

SUPPLEMENT TO EVALUATING STATIONARITY VIA CHANGE-POINT ALTERNATIVES WITH APPLICATIONS TO FMRI DATA*

BY JOHN A. D. ASTON, AND CLAUDIA KIRCH

University of Warwick and Karlsruhe Institute of Technology

This supplement contains additional material including technical details and proofs for *Evaluating stationarity via change-point alternatives with applications to fMRI data* to be published in Annals of Applied Statistics.

Note: All numbers refer to main paper unless prefixed by S.

S.1. Technical Assumptions on the Functional Time Series.

S.1.1. *Basic assumption for the time series and change.* Throughout the paper we make the following very general very weak assumptions on the time series $\{Y_t(\cdot)\}$ and the change $\{\Delta = \mu_2 - \mu_1\}$. The mean functions before the change $\mu = \mu(\cdot)$, the change $\Delta = \Delta(\cdot)$ as well as the functional time series $\{Y_t(\cdot) : 1 \leq t \leq n\}$ are elements of $L^2(\mathcal{U})$. Furthermore we assume that the time series $\{Y_t(\cdot) : t \geq 1\}$ is centered, stationary and ergodic with

$$\mathbb{E} \|Y_t(\cdot)\|^2 = \int \mathbb{E}(Y_t^2(u)) du < \infty.$$

*This work as well as the position of the second author was financed by the Stifterverband für die Deutsche Wissenschaft by funds of the Claussen-Simon-trust. The first author was also supported by the Engineering and Physical Sciences Research Council (UK) through the CRiSM programme grant and by the project grant EP/H016856/1, and thanks SAMSI for hosting the author during which some of the work was carried out.

AMS 2000 subject classifications: Primary 62M10, 62M40, 62H35

Keywords and phrases: Epidemic change, Functional time series, High-dimensional data, Resting state fMRI, Separable covariance structure, Stationarity

S.1.2. *Component time series assumptions for null asymptotics*. To obtain the null asymptotics for our test statistics in (5.5), we need the following stronger assumptions on the projected time series $\boldsymbol{\eta}_t = (\eta_{t,1}, \dots, \eta_{t,d})^T$, where $\eta_{t,l} = \int Y_t(u) v_l(u) du$, $l = 1, \dots, d$, and $v_l(\cdot)$ is as in (5.2).

- The time series $\{\boldsymbol{\eta}_t : t \in \mathbb{Z}\}$ is stationary and short-range dependent i.e.

$$\sum_{t \in \mathbb{Z}} |\text{cov}(\eta_{0,l_1}, \eta_{t,l_2})| < \infty, \quad l_1, l_2 = 1, \dots, d.$$

- $\{\boldsymbol{\eta}_t\}$ fulfills the following functional limit theorem

$$\left\{ \frac{1}{\sqrt{n}} \sum_{1 \leq t \leq nx} \boldsymbol{\eta}_t : 0 \leq x \leq 1 \right\} \xrightarrow{D^d[0,1]} \{\boldsymbol{\Sigma}^{-1/2} \mathbf{W}_d(x) : 0 \leq x \leq 1\},$$

where \mathbf{W}_d is a standard d -dimensional Wiener process and the positive-definite covariance matrix $\boldsymbol{\Sigma}$ is defined by

$$\boldsymbol{\Sigma} = \sum_{k \in \mathbb{Z}} \Gamma(k), \quad \Gamma(j) = \mathbb{E} \boldsymbol{\eta}_t \boldsymbol{\eta}_{t+h}^T \text{ for } h \geq 0, \text{ and } \Gamma(h) = \Gamma(-h)^T \text{ for } h < 0.$$

These assumptions can be verified for general orthonormal systems $\{v_k(\cdot), k = 1, \dots, d\}$ under assumptions on the dependence structure of the original time series $\{Y_t(\cdot)\}$. For instance, it is fulfilled for strong mixing time series as well as $L^p - m$ -approximable time series (in the sense of Hörmann and Kokoszka (2010)). More details can be found in Aston and Kirch (2012).

S.2. Some Theory for Principal Components Based on Separable Covariance Structures. In this section we give some additional theoretic results for estimated principal components based on a separable covariance structure as well as a very general power result when using the corresponding separable dimension reduction in the change-point tests.

S.2.1. *Consistency results for the estimated separable principal components.* In this section we will prove that (5.2) and (5.3) hold when using the separable principal components. In particular, this gives consistency of the estimators in the correctly specified case, when the covariance kernel $c(u, s)$ is truly separable. However even if the true covariance kernel is not separable we still obtain (5.2) and (5.3) just with a different interpretation of the limit orthonormal system as stated in the theorem below.

Recall that we use the following estimator to obtain the separable marginal covariance kernels

$$(S.2.1) \quad \widehat{c}_1(u_1, s_1) = \int_{\mathcal{U}_2} \widehat{c}_n((u_1, z), (s_1, z)) dz$$

and

$$(S.2.2) \quad \widehat{c}_2(u_2, s_2) = \int_{\mathcal{U}_1} \widehat{c}_n((z, u_2), (z, s_2)) dz,$$

where $\widehat{c}_n((u_1, u_2), (s_1, s_2))$ is as in (3.5). Furthermore,

$$(S.2.3) \quad \begin{aligned} \widetilde{c}_1(u_1, s_1) &= \int_{\mathcal{U}_2} c((u_1, z), (s_1, z)) dz, & \widetilde{c}_2(u_2, s_2) &= \int_{\mathcal{U}_1} c((z, u_2), (z, s_2)) dz, \\ \widetilde{c}((u_1, u_2), (s_1, s_2)) &= \widetilde{c}_1(u_1, s_1) \widetilde{c}_2(u_2, s_2) \end{aligned}$$

and define $\widetilde{k}_1, \widetilde{k}_2, \widetilde{k}$ correspondingly by replacing c with the contaminated covariance kernel $k((u_1, u_2), (s_1, s_2))$ as in (5.10). Define

$$(S.2.4) \quad \widehat{v}_{(r,l)}(u_1, u_2) = \widehat{v}_{1,r}(u_1) \widehat{v}_{2,l}(u_2), \quad r = 1, \dots, d_1, l = 1, \dots, d_2,$$

where $\widehat{v}_{i,r}$ is the r th principal component of \widehat{c}_i as in (S.2.1) resp. (S.2.2).

The following theorem shows that the non-contaminated subspace as in (5.2) of this estimator is given as the corresponding subspace of principal components of \widetilde{c} while the contaminated subspace as in (5.3) is given as the corresponding subspace of principal components of \widetilde{k} .

While we do not need that the covariance kernel is separable for the result in this theorem, we do obtain consistency in the separable case without change as then (with the notation of the theorem below) $\widetilde{v}_{(r,l)}(u_1, u_2) = v_{1,r}(u_1) v_{2,l}(u_2)$ and those are the true eigenfunctions in the separable case.

To obtain the theorem we merely require (5.8) to hold under the null hypothesis and (5.9) under the alternative, where c and k are kernels but not necessarily separable. The nonparametric covariance estimator (3.5), for example, fulfills (5.8) as well as (5.9) for a large class of functional time series including mixing sequences or $L^p - m$ -approximable time series (in the sense of Hörmann and Kokoszka (2010)), for more details we refer to Aston and Kirch (2012).

- THEOREM S.2.1.** a) Let (5.8) hold and in addition $\tilde{\lambda}_{i,1} > \tilde{\lambda}_{i,2} > \dots > \tilde{\lambda}_{i,d_i+1} \geq 0$, $i = 1, 2$, where $\tilde{\lambda}_{i,l}$ are the largest eigenvalues of \tilde{c}_i . Let $\tilde{v}_{i,r}$ be the r th principal component of \tilde{c}_i .
Then $\hat{v}_{(r,l)}(u_1, u_2)$ as in (S.2.4) and $\tilde{v}_{(r,l)}(u_1, u_2) = \tilde{v}_{1,r}(u_1)\tilde{v}_{2,l}(u_2)$, $r = 1, \dots, d_1, l = 1, \dots, d_2$, fulfill (5.2).
- b) Let (5.9) hold and in addition $\tilde{\gamma}_{i,1} > \tilde{\gamma}_{i,2} > \dots > \tilde{\gamma}_{i,d_i+1}$, $i = 1, 2$, where $\tilde{\gamma}_{i,l}$ are the largest eigenvalues of \tilde{k}_i . Let $w_{i,r}$ be the r th principal components of \tilde{k}_i .
Then $\hat{w}_{(r,l)}(u_1, u_2)$ as in (S.2.4) and $\tilde{w}_{(r,l)}(u_1, u_2) = \tilde{w}_{1,r}(u_1)\tilde{w}_{2,l}(u_2)$, $r = 1, \dots, d_1, l = 1, \dots, d_2$, fulfill (5.3).

REMARK S.2.1. Note that $\tilde{v}_{(r,l)}$ resp. $\tilde{w}_{(r,l)}$ are eigenfunctions belonging to the eigenvalue $\tilde{\lambda}_{1,r}\tilde{\lambda}_{2,l}$ of \tilde{c} resp. to the eigenvalue $\tilde{\gamma}_{1,r}\tilde{\gamma}_{2,l} \geq 0$ of \tilde{k} . This 'balanced' choice of basis is guaranteed to include at least the eigenfunctions belonging to the largest $d = \min(d_1, d_2)$ eigenvalues of the joint covariance kernel \tilde{c} (resp. \tilde{k}).

Furthermore, this is preferable to choosing just the eigenfunctions belonging to the largest d joint eigenvalues for the following reason: Corollary 5.1 b) shows that any large enough separable change has a tendency to switch the eigenfunctions in such a way that it becomes detectable. However, if the change is large in one of the components and small in the other one examples can be constructed where the change is not orthogonal to the d joint noncontaminated subspace but orthogonal to the contaminated subspace. The reason is that the contaminated eigenvalues in one component may become very large while only increasing slightly in the other component. As a result it is possible for all $\hat{v}_{1,k}(u_1) \cdot \hat{v}_{2,l}(u_2)$ corresponding to the joint largest eigenvalues that the first change Δ_1 , e.g., is not orthogonal to $w_{1,k}$ but Δ_2 is orthogonal to $w_{2,l}$. But then $\Delta(u_1, u_2) = \Delta_1(u_1)\Delta_2(u_2)$ is orthogonal to the joint eigenspace belonging to the largest contaminated eigenvalues and hence not detectable. This cannot happen if one chooses a balanced basis of the products of the eigenfunctions belonging to the largest d_1 eigenvalues in the first dimension and largest d_2 eigenvalues in the second dimension (cf. Corollary 5.1 a)).

S.2.2. Power analysis for a general change. The following theorem characterizes the contaminated eigenfunction basis and its relation to the change for the separable estimation procedure. In part a) of the theorem a characterisation via the contaminated projection subspace is given for all changes that

are detectable by the uncontaminated projection subspace, i.e. for which

$$\int_{\mathcal{U}_1} \int_{\mathcal{U}_2} \Delta(u_1, u_2) \tilde{v}_{1,r}(u_1) \tilde{v}_{2,l}(u_2) du_1 du_2 \neq 0$$

for some r, l . Unlike in the situation where one uses a fully nonparametric covariance estimator to obtain the projection subspace (cf. Theorem 3.2 in [Aston and Kirch \(2012\)](#)) we cannot conclude that the change is also detectable by the contaminated (separable) projection subspace, i.e. in general it does not hold that

$$\int_{\mathcal{U}_1} \int_{\mathcal{U}_2} \Delta(u_1, u_2) \tilde{w}_{1,r}(u_1) \tilde{w}_{2,l}(u_2) du_1 du_2 \neq 0$$

for some r, l . Only the weaker statement given in the theorem can be derived. However, usually this will not cause a problem and the change will remain detectable. In the special case of a separable change, i.e. $\Delta(u_1, u_2) = \Delta_1(u_1)\Delta_2(u_2)$, detectability of a change by the uncontaminated projection subspace and the assertion obtained in the theorem are equivalent, showing that any separable change detectable by the uncontaminated projection subspace is also detectable by the contaminated subspace. This is the separable analogue of Theorem 3.2 a) from [Aston and Kirch \(2012\)](#).

Part b) of the theorem gives a sufficient condition for which a large enough change is detectable using only a one-dimensional projection subspace. In case of a separable change this condition is fulfilled for any change, showing that any large enough separable change is detectable even when using only a one-dimensional subspace. This is the separable analogue of Theorem 3.2 b) in [Aston and Kirch \(2012\)](#).

Note that it is not needed that c is separable.

THEOREM S.2.2. *a) Let $\tilde{v}_{j,r}$ be the r th principal component of \tilde{c}_j and $\tilde{w}_{j,r}$ be the r th principal component belonging to \tilde{k}_j and let*

$$k((u_1, u_2), (s_1, s_2)) = c((u_1, u_2), (s_1, s_2)) + \theta(1 - \theta)\Delta(u_1, u_2)\Delta(s_1, s_2)$$

as in (5.10). Then it holds:

$$\int_{\mathcal{U}_1} \int_{\mathcal{U}_2} \Delta(u_1, u_2) \tilde{v}_{1,r}(u_1) \tilde{v}_{2,l}(u_2) du_1 du_2 \neq 0, \quad \text{for some } 1 \leq r \leq d_1, 1 \leq l \leq d_2$$

$$\implies \int_{\mathcal{U}_1} \Delta(u_1, u_2) \tilde{w}_{1,r}(u_1) du_1 \neq 0 \quad \text{for some } 1 \leq r \leq d_1$$

$$\text{and } \int_{\mathcal{U}_2} \Delta(u_1, u_2) w_{2,r}(u_2) du_2 \neq 0 \quad \text{for some } 1 \leq l \leq d_2.$$

b) Let $\Delta_D(u_1, u_2) = D\Delta(u_1, u_2)$ for some $\Delta(u_1, u_2)$ with $\int_{\mathcal{U}_1} \int_{\mathcal{U}_2} \Delta^2(u_1, u_2) du_1 du_2 \neq 0$. Let $\tilde{w}_{j,k,D}$, $k = 1, \dots, d_j$, be the normalized principal components of the covariance kernel $\tilde{k}_{j,D}$ obtained analogously to (S.2.3) with

$$k_D((u_1, u_2), (s_1, s_2)) = c((u_1, u_2), (s_1, s_2)) + \theta(1-\theta)\Delta_D(u_1, u_2)\Delta_D(s_1, s_2).$$

Let $\psi_j(u_1, s_1) = \int \Delta(u_1, z)\Delta(s_1, z) dz$ resp. $\psi_2(u_2, s_2) = \int \Delta(z, u_2)\Delta(z, s_2) dz$ have decreasing eigenvalues $\xi_{j,1} > \xi_{j,2} \geq \xi_{j,3} \geq \dots$ with corresponding normalised eigenfunctions $x_{j,1}, \dots$

Then, for $j = 1, 2$, as $D \rightarrow \infty$,

$$\|s_{j,1}\tilde{w}_{j,1,D}(\cdot) - x_{j,1}(\cdot)\| \rightarrow 0, \quad s_{j,1} = \text{sgn}\left(\int \tilde{w}_{j,1,D}(z)x_{j,1}(z) dz\right).$$

In particular, there exists $D_0 > 0$ such that

$$\int \Delta_D(u_1, u_2)\tilde{w}_{1,1,D}(u_1)\tilde{w}_{2,1,D}(u_2) du_1 du_2 \neq 0,$$

for all $|D| \geq D_0$, if

$$\int \Delta(u_1, u_2)x_{1,1}(u_1)x_{2,1}(u_2) du_1 du_2 \neq 0.$$

S.3. Some additional aspects of estimation of the temporal covariance matrix. In this section we give some additional information on possible adaptations of the temporal long-run covariance matrix needed for the change-point procedure.

As pointed out in Section 6.1 most estimators for the long-run covariance matrix are based on

$$\hat{\Sigma} = \sum_{|h| \leq b_n} w_q(h/b_n)\hat{\Gamma}(h),$$

for some appropriate weight function w_q and bandwidth b_n where $\hat{\Gamma}(\cdot)$ is an estimator for the autocovariance matrix of the (uncontaminated) projected data vector.

Recall that

$$(\hat{m}_{1,l}, \hat{m}_{2,l}) = \arg \max_{k_1, k_2} \left(\left| \sum_{t=k_1}^{k_2} \hat{\eta}_{t,l} - \frac{k_2 - k_1}{n} \sum_{t=1}^n \hat{\eta}_{t,l} \right| \right)$$

are the estimated change-points that are estimated separately in each component and let

(S.3.1)

$$\widehat{\mathbf{e}}(t) = (\widehat{e}_1(t), \dots, \widehat{e}_d(t))^T,$$

$$\text{where } \widehat{e}_l(t) = \widehat{\eta}_{t,l} - \widetilde{\eta}_{\widehat{m}_{1,l}, \widehat{m}_{2,l}} \mathbf{1}_{\{\widehat{m}_{1,l} < t \leq \widehat{m}_{2,l}\}} - \widetilde{\eta}_{\widehat{m}_{1,l}, \widehat{m}_{2,l}}^{\circ} \mathbf{1}_{\{t \leq \widehat{m}_{1,l}, \widehat{m}_{2,l} < t\}},$$

$$\widetilde{\eta}_{\widehat{m}_{1,l}, \widehat{m}_{2,l}} = \frac{1}{\widehat{m}_{2,l} - \widehat{m}_{1,l}} \sum_{t=\widehat{m}_{1,l}+1}^{\widehat{m}_{2,l}} \widehat{\eta}_{t,l},$$

$$\widetilde{\eta}_{\widehat{m}_{1,l}, \widehat{m}_{2,l}}^{\circ} = \frac{1}{n - \widehat{m}_{2,l} + \widehat{m}_{1,l}} \sum_{1 \leq t \leq \widehat{m}_{1,l}, \widehat{m}_{2,l} < t \leq n} \widehat{\eta}_{t,l},$$

be the estimated uncontaminated data. With this notation, we obtain an estimator of the uncontaminated autocovariance matrix as

$$\widehat{\Gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-r} \widehat{\mathbf{e}}(t) \widehat{\mathbf{e}}(t+h)^T, \quad h \geq 0, \quad \widehat{\Gamma}(h) = \widehat{\Gamma}(-h), \quad h < 0.$$

We use the following flat-top kernel

$$w(x) = \begin{cases} 1, & |x| \leq 1/2, \\ 2(1 - |x|), & 1/2 < |x| < 1, \\ 0, & |x| \geq 1, \end{cases}$$

and the bandwidth $B_{l,k} = B_{k,l} = 2 \max(\widehat{b}_{l,k}, \widehat{b}_{k,l})$, where $\widehat{b}_{l,k}$ is the smallest positive integer such that

$$\left| \widehat{\Gamma}_{l,k}(\widehat{b}_{l,k} + j) / \sqrt{\widehat{\Gamma}_{l,l}(0) \widehat{\Gamma}_{k,k}(0)} \right| < 1.4 \sqrt{\log_{10} n/n}, \quad \text{for } j = 1, \dots, 3.$$

Using the matrix of entries

$$\widehat{\Sigma}^{(1)} = \left(\sum_{|h| \leq b_{k,l}} w_q(h/b_{k,l}) \widehat{\Gamma}_{k,l}(h) \right)_{k,l}$$

gives a symmetric estimator. However, while it is asymptotically positive definite, this is not necessarily true for small samples, where it can even have negative eigenvalues. In fact, in the fMRI application described in this paper

the long-run covariance matrix had dimension 64×64 and estimation was conducted based on 225 data vectors only. The effect is that the estimation error can become rather large resulting in as much as thirty percent negative eigenvalues.

If one is only interested in a good positive definite estimator of the long-run covariance matrix but not its inverse as in our case, then the following approach yields reasonable results and was suggested by Politis (2009) to overcome this deficiency. Consider the orthogonal diagonalization

$$\widehat{\Sigma}^{(1)} = UDU^T$$

where U is an orthogonal matrix, $D = \text{diag}(\delta_1, \dots, \delta_d)$ with $\delta_1 \geq \delta_2 \geq \dots \geq \delta_d$, and $D^+ = \text{diag}(\delta_1^+, \dots, \delta_d^+)$, $\delta_j^+ = \max(\delta_j, M_C/a_n)$, where $a_n \rightarrow \infty$ and

$$M_C = \text{Median}(m_1, \dots, m_d),$$

where m_1, \dots, m_d are the eigenvalues of the estimated non-contaminated covariance matrix

$$\left(\frac{1}{n-1} \sum_{t=1}^n e_l(t)e_k(t) \right)_{k,l}.$$

This choice ensures that D^+ is scale invariant and asymptotically equal to D . A symmetric and positive definite estimator for Σ is thus given by

$$\widehat{\Sigma}^{(2)} = UD^+U^T$$

For the change-point procedure in this paper an estimator for the inverse of Σ is needed, which can easily be obtained from the above decomposition as

$$\widehat{\Sigma}^{(2)-1} = U(D^+)^{-1}U^T.$$

If the eigenvalues of the estimator are arbitrarily small, then the eigenvalues of the inverse become arbitrarily large, which in turn causes the change-point statistic to become very large. A cut-off point M_C/a_n as above solves the problem in principle, but if it is chosen too small it will cause the change-point statistic to reject. For this reason, we used the following estimator in

our data example for the inverse

$$\begin{aligned}\widehat{\Sigma}^{(3)-1} &= U\widetilde{D}^{-1}U^T, \\ \text{where } \widetilde{D}^{-1} &= \text{diag}\left(\widetilde{\delta}_1^{-1}, \dots, \widetilde{\delta}_d^{-1}\right), \\ \widetilde{\delta}_j^{-1} &= \begin{cases} \delta_j^{-1}, & \delta_j \geq M_C/a_n, \\ 0, & \text{else.} \end{cases}\end{aligned}$$

This is a conservative estimator because the value of the change-point statistic using this estimator will be smaller than any instance where δ_j had been set to any fixed small value. In the application a_n was set equal to $\log n$ and in all cases $M_C/a_n \approx 0.5$. Nevertheless even with this conservative cut-off point the null hypothesis of stationarity was rejected for all subjects in the Connectome data set.

In conclusion, if the dimension of the projection subspace is large in comparison to the length of the time series or if the time series deviates from stationarity in a different way than exhibiting an epidemic change, this estimator does not perform satisfactory. In this case it leads to a more stable and conservative change detection procedure if one only corrects for the long-run variance, setting all non-diagonal elements of the matrix equal to zero leading to the statistics discussed in Section 6.1.

S.4. Proofs.

OF THEOREM S.2.1. By the Cauchy-Schwarz inequality it holds

$$\begin{aligned}& n \int_{\mathcal{U}_1} \int_{\mathcal{U}_1} \left(\int_{\mathcal{U}_2} \widehat{c}_n((u_1, z), (s_1, z)) - c((u_1, z), (s_1, z)) dz \right)^2 du_1 ds_1 \\ & \leq \int_{\mathcal{U}_2} 1 dz n \int_{\mathcal{U}_1} \int_{\mathcal{U}_1} \int_{\mathcal{U}_2} (\widehat{c}_n((u_1, z), (s_1, z)) - c((u_1, z), (s_1, z)))^2 dz du_1 ds_1 \xrightarrow{P} 0,\end{aligned}$$

where the convergence follows from the assumptions of the theorem as well as the continuity of c and *a.s.* continuity of \widehat{c}_n . By Theorem 2.1 of [Aston and Kirch \(2012\)](#) we can conclude that

$$\int_{\mathcal{U}_1} (\widehat{v}_{1,l}(u_1) - \text{sgn}_{1,l} \widetilde{v}_{1,l}(u_1))^2 du_1 = O_P(n^{-1}),$$

where $\text{sgn}_{1,l} = \pm 1$. Similarly one gets

$$\int_{\mathcal{U}_2} (\widehat{v}_{2,l}(u_2) - \text{sgn}_{2,l} \widetilde{v}_{2,l}(u_2))^2 du_2 = O_P(n^{-1}).$$

Putting the two together yields an orthonormal system fulfilling (5.2). The proof of b) is analogous. \square

OF THEOREM S.2.2. It holds

$$\tilde{k}_1(u_1, u_2) = \tilde{c}_1(u_1, u_2) + \theta(1 - \theta) \int_{\mathcal{U}_2} \Delta(u_1, z) \Delta(s_1, z) dz,$$

Analogously

$$\tilde{k}_2(s_1, s_2) = \tilde{c}_2(s_1, s_2) + \theta(1 - \theta) \int_{\mathcal{U}_1} \Delta(z, u_2) \Delta(z, s_2) dz.$$

We are now ready to prove a).

We prove the assertion by contradiction. Thus either

$$\int_{\mathcal{U}_1} \Delta(u_1, u_2) \tilde{w}_{1,r}(u_1) du_1 \equiv 0 \quad \text{for all } r = 1, \dots, d_1,$$

or

$$\int_{\mathcal{U}_2} \Delta(u_1, u_2) \tilde{w}_{2,l}(u_2) du_2 \equiv 0 \quad \text{for all } l = 1, \dots, d_2.$$

Analogous to the proof of Theorem 3.2 in [Aston and Kirch \(2012\)](#) we obtain

$$\int_{\mathcal{U}_1} \Delta(u_1, u_2) \tilde{v}_{1,r}(u_1) du_1 \equiv 0 \quad \text{for all } r = 1, \dots, d_1,$$

or

$$\int_{\mathcal{U}_2} \Delta(u_1, u_2) \tilde{v}_{2,l}(u_2) du_2 \equiv 0 \quad \text{for all } l = 1, \dots, d_2.$$

This in turn implies

$$\int_{\mathcal{U}_1} \int_{\mathcal{U}_2} \Delta(u_1, u_2) \tilde{v}_{1,r}(u_1) \tilde{v}_{2,l}(u_2) du_1 du_2 = 0, \quad \text{for all } 1 \leq r \leq d_1, 1 \leq l \leq d_2.$$

For b) we show the result for $j = 1$, the assertion for $j = 2$ follows analogously. The eigenvectors of \tilde{k}_1/D^2 are the same as those of \tilde{k}_1 , while the eigenvalues are also divided by D^2 hence remain in the same order. As $D \rightarrow \infty$ it holds

$$\begin{aligned} \frac{\tilde{k}_1(u_1, s_1)}{D^2} &= \frac{\tilde{c}_1(u_1, s_1)}{D^2} + \theta(1 - \theta)\psi_1(u_1, s_1) \\ &\rightarrow \theta(1 - \theta)\psi_1(u_1, s_1). \end{aligned}$$

By assumption, and Theorem 2.1 in [Aston and Kirch \(2012\)](#), it holds

$$\pm \tilde{w}_{1,k,D} \rightarrow x_{1,k}, \quad k = 1, \dots, d_1.$$

The second assertion of b) follows by the continuity of the scalar product in a Hilbert space. \square

OF COROLLARY 5.1. In case of separability it holds

$$\begin{aligned} \int_{\mathcal{U}_1} \Delta(u_1, u_2) \tilde{w}_{1,r}(u_1) du_1 &= \Delta_2(u_2) \int_{\mathcal{U}_1} \Delta_1(u_1) \tilde{w}_{1,r}(u_1) du_1 \neq 0 \\ \implies \int_{\mathcal{U}_1} \Delta_1(u_1) \tilde{w}_{1,r}(u_1) du_1 &\neq 0. \end{aligned}$$

Analogously

$$\int_{\mathcal{U}_2} \Delta_2(u_2) \tilde{w}_{2,l}(u_2) du_2 \neq 0,$$

hence assertion a) follows from Theorem S.2.2 a).

For b) note that in case of separability

$$\int_{\mathcal{U}_1} \int_{\mathcal{U}_2} \Delta(u_1, u_2) \Delta_1(u_1) \Delta_2(u_2) du_1 du_2 = \int_{\mathcal{U}_1} \int_{\mathcal{U}_2} \Delta_1(u_1)^2 \Delta_2(u_2)^2 du_1 du_2 \neq 0,$$

since $\Delta \neq 0$. Hence

$$\int_{\mathcal{U}_2} \Delta(u_1, z) \Delta(s_1, z) dz = \int_{\mathcal{U}_2} \Delta_2^2(z) dz \Delta_1(u_1) \Delta_1(s_1),$$

has rank 1, i.e. only one non-zero eigenvalue. The corresponding eigenfunction is $\Delta_1(u_1)$ and the normalised eigenfunction is $\Delta_j / \|\Delta_j\|$, $j = 1, 2$, so that b) follows from Theorem S.2.2 b). \square

OF THEOREM 8.1. Let $\hat{F}_m(x) = \frac{1}{m} \sum_{i=1}^m 1_{\{\vartheta_i \leq x\}}$ be the (unobservable) empirical distribution function of F_ϑ . By the Glivenko-Cantelli lemma we know as $m \rightarrow \infty$

$$(S.4.1) \quad \sup_x |\hat{F}_m(x) - F_\vartheta(x)| \rightarrow 0 \quad a.s.$$

Hence it is sufficient to show that $\sup_x |\hat{F}_{\hat{\vartheta},m}(x) - \hat{F}_m(x)| \rightarrow 0$ a.s. It holds

$$\left| \hat{F}_{\hat{\vartheta},m}(x) - \hat{F}_m(x) \right| \leq \frac{1}{m} \sum_{i=1}^m 1_{\{A_i(x)\}},$$

where

$$A_i(x) = \begin{cases} \{\vartheta_i < x \leq \widehat{\vartheta}_i\}, & \vartheta_i \leq \widehat{\vartheta}_i, \\ \{\widehat{\vartheta}_i < x \leq \vartheta_i\}, & \vartheta_i > \widehat{\vartheta}_i. \end{cases}$$

For any $\epsilon > 0$ it holds

$$A_i(x) \subset \{|\vartheta_i - x| < \epsilon\} \cup \{|\widehat{\vartheta}_i - \vartheta_i| \geq \epsilon\},$$

hence

$$\mathbf{1}_{\{A_i(x)\}} \leq \mathbf{1}_{\{|\vartheta_i - x| < \epsilon\}} + \mathbf{1}_{\{|\widehat{\vartheta}_i - \vartheta_i| \geq \epsilon\}}.$$

Consider $Y_m(i) := \mathbf{1}_{\{|\widehat{\vartheta}_i - \vartheta_i| \geq \epsilon\}} - P(|\widehat{\vartheta}_i - \vartheta_i| \geq \epsilon)$, a triangular array of rowwise i.i.d. centered random variables with $|Y_m(i)| \leq 2$ and

$$\mathbb{E} \left| \frac{1}{m} \sum_{i=1}^m Y_m(i) \right|^4 \leq C \frac{1}{m^2}$$

for some constant $C > 0$. An application of the Markov inequality as well as of the Borel-Cantelli Lemma thus shows $\frac{1}{m} \sum_{i=1}^m Y_m(i) \rightarrow 0$ *a.s.*, which implies in turn

$$\frac{1}{m} \sum_{i=1}^m \mathbf{1}_{\{|\widehat{\vartheta}_i - \vartheta_i| \geq \epsilon\}} \rightarrow 0 \quad \text{a.s.},$$

since by (8.1) and the dominated convergence theorem

$$P(|\widehat{\vartheta}_i - \vartheta_i| \geq \epsilon) = \mathbb{E} P^* (|\widehat{\vartheta}_i - \vartheta_i| \geq \epsilon) \rightarrow 0.$$

Noting that

$$\mathbf{1}_{\{|\vartheta_i - x| < \epsilon\}} \leq \mathbf{1}_{\{\vartheta_i \leq x + \epsilon\}} - \mathbf{1}_{\{\vartheta_i \leq x - \epsilon\}},$$

an application of (S.4.1) shows that uniformly in ϵ

$$\begin{aligned} & \sup_x \left| \frac{1}{m} \sum_{i=1}^m \mathbf{1}_{\{|\vartheta_i - x| < \epsilon\}} \right| \\ & \leq \sup_x \left| \widehat{F}_m(x + \epsilon) - \widehat{F}_m(x - \epsilon) \right| \leq \sup_x |F_\vartheta(x + \epsilon) - F_\vartheta(x - \epsilon)| + o(1) \quad \text{a.s.} \end{aligned}$$

which becomes arbitrarily small for $\epsilon \rightarrow 0$ by the uniform continuity of F_ϑ . The uniform continuity holds by the continuity and the fact that $F_\vartheta(x) = 0$ for $x \leq 0$ and $F_\vartheta(x) = 1$ for $x \geq 1$. Putting everything together yields the assertion. \square

OF THEOREM 8.2. Note that

$$\mathbb{E} \left| \widehat{f}_{\widehat{\vartheta},m}(x) - \widehat{f}_m(x) \right|^2 = \text{var} \left(\widehat{f}_{\widehat{\vartheta},m}(x) - \widehat{f}_m(x) \right) + \left(\mathbb{E} \left(\widehat{f}_{\widehat{\vartheta},m}(x) - \widehat{f}_m(x) \right) \right)^2.$$

By the boundedness of K and since $\int \frac{1}{h} K \left(\frac{x-c}{h} \right) dx = \int K(x) dx = 1$ it holds for the first term

$$\begin{aligned} & \int \text{var} \left(\widehat{f}_{\widehat{\vartheta},m}(x) - \widehat{f}_m(x) \right) dx \\ & \leq \frac{1}{mh^2} \int \mathbb{E} \left(K \left(\frac{x - \widehat{\vartheta}_1}{h} \right) - K \left(\frac{x - \vartheta_1}{h} \right) \right)^2 dx \\ & \leq \frac{1}{mh} \left(\mathbb{E} \int \frac{1}{h} K \left(\frac{x - \widehat{\vartheta}_1}{h} \right) dx + \mathbb{E} \int \frac{1}{h} K \left(\frac{x - \vartheta_1}{h} \right) dx \right) = \frac{1}{mh} \rightarrow 0. \end{aligned}$$

It holds

$$\int \left| \mathbb{E} \left(K(y) - K \left(y + \frac{\vartheta_1 - \widehat{\vartheta}_1}{h} \right) \right) \right| dy \leq 2,$$

as well as by the boundedness and Lipschitz-continuity of the kernel K

$$\begin{aligned} & \left| \mathbb{E} \left(K(y) - K \left(y + \frac{\vartheta_1 - \widehat{\vartheta}_1}{h} \right) \right) \right| \\ & = O(1) \mathbb{E} \left(\min \left(1, \left| \frac{\vartheta_1 - \widehat{\vartheta}_1}{h} \right| \right) \right) \\ & = O(1) \mathbb{E} \left(1_{\{|\vartheta_1 - \widehat{\vartheta}_1| \geq \epsilon h\}} \right) + O(1) \mathbb{E} \left(\left| \frac{\widehat{\vartheta}_1 - \vartheta_1}{h} \right| 1_{\{|\vartheta_1 - \widehat{\vartheta}_1| < \epsilon h\}} \right) \\ & \leq \mathbb{E} P^*(|\vartheta_1 - \widehat{\vartheta}_1| \geq \epsilon h) + \epsilon = o(1) + \epsilon \end{aligned}$$

by (8.2) and the dominated convergence theorem. We now conclude

$$\begin{aligned} & \int \left(\mathbb{E} \left(\widehat{f}_m(x) - \widehat{f}_{\widehat{\vartheta},m}(x) \right) \right)^2 dx \\ & = \int \left[\frac{1}{h} \mathbb{E} \left(K \left(\frac{x - \vartheta_1}{h} \right) - K \left(\frac{x - \widehat{\vartheta}_1}{h} \right) \right) \right]^2 dx \\ & = \int \left[\mathbb{E} \left(K(y) - K \left(y + \frac{\vartheta_1 - \widehat{\vartheta}_1}{h} \right) \right) \right]^2 dy = o(1). \end{aligned}$$

□

References.

- ASTON, J. A. D. and KIRCH, C. (2012). Detecting and estimating changes in dependent functional data. *Journal of Multivariate Analysis* **109** 204–220.
- HÖRMANN, S. and KOKOSZKA, P. (2010). Weakly dependent functional data. *Ann. Statist.* **38** 1845–1884.
- POLITIS, D. N. (2009). Higher-order accurate, positive semi-definite estimation of large-sample covariance and spectral density matrices. Preprint: Department of Economics, UCSD, Paper 2005-03R, <http://repositories.cdlib.org/ucsdecon/2005-03R>.

CRISM,
DEPARTMENT OF STATISTICS,
UNIVERSITY OF WARWICK,
COVENTRY,
CV4 7AL,
UK
E-MAIL: j.a.d.aston@warwick.ac.uk

KARLSRUHE INSTITUTE OF TECHNOLOGY (KIT),
INSTITUTE FOR STOCHASTICS,
KAISERSTR. 89,
D – 76133 KARLSRUHE,
GERMANY
E-MAIL: claudia.kirch@kit.edu