big(s_l^i\big)\big)$$
$$+ Y_{i0} + \sum_{k=1}^{l} \big| f_i\big(s_{k-1}^i\big) - f_i\big(s_k^i\big) \big|$$
$$= \big| Y_{ij(n)} - f_i\big(s_l^i\big) \big| + Y_{i0}$$
$$+ \sum_{k=1}^{l} \big| f_i\big(s_{k-1}^i\big) - f_i\big(s_k^i\big) \big|$$
$$= A + Y_{i0} + B.$$

Write $\forall k = 1, \dots, l+1$, $s_{k-1}^i \leq t_{j_k - p_k} < \dots < t_{j_k} \leq s_k^i$, we get that

$$\big| f_i\big(s_{k-1}^i\big) - f_i\big(s_k^i\big) \big|$$
$$= \Big| f_i\big(s_{k-1}^i\big) - f_i\big(t_{j_k - p_k}\big)$$
$$+ \sum_{q=1}^{p_k} (f_i(t_{j_k - q}) - f_i(t_{j_k - q + 1})) + f_i(t_{j_k}) - f_i\big(s_k^i\big) \Big|$$
$$= \big| f_i\big(s_{k-1}^i\big) - f_i\big(t_{j_k - p_k}\big) \big|$$
$$+ \sum_{q=1}^{p_k} |f_i(t_{j_k - q}) - f_i(t_{j_k - q + 1})| + \big| f_i\big(t_{j_k}\big) - f_i\big(s_k^i\big) \big|$$
$$= \big| f_i\big(s_{k-1}^i\big) - Y_{ij_k - p_k} \big|$$
$$+ \sum_{q=1}^{p_k} |Y_{ij_k - q + 1} - Y_{ij_k - q}| + \big| Y_{ij_k} - f_i\big(s_k^i\big) \big|.$$

Hence

$$B = \sum_{k=1}^{l} \big| f_i\big(s_{k-1}^i\big) - f_i\big(s_k^i\big) \big|$$
$$= \sum_{k=1}^{l} \big| f_i\big(s_{k-1}^i\big) - Y_{ij_k - p_k} \big| + \sum_{k=1}^{l} \big| Y_{ij_k} - f_i\big(s_k^i\big) \big|$$
$$+ \sum_{k=1}^{l} \sum_{q=1}^{p_k} |Y_{ij_k - q + 1} - Y_{ij_k - q}|$$
$$= \sum_{k=1}^{l} \big| f_i\big(s_{k-1}^i\big) - Y_{ij_k - p_k} \big| + \sum_{k=1}^{l} \big| Y_{ij_k} - f_i\big(s_k^i\big) \big|$$
$$+ \sum_{k=1}^{j_l} |Y_{ik} - Y_{ik-1}|$$
$$- \sum_{k=1}^{l-1} |Y_{ij_{k+1} - p_{k+1}} - Y_{ij_{k+1} - p_{k+1} - 1}| - |Y_{i1} - Y_{i0}|.$$

With the same ideas, we can write

$$A = \big| Y_{ij(n)} - f_i\big(s_l^i\big) \big|$$
$$= \sum_{q = j_{l+1} - j(n) + 1}^{p_{l+1} + 1} |Y_{ij_{l+1} - q + 1} - Y_{ij_{l+1} - q}|$$
$$+ \big| Y_{ij_{l+1} - p_{l+1}} - f_i\big(s_l^i\big) \big|$$
$$- |Y_{ij_{l+1} - p_{l+1}} - Y_{ij_{l+1} - p_{l+1} - 1}|.$$

As a result,

$$
\begin{aligned}
& Z_{ij(n)} - \tilde{Z}_{ij(n)} \\
& \quad = \sum_{k=1}^{l+1} \left| f_i\left(s_{k-1}^i\right) - Y_{ij_k - p_k} \right| + \sum_{k=1}^{l} \left| Y_{ij_k} - f_i\left(s_k^i\right) \right| \\
& \quad\quad - \sum_{k=1}^{l} |Y_{ij_{k+1}-p_{k+1}} - Y_{ij_{k+1}-p_{k+1}-1}| - |Y_{i1} - Y_{i0}|.
\end{aligned}
$$

By continuity of f, f_i is also continuous, hence

$$
Z_{ij(n)} - \tilde{Z}_{ij(n)} \underset{n\to+\infty}{\overset{a.s.}{\longrightarrow}} 0.
$$

For $t \in \{s_0^i, \ldots, s_{r+1}^i\}$, we get similar results, leading to the conclusion. □

References

Bigot, J.: A scale-space approach with wavelets to singularity estimation. ESAIM Probab. Stat. **9**, 143–164 (2005) (electronic)

Billingsley, P.: Convergence of Probability Measures. Wiley, New York (1968)

Bouroche, J.-M., Saporta, G.: L'analyse des données. Que Sais-Je? Presses Universitaires de France, Paris (1980)

Gamboa, F., Loubes, J.-M., Maza, E.: Semi-parametric estimation of shifts. Electron. J. Stat. **1**, 616–640 (2007)

Gervini, D., Gasser, T.: Nonparametric maximum likelihood estimation of the structural mean of a sample of curves. Biometrika **92**(4), 801–820 (2005)

Kneip, A., Gasser, T.: Statistical tools to analyze data representing a sample of curves. Ann. Stat. **20**(3), 1266–1305 (1992)

Kneip, A., Li, X., MacGibbon, K.B., Ramsay, J.O.: Curve registration by local regression. Can. J. Stat. **28**(1), 19–29 (2000)

Ledoux, M., Talagrand, M.: Probability in Banach Spaces. Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 23. Springer, Berlin (1991). Isoperimetry and processes

Liu, X., Müller, H.-G.: Functional convex averaging and synchronization for time-warped random curves. J. Am. Stat. Assoc. **99**(467), 687–699 (2004)

Loubes, J.-M., Maza, É., Lavielle, M., Rodríguez, L.: Road trafficking description and short term travel time forecasting, with a classification method. Can. J. Stat. **34**(3), 475–491 (2006)

Ramsay, J.O., Li, X.: Curve registration. J. R. Stat. Soc. Ser. B Stat. Methodol. **60**(2), 351–363 (1998)

Ramsay, J.O., Silverman, B.W.: Applied Functional Data Analysis. Springer Series in Statistics. Springer, New York (2002). Methods and case studies

Rønn, B.B.: Nonparametric maximum likelihood estimation for shifted curves. J. R. Stat. Soc. Ser. B Stat. Methodol. **63**(2), 243–259 (2001)

Rudin, W.: Real and Complex Analysis, 3rd edn. McGraw-Hill, New York (1987)

Sakoe, H., Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans. ASSP-26 **1**, 43–49 (1978)

van der Vaart, A.W.: Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge (1998)

van der Vaart, A.W., Wellner, J.A.: Weak Convergence and Empirical Processes. Springer Series in Statistics. Springer, New York (1996)

Wang, K., Gasser, T.: Synchronizing sample curves nonparametrically. Ann. Stat. **27**(2), 439–460 (1999)