Biometrika (2012), **xx**, x, *pp*. 1–8

Supplementary file: Dispersion operators and resistant second-order functional data analysis

BY DAVID KRAUS AND VICTOR M. PANARETOS
Section de Mathématiques, Ecole Polytechnique Fédérale de Lausanne, EPFL Station 8, 1015 Lausanne, Switzerland
david.kraus@epfl.ch victor.panaretos@epfl.ch

SUMMARY

This supplementary file contains proofs of Proposition 1, Corollary 1, Proposition 2, Theorem 1 and a technical lemma needed in the proof of Theorem 1. Equations in this supplement are numbered (S1), (S2), . . . ; equation numbers such as (1), (2), . . . or (A1), (A2), . . . refer to the main body of the paper.

PROOF OF PROPOSITION 1

It suffices to prove that the finitely-valued objective functional $M(\mathsf{P};\mathscr{R},\mu)$ given in equation (2) in the paper admits a unique minimizer on the space of Hilbert–Schmidt operators acting on $\mathcal{H}$. By the triangle inequality, monotonicity and convexity of ρ we have that

$$\begin{aligned} E_{\mathsf{P}}(\rho[\|\mathscr{P}(X;\mu) - \{\lambda\mathscr{R} + (1-\lambda)\mathscr{R}'\}\|] - \rho\{\|\mathscr{P}(X;\mu)\|\}) \\ \leq E_{\mathsf{P}}[\rho\{\lambda\|\mathscr{P}(X;\mu) - \mathscr{R}\| + (1-\lambda)\|\mathscr{P}(X;\mu) - \mathscr{R}'\|\} - \rho\{\|\mathscr{P}(X;\mu)\|\}] \\ \leq \lambda\, E_{\mathsf{P}}[\rho\{\|\mathscr{P}(X;\mu) - \mathscr{R}\|\} - \rho\{\|\mathscr{P}(X;\mu)\|\}] \\ + (1-\lambda)\, E_{\mathsf{P}}[\rho\{\|\mathscr{P}(X;\mu) - \mathscr{R}'\|\} - \rho\{\|\mathscr{P}(X;\mu)\|\}] \end{aligned}$$

for any $\lambda \in [0,1]$ and arbitrary Hilbert–Schmidt operators $\mathscr{R},\mathscr{R}'$. Notice that since ρ is strictly increasing, the first inequality is strict unless $\mathscr{P}(X;\mu) - \mathscr{R}$ and $\mathscr{P}(X;\mu) - \mathscr{R}'$ are collinear almost surely. Equivalently, the inequality is strict whenever the distribution of $\mathscr{P}(X;\mu)$ is not concentrated on the line $\{t\mathscr{R} + (1-t)\mathscr{R}' : t \in \mathbb{R}\}$.

We now investigate what this condition means geometrically in the space $\mathcal{H}$. First, notice that as the rank of $\mathscr{P}(X;\mu)$ is 1, the rank of $t\mathscr{R} + (1-t)\mathscr{R}'$ has to be 1 also. Now we distinguish two cases.

First, if $\mathscr{R},\mathscr{R}'$ are collinear, then the line is of the form $\{\alpha\mathscr{R} : \alpha \in \mathbb{R}\}$, which by the condition on the rank is $\{\alpha u \otimes u : \alpha \in \mathbb{R}\}$ for some $u \in \mathcal{H}$. Since $\mathscr{P}(X;\mu)$ is positive semidefinite, we in fact have $\{\alpha u \otimes u : \alpha \geq 0\}$. Thus, the operator $\mathscr{P}(X;\mu)$ lying on this line is equivalent to X lying on the line $\{\mu + \beta u : \beta \in \mathbb{R}\}$.

Second, if $\mathscr{R},\mathscr{R}'$ are not collinear, then operators of the form $t\mathscr{R} + (1-t)\mathscr{R}'$ have rank 1 for at most two values of t. To see this, notice that the rank condition implies that for all $i < j$,

$$\det\left\{ t \begin{pmatrix} R_{ii} & R_{ij} \\ R_{ji} & R_{jj} \end{pmatrix} + (1-t) \begin{pmatrix} R'_{ii} & R'_{ij} \\ R'_{ji} & R'_{jj} \end{pmatrix} \right\} = 0,$$

where $R_{ij} = \langle e_i, \mathscr{R} e_j\rangle$, $R'_{ij} = \langle e_i, \mathscr{R}' e_j\rangle$. This system of quadratic equations has at most two solutions. Thus, the set $\{t\mathscr{R} + (1-t)\mathscr{R}' : t \in \mathbb{R}\}$ reduces at most to the set $\{\alpha_1 u_1 \otimes u_1, \alpha_2 u_2 \otimes$

$u_2\}$ for some nonnegative α_1, α_2 and some $u_1, u_2 \in \mathcal{H}$. Hence, the operator $\mathscr{P}(X;\mu)$ belonging to this set is equivalent to X belonging to the set of at most four points $\{\mu \pm \beta_1 u_1, \mu \pm \beta_2 u_2\}$.

Therefore, if the distribution P is not concentrated on a line or on four points, the objective function to be minimized is strictly convex. It follows that the minimum of the functional exists and is unique.

PROOF OF COROLLARY 1

The empirical version of the functional defining the dispersion operator is the expectation with respect to the empirical distribution $\hat{\mathsf{P}}$. Under our assumptions on P, the empirical distribution $\hat{\mathsf{P}}$ is almost surely not concentrated on a line or on four points. Therefore, strict convexity, and thus existence and uniqueness, follows with probability 1 by applying Proposition 1 to the empirical distribution $\hat{\mathsf{P}}$. Consistency then follows from strict convexity and the consistency of $\hat{\mu}$, using standard arguments.

PROOF OF PROPOSITION 2

Consider $\mathscr{R}$ of the form $\sum_{k=1}^{\infty} \delta_k \varphi_k \otimes \varphi_k$ for some sequence $\delta_1, \delta_2, \ldots$ We will prove that such an operator solves the estimating equation (5) showing that $\mathscr{R}$ and $\mathscr{C}$ have the same set of eigenfunctions, and that the sequence $\delta_1, \delta_2, \ldots$ satisfies the condition (6).

We investigate the coordinates of the left-hand side of (5), with the aim of showing that the values

$$\left\langle \varphi_j, E_{\mathsf{P}}\left[\frac{\rho'\{\|\mathscr{R}-\mathscr{P}(X;\mu)\|\}}{\|\mathscr{R}-\mathscr{P}(X;\mu)\|}\{\mathscr{R}-\mathscr{P}(X;\mu)\}\right]\varphi_k\right\rangle \tag{S1}$$

are zero for all j, k. By the orthonormality of $\varphi_1, \varphi_2, \ldots$, we have that

$$\begin{aligned}\|\mathscr{R}-\mathscr{P}(X;\mu)\|^2 &= \left\|\sum_{k=1}^{\infty}\delta_k\varphi_k\otimes\varphi_k - \sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\lambda_j^{1/2}\lambda_k^{1/2}\beta_j\beta_k\varphi_j\otimes\varphi_k\right\|^2 \\ &= \sum_k(\delta_k-\lambda_k\beta_k^2)^2 + \sum_{k\neq j}\lambda_j\lambda_k\beta_j^2\beta_k^2.\end{aligned}$$

First, we compute the off-diagonal coordinates with $j \neq k$. The first summand in (S1) is zero because $\langle\varphi_j, \mathscr{R}\varphi_k\rangle = 0$. To show that the second summand in (S1) is zero, we use the fact that, by assumption, the sequence $\{s_i\beta_i\}_{i=1}^{\infty}$ with $s_i = (-1)^{1\{i=j\}}$ has the same joint distribution as $\{\beta_i\}_{i=1}^{\infty}$. Compute

$$\begin{aligned}A_{jk} &= \left\langle \varphi_j, E_{\mathsf{P}}\left[\frac{\rho'\{\|\mathscr{R}-\mathscr{P}(X;\mu)\|\}}{\|\mathscr{R}-\mathscr{P}(X;\mu)\|}\mathscr{P}(X;\mu)\right]\varphi_k\right\rangle \\ &= E\left(\frac{\rho'[\{\sum_i(\delta_i-\lambda_i\beta_i^2)^2+\sum_{i\neq l}\lambda_i\lambda_l\beta_i^2\beta_l^2\}^{1/2}]}{\{\sum_i(\delta_i-\lambda_i\beta_i^2)^2+\sum_{i\neq l}\lambda_i\lambda_l\beta_i^2\beta_l^2\}^{1/2}}\lambda_j^{1/2}\lambda_k^{1/2}\beta_j\beta_k\right) \\ &= E\left\{\frac{\rho'([\sum_i\{\delta_i-\lambda_i(s_i\beta_i)^2\}^2+\sum_{i\neq l}\lambda_i\lambda_l(s_i\beta_i)^2(s_l\beta_l)^2]^{1/2})}{[\sum_i\{\delta_i-\lambda_i(s_i\beta_i)^2\}^2+\sum_{i\neq l}\lambda_i\lambda_l(s_i\beta_i)^2(s_l\beta_l)^2]^{1/2}}\lambda_j^{1/2}\lambda_k^{1/2}s_j\beta_js_k\beta_k\right\} \\ &= -A_{jk}.\end{aligned}$$

Thus, $A_{jk} = 0$. Therefore, the operator $\mathscr{R}$ is diagonalized by the same functions $\varphi_1, \varphi_2, \ldots$ as $\mathscr{C}$. By computing the diagonal coordinates with $j = k$ in (5) we obtain (6).

A TECHNICAL LEMMA

LEMMA 1. *Under the assumptions of Theorem 1,*

(a) the linear operator $\mathscr{D}(\mathsf{P};\mu)$ defined in equation (A1) is a bijection of $\mathcal{H}$ onto itself, it is bounded and has bounded inverse,

(b) the linear operator $\mathfrak{D}(\mathsf{P};\mathscr{R},\mu)$ defined in equation (A2) is a bijection of $\mathrm{HS}(\mathcal{H},\mathcal{H})$ onto itself, it is bounded and has bounded inverse.

Proof. We prove part (a); the proof of part (b) is similar. The proof uses and extends the steps of the proof of Lemma 1 (iii) of Gervini (2008) modified for the present context of general ρ and generalized to the case of infinitely many components in the Karhunen–Loève expansion.

Recall that

$$\mathscr{D}(\mathsf{P};\mu) = E_{\mathsf{P}}\left[\frac{\rho'(\|X-\mu\|)}{\|X-\mu\|}\mathscr{I} + \left\{\frac{\rho''(\|X-\mu\|)}{\|X-\mu\|^2} - \frac{\rho'(\|X-\mu\|)}{\|X-\mu\|^3}\right\}\mathscr{P}(X;\mu)\right];$$

see the appendix of the main body of the paper. To show that $\mathscr{D}(\mathsf{P};\mu)$ is a bijection, we need to find for any $h \in \mathcal{H}$ a unique element $f \in \mathcal{H}$ such that $\mathscr{D}(\mathsf{P};\mu)f = h$. The set of orthonormal eigenfunctions $\{\varphi_k\}_{k=1}^{\infty}$ of $\mathscr{C}$ can be extended to an orthonormal basis of $\mathcal{H}$ by possibly adding some functions $\{\psi_k\}_{k=1}^{q}$ with q finite or infinite or zero. It is then enough to verify the relation $\mathscr{D}(\mathsf{P};\mu)f = h$ in terms of the Fourier coefficients of both sides with respect to the basis $\{\varphi_k\}_{k=1}^{\infty} \cup \{\psi_k\}_{k=1}^{q}$, i.e., to show that $\langle \mathscr{D}(\mathsf{P};\mu)f, \varphi_k\rangle = \langle h, \varphi_k\rangle$ for all $k = 1, 2, \ldots$ and $\langle \mathscr{D}(\mathsf{P};\mu)f, \psi_k\rangle = \langle h, \psi_k\rangle$ for all $k = 1, \ldots, q$. As $\langle \mathscr{D}(\mathsf{P};\mu)f, \varphi_k\rangle = \langle f, \mathscr{D}(\mathsf{P};\mu)\varphi_k\rangle$ and $\langle \mathscr{D}(\mathsf{P};\mu)f, \psi_k\rangle = \langle f, \mathscr{D}(\mathsf{P};\mu)\psi_k\rangle$, we first investigate $\mathscr{D}(\mathsf{P};\mu)\varphi_k$ and $\mathscr{D}(\mathsf{P};\mu)\psi_k$.

We begin by exploring the structure of the operator $\mathscr{D}(\mathsf{P};\mu)$. We can rewrite

$$E_{\mathsf{P}}\left\{\frac{\rho'(\|X-\mu\|)}{\|X-\mu\|^3}\mathscr{P}(X;\mu)\right\} = E_{\mathsf{P}}(\tilde{\varepsilon}\otimes\tilde{\varepsilon}),$$

where

$$\tilde{\varepsilon} = \frac{\rho'(\|X-\mu\|)^{1/2}}{\|X-\mu\|^{3/2}}(X-\mu) = \sum_{k=1}^{\infty}\lambda_k^{1/2}\frac{\rho'(\|X-\mu\|)^{1/2}}{\|X-\mu\|^{3/2}}\beta_k\varphi_k = \sum_{k=1}^{\infty}\tilde{\lambda}_k^{1/2}\tilde{\beta}_k\varphi_k \qquad \text{(S2)}$$

with

$$\tilde{\lambda}_k = \lambda_k\, E_{\mathsf{P}}\left\{\frac{\rho'(\|X-\mu\|)}{\|X-\mu\|^3}\beta_k^2\right\},$$
$$\tilde{\beta}_k = \frac{\rho'(\|X-\mu\|)^{1/2}}{\|X-\mu\|^{3/2}}\beta_k\Big/\left[E_{\mathsf{P}}\left\{\frac{\rho'(\|X-\mu\|)}{\|X-\mu\|^3}\beta_k^2\right\}\right]^{1/2}.$$

Thus, we need to find the covariance operator of $\tilde{\varepsilon}$. The series expansion (S2) of $\tilde{\varepsilon}$ is a Karhunen–Loève expansion because the coefficients $\tilde{\beta}_k$ have zero mean and unit variance and are uncorrelated (which follows from the fact that the distribution of $\{\beta_k\}$ is invariant under the change of the sign of any component). Therefore, since $E_{\mathsf{P}}(\|\tilde{\varepsilon}\|^2) < \infty$, which follows immediately from the assumption that

$$E_{\mathsf{P}}\left\{\frac{\rho'(\|X-\mu\|)}{\|X-\mu\|}\right\} < \infty,$$

the operator of interest, as the covariance operator of $\tilde{\varepsilon}$, takes the form

$$E_{\mathsf{P}}\left\{\frac{\rho'(\|X-\mu\|)}{\|X-\mu\|^3}\mathscr{P}(X;\mu)\right\}=\sum_{k=1}^{\infty}\tilde{\lambda}_k\varphi_k\otimes\varphi_k=\sum_{k=1}^{\infty}E_{\mathsf{P}}\left\{\frac{\rho'(\|X-\mu\|)}{\|X-\mu\|^3}\lambda_k\beta_k^2\right\}\varphi_k\otimes\varphi_k.$$

Using analogous arguments for

$$\dot{\varepsilon}=\frac{\rho''(\|X-\mu\|)^{1/2}}{\|X-\mu\|}(X-\mu),$$

we can show that

$$E_{\mathsf{P}}\left\{\frac{\rho''(\|X-\mu\|)}{\|X-\mu\|^2}\mathscr{P}(X;\mu)\right\}=\sum_{k=1}^{\infty}E_{\mathsf{P}}\left\{\frac{\rho''(\|X-\mu\|)}{\|X-\mu\|^2}\lambda_k\beta_k^2\right\}\varphi_k\otimes\varphi_k.$$

Hence, we finally obtain $\mathscr{D}(\mathsf{P};\mu)$ in the form

$$\begin{aligned}\mathscr{D}(\mathsf{P};\mu)&=E_{\mathsf{P}}\left\{\frac{\rho'(\|X-\mu\|)}{\|X-\mu\|}\right\}\mathscr{I}\\&\quad+\sum_{k=1}^{\infty}E_{\mathsf{P}}\left[\left\{\frac{\rho''(\|X-\mu\|)}{\|X-\mu\|^2}-\frac{\rho'(\|X-\mu\|)}{\|X-\mu\|^3}\right\}\lambda_k\beta_k^2\right]\varphi_k\otimes\varphi_k.\end{aligned}$$

Therefore, for $k=1,2,\dots$ we have

$$\mathscr{D}(\mathsf{P};\mu)\varphi_k=E_{\mathsf{P}}\left\{\frac{\rho'(\|X-\mu\|)}{\|X-\mu\|}\right\}\varphi_k+E_{\mathsf{P}}\left[\left\{\frac{\rho''(\|X-\mu\|)}{\|X-\mu\|^2}-\frac{\rho'(\|X-\mu\|)}{\|X-\mu\|^3}\right\}\lambda_k\beta_k^2\right]\varphi_k$$

and, for $k=1,\dots,q$, we have

$$\mathscr{D}(\mathsf{P};\mu)\psi_k=E_{\mathsf{P}}\left\{\frac{\rho'(\|X-\mu\|)}{\|X-\mu\|}\right\}\psi_k.$$

Thus, we obtain

$$\begin{aligned}\langle\mathscr{D}(\mathsf{P};\mu)f,\varphi_k\rangle&=\nu_k\langle f,\varphi_k\rangle\quad(k=1,2,\dots),\\\langle\mathscr{D}(\mathsf{P};\mu)f,\psi_k\rangle&=\eta\langle f,\psi_k\rangle\quad(k=1,\dots,q),\end{aligned}$$

where

$$\nu_k=E_{\mathsf{P}}\left\{\frac{\rho'(\|X-\mu\|)}{\|X-\mu\|}\right\}+\lambda_k\,E_{\mathsf{P}}\left[\left\{\frac{\rho''(\|X-\mu\|)}{\|X-\mu\|^2}-\frac{\rho'(\|X-\mu\|)}{\|X-\mu\|^3}\right\}\beta_k^2\right]\quad(k=1,2,\dots)$$

and

$$\eta=E_{\mathsf{P}}\left\{\frac{\rho'(\|X-\mu\|)}{\|X-\mu\|}\right\}.$$

So f, the candidate for $\mathscr{D}(\mathsf{P};\mu)^{-1}h$, should have Fourier coefficients $\langle f,\varphi_k\rangle$, $\langle f,\psi_k\rangle$ satisfying the system of equations

$$\nu_k\langle f,\varphi_k\rangle=\langle h,\varphi_k\rangle\quad(k=1,2,\dots),\quad\eta\langle f,\psi_k\rangle=\langle h,\psi_k\rangle\quad(k=1,\dots,q).$$

To be able to write $\langle f,\varphi_k\rangle=\langle h,\varphi_k\rangle/\nu_k$, we need to show that ν_k $(k=1,2,\dots)$ and η are nonzero and finite. Then, f will be uniquely determined by the formula

$$f=\sum_{k=1}^{\infty}\frac{\langle h,\varphi_k\rangle}{\nu_k}\varphi_k+\sum_{k=1}^{q}\frac{\langle h,\psi_k\rangle}{\eta}\psi_k$$

provided that f is a well-defined element of $\mathcal{H}$, that is,

$$\|f\|^2 = \sum_{k=1}^{\infty} \frac{\langle h, \varphi_k \rangle^2}{\nu_k^2} + \sum_{k=1}^{q} \frac{\langle h, \psi_k \rangle^2}{\eta^2} < \infty. \tag{S3}$$

We assumed that $\eta < \infty$ and we immediately see that $\eta > 0$ because ρ is strictly increasing. We now deal with ν_k ($k = 1, 2, \ldots$). We will show that there exist $0 < a \leq b < \infty$ such that $\nu_k \in [a, b]$ for all $k = 1, 2, \ldots$

First we establish the lower bound a. Using the Karhunen–Loève expansion (S2) we can rewrite

$$E_{\mathsf{P}}\left\{ \frac{\rho'(\|X-\mu\|)}{\|X-\mu\|} \right\} = E_{\mathsf{P}}(\|\tilde{\varepsilon}\|^2) = \sum_{k=1}^{\infty} \tilde{\lambda}_k = \sum_{k=1}^{\infty} \lambda_k \, E_{\mathsf{P}}\left\{ \frac{\rho'(\|X-\mu\|)}{\|X-\mu\|^3} \beta_k^2 \right\}. \tag{S4}$$

Each term in the series on the right hand side of (S4) is obviously positive and by finiteness of the left hand side it is finite, and thus the differences

$$E_{\mathsf{P}}\left\{ \frac{\rho'(\|X-\mu\|)}{\|X-\mu\|} \right\} - \lambda_k \, E_{\mathsf{P}}\left\{ \frac{\rho'(\|X-\mu\|)}{\|X-\mu\|^3} \beta_k^2 \right\}, \tag{S5}$$

which appear in the expression for ν_k, are positive and bounded away from zero by a constant a. The remaining term

$$\lambda_k \, E_{\mathsf{P}}\left\{ \frac{\rho''(\|X-\mu\|)}{\|X-\mu\|^2} \beta_k^2 \right\} \tag{S6}$$

appearing in ν_k is nonnegative as $\rho'' \geq 0$ because ρ is convex. It follows that $\nu_k \geq a$ for all $k = 1, 2, \ldots$

Now we find the upper bound b. By applying the same idea as in (S4) to $\dot{\varepsilon}$, we obtain

$$E_{\mathsf{P}}\{\rho''(\|X-\mu\|)\} = \sum_{k=1}^{\infty} \lambda_k \, E_{\mathsf{P}}\left\{ \frac{\rho''(\|X-\mu\|)}{\|X-\mu\|^2} \beta_k^2 \right\}. \tag{S7}$$

In view of (S7), the terms (S6) are smaller than or equal to $E_{\mathsf{P}}\{\rho''(\|X-\mu\|)\}$. The differences (S5) are smaller than

$$E_{\mathsf{P}}\left\{ \frac{\rho'(\|X-\mu\|)}{\|X-\mu\|} \right\}.$$

Therefore, we have that $\nu_k \leq b$ for all $k = 1, 2, \ldots$ with

$$b = E_{\mathsf{P}}\left\{ \frac{\rho'(\|X-\mu\|)}{\|X-\mu\|} \right\} + E_{\mathsf{P}}\{\rho''(\|X-\mu\|)\}.$$

Finally, it remains to show (S3), which is now straightforward because

$$\|f\|^2 = \sum_{k=1}^{\infty} \frac{\langle h, \varphi_k \rangle^2}{\nu_k^2} + \sum_{k=1}^{q} \frac{\langle h, \psi_k \rangle^2}{\eta^2} \leq \frac{\sum_{k=1}^{\infty} \langle h, \varphi_k \rangle^2 + \sum_{k=1}^{q} \langle h, \psi_k \rangle^2}{\min(a, \eta)} = \frac{\|h\|^2}{\min(a, \eta)} < \infty.$$

This shows that f is a well defined element of $\mathcal{H}$ and thus the linear operator $\mathscr{D}(\mathsf{P}; \mu)$ is a bijection of $\mathcal{H}$ onto itself. It also shows that the inverse $\mathscr{D}(\mathsf{P}; \mu)^{-1}$ is a bounded operator. Hence also the operator $\mathscr{D}(\mathsf{P}; \mu)$ is bounded by the bounded inverse theorem or by direct verification. $\square$

Remark: As ν_k are bounded away from zero and bounded from above, the operator $\mathscr{D}(\mathsf{P};\mu)$ is only a small perturbation of a multiple of the identity. This gives an intuitive explanation why it inherits its bijectivity and boundedness.

PROOF OF THEOREM 1

It is enough to prove the weak convergence of $n^{1/2}\mathscr{B}(\hat{\mathsf{P}}_1,\hat{\mathsf{P}}_2,a_n;\hat{\mathscr{R}},\hat{\mu}_1,\hat{\mu}_2)$. The weak convergence of the vector with components S_l will then follow directly from Slutsky's theorem. The continuous mapping theorem and Slutsky's theorem will then establish the weak convergence of the statistic T. Applying a Taylor expansion (Nelson, 1969, Theorem 6, p. 12) of $\mathscr{B}(\hat{\mathsf{P}}_1,\hat{\mathsf{P}}_2,a_n;\hat{\mathscr{R}},\hat{\mu}_1,\hat{\mu}_2)$ around the true values of the parameters yields

$$\begin{aligned} n^{1/2}\mathscr{B}(\hat{\mathsf{P}}_1,\hat{\mathsf{P}}_2,a_n;\hat{\mathscr{R}},\hat{\mu}_1,\hat{\mu}_2) = {}& n^{1/2}\mathscr{B}(\hat{\mathsf{P}}_1,\hat{\mathsf{P}}_2,a_n;\mathscr{R},\mu_1,\mu_2)\\ &+\mathfrak{D}_1(\hat{\mathsf{P}}_1,\hat{\mathsf{P}}_2,a_n;\mathscr{R}^\diamond,\mu_1^\diamond,\mu_2^\diamond)n^{1/2}(\hat{\mathscr{R}}-\mathscr{R})\\ &+a_n^{1/2}\mathbb{D}(\hat{\mathsf{P}}_1;\mathscr{R}^\diamond,\mu_1^\diamond)n_1^{1/2}(\hat{\mu}_1-\mu_1)\\ &-(1-a_n)^{1/2}\mathbb{D}(\hat{\mathsf{P}}_2;\mathscr{R}^\diamond,\mu_2^\diamond)n_2^{1/2}(\hat{\mu}_2-\mu_2), \end{aligned} \tag{S8}$$

where

$$\begin{aligned} \mathfrak{D}_1(\mathsf{P}_1,\mathsf{P}_2,a;\mathscr{R},\mu_1,\mu_2) &= \frac{\partial}{\partial\mathscr{R}}\mathscr{B}(\mathsf{P}_1,\mathsf{P}_2,a;\mathscr{R},\mu_1,\mu_2)\\ &= a\mathfrak{D}(\mathsf{P}_1;\mathscr{R},\mu_1)-(1-a)\mathfrak{D}(\mathsf{P}_2;\mathscr{R},\mu_2) \end{aligned}$$

and

$$\mathfrak{D}(\mathsf{P};\mathscr{R},\mu)=\frac{\partial}{\partial\mathscr{R}}\mathscr{G}(\mathsf{P};\mathscr{R},\mu),\quad \mathbb{D}(\mathsf{P};\mathscr{R},\mu)=\frac{\partial}{\partial\mu}\mathscr{G}(\mathsf{P};\mathscr{R},\mu).$$

See the Appendix in the main body of the paper for explicit formulae.

We now turn to develop certain asymptotic representations for $\hat{\mu}_1$, $\hat{\mu}_2$ and $\hat{\mathscr{R}}$. Using the Taylor expansion, law of large numbers and consistency of $\hat{\mu}_1$ we get

$$\begin{aligned} 0=n_1^{1/2}G(\hat{\mathsf{P}}_1;\hat{\mu}_1) &= n_1^{1/2}G(\hat{\mathsf{P}}_1;\mu_1)+\mathscr{D}(\hat{\mathsf{P}}_1;\mu_1^\dagger)n_1^{1/2}(\hat{\mu}_1-\mu_1)\\ &= n_1^{1/2}G(\hat{\mathsf{P}}_1;\mu_1)+\mathscr{D}(\mathsf{P}_1;\mu_1)n_1^{1/2}(\hat{\mu}_1-\mu_1)+o_P(1), \end{aligned}$$

where the term $o_P(1)$ is due to the fact that we replace $\mathscr{D}(\hat{\mathsf{P}}_1;\mu_1)$ by its limit $\mathscr{D}(\mathsf{P}_1;\mu_1)$. From this and an analogous expansion for μ_2 we obtain

$$\begin{aligned} n_1^{1/2}(\hat{\mu}_1-\mu_1) &= -\mathscr{D}(\mathsf{P}_1;\mu_1)^{-1}n_1^{1/2}G(\hat{\mathsf{P}}_1;\mu_1)+o_P(1),\\ n_2^{1/2}(\hat{\mu}_2-\mu_2) &= -\mathscr{D}(\mathsf{P}_2;\mu_2)^{-1}n_2^{1/2}G(\hat{\mathsf{P}}_2;\mu_2)+o_P(1). \end{aligned} \tag{S9}$$

The existence of the bounded inverse operators in the above equations, as well as of other inverse operators appearing later in the proof, is shown in Lemma 1. The Taylor expansion of the estimating score for $\mathscr{R}$ around the true values is

$$\begin{aligned} \mathscr{O}=n^{1/2}\mathscr{G}(\hat{\mathsf{P}}_1,\hat{\mathsf{P}}_2,a_n;\hat{\mathscr{R}},\hat{\mu}_1,\hat{\mu}_2) = {}& n^{1/2}\mathscr{G}(\hat{\mathsf{P}}_1,\hat{\mathsf{P}}_2,a_n;\mathscr{R},\mu_1,\mu_2)\\ &+\mathfrak{D}_0(\hat{\mathsf{P}}_1,\hat{\mathsf{P}}_2,a_n;\mathscr{R}^\ddagger,\mu_1^\ddagger,\mu_2^\ddagger)n^{1/2}(\hat{\mathscr{R}}-\mathscr{R})\\ &+a_n^{1/2}\mathbb{D}(\hat{\mathsf{P}}_1;\mathscr{R}^\ddagger,\mu_1^\ddagger)n_1^{1/2}(\hat{\mu}_1-\mu_1)\\ &+(1-a_n)^{1/2}\mathbb{D}(\hat{\mathsf{P}}_2;\mathscr{R}^\ddagger,\mu_2^\ddagger)n_2^{1/2}(\hat{\mu}_2-\mu_2), \end{aligned}$$

where $\mathfrak{D}_0(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2) = a\mathfrak{D}(\mathsf{P}_1; \mathscr{R}, \mu_1) + (1-a)\mathfrak{D}(\mathsf{P}_2; \mathscr{R}, \mu_2)$. This yields

$$\begin{aligned} n^{1/2}(\hat{\mathscr{R}} - \mathscr{R}) = &-\mathfrak{D}_0(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2)^{-1} \\ &\{n^{1/2}\mathscr{G}(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \mathscr{R}, \mu_1, \mu_2) + a_n^{1/2}\mathbb{D}(\mathsf{P}_1; \mathscr{R}^\ddagger, \mu_1^\ddagger) n_1^{1/2}(\hat{\mu}_1 - \mu_1) \\ &+ (1-a_n)^{1/2}\mathbb{D}(\mathsf{P}_2; \mathscr{R}^\ddagger, \mu_2^\ddagger) n_2^{1/2}(\hat{\mu}_2 - \mu_2)\} \\ &+ o_P(1); \end{aligned} \tag{S10}$$

here again the term $o_P(1)$ is present because we replace the empirical distributions by their theoretical counterparts in $\mathfrak{D}_0$ and $\mathbb{D}$.

The different Taylor expansions we have used contain various elements denoted by $^\diamond$, †, ‡ which lie on the line segments between the true and estimated corresponding parameters. We will replace all of these elements by the true values of the parameters. Due to the consistency of the estimators, the difference between a quantity at the true value of the parameters and at a value on the line segment between the true value and the estimator converges in probability to zero. Moreover, the quantities involving elements marked with $^\diamond$, † or ‡ are always multiplied by a term that is bounded in probability (by its convergence in distribution which will be seen later). Hence, the change we make by replacing the elements marked with $^\diamond$, † or ‡ by their true values is asymptotically negligible. The reason for doing this is that we obtain simpler formulas.

Denote

$$\begin{aligned} \mathfrak{H}_1(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2) &= \mathfrak{I} - \mathfrak{D}_1(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2)\mathfrak{D}_0(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2)^{-1}, \\ \mathbb{H}_1(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2) &= \mathfrak{H}_1(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2)\mathbb{D}(\mathsf{P}_1; \mathscr{R}, \mu_1)\mathscr{D}(\mathsf{P}_1; \mu_1)^{-1}, \\ \mathfrak{H}_2(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2) &= \mathfrak{I} + \mathfrak{D}_1(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2)\mathfrak{D}_0(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2)^{-1}, \\ \mathbb{H}_2(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2) &= \mathfrak{H}_2(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2)\mathbb{D}(\mathsf{P}_2; \mathscr{R}, \mu_2)\mathscr{D}(\mathsf{P}_2; \mu_2)^{-1}, \end{aligned}$$

where $\mathfrak{I}$ stands for the identity operator on $\mathrm{HS}(\mathcal{H}, \mathcal{H})$. Inserting (S9) and (S10) into (S8), we obtain

$$\begin{aligned} n^{1/2}\mathscr{B}(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \hat{\mathscr{R}}, \hat{\mu}_1, \hat{\mu}_2) = &\, a_n^{1/2}\mathfrak{H}_1(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2) n_1^{1/2}\mathscr{G}(\hat{\mathsf{P}}_1; \mathscr{R}, \mu_1) \\ &- a_n^{1/2}\mathbb{H}_1(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2) n_1^{1/2} G(\hat{\mathsf{P}}_1; \mu_1) \\ &- (1-a_n)^{1/2}\mathfrak{H}_2(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2) n_2^{1/2}\mathscr{G}(\hat{\mathsf{P}}_2; \mathscr{R}, \mu_2) \\ &+ (1-a_n)^{1/2}\mathbb{H}_2(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2) n_2^{1/2} G(\hat{\mathsf{P}}_2; \mu_2) \\ &+ o_P(1). \end{aligned}$$

The term $o_P(1)$ is due to the fact that we have replaced the quantities marked with $^\diamond$, †, ‡ by their true counterparts.

By the central limit theorem for Hilbert spaces (Bosq, 2000, Theorem 2.7), the operators $n_1^{1/2}\mathscr{G}(\hat{\mathsf{P}}_1; \mathscr{R}, \mu_1)$, $n_1^{1/2}G(\hat{\mathsf{P}}_1; \mu_1)$ jointly converge in distribution to a zero-mean Gaussian random variable in $\mathrm{HS}(\mathcal{H}, \mathcal{H}) \times \mathcal{H}$. The asymptotic covariance operator of $n_1^{1/2}\mathscr{G}(\hat{\mathsf{P}}_1; \mathscr{R}, \mu_1)$, i.e., an operator on operators on $\mathcal{H}$, can be estimated by the empirical covariance $\mathfrak{J}(\hat{\mathsf{P}}_1; \hat{\mathscr{R}}, \hat{\mu}_1)$, where

$$\mathfrak{J}(\mathsf{P}; \mathscr{R}, \mu) = E_{\mathsf{P}}\left(\left[\frac{\rho'\{\|\mathscr{P}(X;\mu) - \mathscr{R}\|\}}{\|\mathscr{P}(X;\mu) - \mathscr{R}\|}\{\mathscr{R} - \mathscr{P}(X;\mu)\} - \mathscr{G}(\mathsf{P}; \mathscr{R}, \mu)\right]^{\otimes 2}\right)$$

with the notation $\mathscr{A}^{\otimes 2} = \mathscr{A} \otimes \mathscr{A}$ for $\mathscr{A} \in \mathrm{HS}(\mathcal{H}, \mathcal{H})$, the asymptotic covariance operator of $n_1^{1/2} G(\hat{\mathsf{P}}_1; \mu_1)$, i.e., an operator on $\mathcal{H}$, can be estimated by $\mathscr{J}(\hat{\mathsf{P}}_1; \hat{\mu}_1)$, where

$$\mathscr{J}(\mathsf{P}; \mu) = E_{\mathsf{P}}\left[\left\{\frac{\rho'(\|X-\mu\|)}{\|X-\mu\|}(\mu - X) - G(\mathsf{P}; \mu)\right\}^{\otimes 2}\right]$$

with $f^{\otimes 2} = f \otimes f$ for $f \in \mathcal{H}$, and the asymptotic cross-covariance operator of $n_1^{1/2} \mathscr{G}(\hat{\mathsf{P}}_1; \mathscr{R}, \mu_1)$ and $n_1^{1/2} G(\hat{\mathsf{P}}_1; \mu_1)$, i.e., an operator from $\mathcal{H}$ to operators on $\mathcal{H}$, can be estimated by $\mathbb{J}(\hat{\mathsf{P}}_1; \hat{\mathscr{R}}, \hat{\mu}_1)$, where

$$\begin{aligned}\mathbb{J}(\mathsf{P}; \mathscr{R}, \mu) = E_{\mathsf{P}}\Bigg(&\left[\frac{\rho'\{\|\mathscr{P}(X;\mu) - \mathscr{R}\|\}}{\|\mathscr{P}(X;\mu) - \mathscr{R}\|}\{\mathscr{R} - \mathscr{P}(X;\mu)\} - \mathscr{G}(\mathsf{P}; \mathscr{R}, \mu)\right] \\ &\otimes \left\{\frac{\rho'(\|X-\mu\|)}{\|X-\mu\|}(\mu - X) - G(\mathsf{P}; \mu)\right\}\Bigg).\end{aligned}$$

Similarly, $n_2^{1/2} \mathscr{G}(\hat{\mathsf{P}}_2; \mathscr{R}, \mu_2)$, $n_2^{1/2} G(\hat{\mathsf{P}}_2; \mu_2)$ jointly converge in distribution to a zero-mean Gaussian random element with covariance estimators analogous to those mentioned above for the sample from P_1. As the samples are independent, all four random variables jointly converge in distribution.

Finally, it follows by Slutsky's theorem that the test operator $n^{1/2} \mathscr{B}(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \hat{\mathscr{R}}, \hat{\mu}_1, \hat{\mu}_2)$ is asymptotically distributed as a zero-mean Gaussian operator whose covariance operator can be consistently estimated by

$$\begin{aligned}\mathfrak{W}(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \hat{\mathscr{R}}, \hat{\mu}_1, \hat{\mu}_2) = &\, a_n \mathfrak{W}_1(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \hat{\mathscr{R}}, \hat{\mu}_1, \hat{\mu}_2) \\ &+ (1 - a_n)\mathfrak{W}_2(\hat{\mathsf{P}}_1, \hat{\mathsf{P}}_2, a_n; \hat{\mathscr{R}}, \hat{\mu}_1, \hat{\mu}_2),\end{aligned}$$

where

$$\begin{aligned}&\mathfrak{W}_1(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2) \\ &\quad = \mathfrak{H}_1(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2)\mathfrak{J}(\mathsf{P}_1; \mathscr{R}, \mu_1)\mathfrak{H}_1(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2)^* \\ &\quad\quad - \mathfrak{H}_1(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2)\mathbb{J}(\mathsf{P}_1; \mathscr{R}, \mu_1)\mathbb{H}_1(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2)^* \\ &\quad\quad - \mathbb{H}_1(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2)\mathbb{J}(\mathsf{P}_1; \mathscr{R}, \mu_1)^*\mathfrak{H}_1(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2)^* \\ &\quad\quad + \mathbb{H}_1(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2)\mathscr{J}(\mathsf{P}_1; \mathscr{R}, \mu_1)\mathbb{H}_1(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2)^*\end{aligned}$$

with * denoting adjoint operators, and $\mathfrak{W}_2(\mathsf{P}_1, \mathsf{P}_2, a; \mathscr{R}, \mu_1, \mu_2)$ is defined analogously with $\mathbb{H}_2, \mathfrak{H}_2$ in place of $\mathbb{H}_1, \mathfrak{H}_1$, respectively, and P_2 instead of P_1 in $\mathfrak{J}, \mathbb{J}, \mathscr{J}$.

REFERENCES

BOSQ, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*. New York: Springer.

GERVINI, D. (2008). Robust functional estimation using the median and spherical principal components. *Biometrika* **95**, 587–600.

NELSON, E. (1969). *Topics in Dynamics. I: Flows*. Princeton: Princeton University Press.