Estimation of a change–point in the mean function of functional data*

Alexander Aue[†] Robertas Gabrys[‡] Lajos Horváth[§] Piotr Kokoszka[¶]

Abstract

The paper develops a comprehensive asymptotic theory for the estimation of a change–point in the mean function of functional observations. We consider both the case of a constant change size, and the case of a change whose size approaches zero, as the sample size tends to infinity. We show how the limit distribution of a suitably defined change–point estimator depends on the size and location of the change. The theoretical insights are confirmed by a simulation study which illustrates the behavior of the estimator in finite samples.

Keywords: Change-Point estimation; Mean function; Functional data analysis;

Abbreviated Title: Functional Change-Point Estimators;

AMS 2000 Subject Classification: Primary 62H25, 62G05; Secondary 62G20

*Research partially supported by NSF grants DMS 0604670 and DMS 0652420

[†]Department of Statistics, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA, email: alexaue@wald.ucdavis.edu

[‡]Department of Mathematics and Statistics, Utah State University, 3900 Old Main Hill, Logan, UT 84322–3900 USA, email: gabrys.robertas@gmail.com

[§]Department of Mathematics, University of Utah, 155 South 1440 East, Salt Lake City, UT 84112–0090, USA, email: horvath@math.utah.edu

[¶]Department of Mathematics and Statistics, Utah State University, 3900 Old Main Hill, Logan, UT 84322-3900, USA email: Piotr.Kokoszka@usu.edu

1 Introduction

Functional data analysis (FDA) has been enjoying increased attention over the last decade due to its applicability to problems which are difficult to cast into a framework of scalar or vector observations. Even if such standard approaches are available, the functional approach often leads to a more natural and parsimonious description of the data, and to more accurate inference and prediction results, see, for example, Antoniadis and Sapatinas (2003), Chiou *et al.* (2004), Fernández de Castro *et al.* (2005), Laukaitis and Račkauskas (2005), Müller and Stadtmüller (2005), Yao *et al.* (2005), Glendinning and Fleet (2007), Kokoszka *et al.* (2008a, 2008b). Both inferential and exploratory tools of FDA can however be severely biased if the stochastic structure of the data changes at some unknown point within the sample. In the scalar context, this issue has received considerable attention, see Cobb (1978), Inclán and Tiao (1994), Davis *et al.* (1995), Antoch *et al.* (1997), Garcia and Ghysels (1998), Horváth *et al.* (1999), Kokoszka and Leipus (2000), among many others.

The most important change that can occur in the functional context is the change of the mean function. This paper investigates large sample properties of an estimator of such a change–point. We consider both the case of a fixed size change and a contiguous change whose size approaches zero as the sample size increases. Specifically, we assume that the functional observations $X_1, \ldots, X_n$ are defined on a compact set $\mathcal{T}$ and follow the model

$$X_i = \mu + \Delta \mathbb{I}\{i > k^*\} + Y_i, \qquad i = 1, \ldots, n, \tag{1.1}$$

where μ and $\Delta \neq 0$ are unknown, square integrable and deterministic functions over $\mathcal{T}$, and $Y_1, \ldots, Y_n$ are independent, identically distributed zero mean random elements of $\mathcal{L}^2(\mathcal{T})$ with covariance function

$$K(s,t) = \mathbb{E}[Y_1(s)Y_1(t)], \qquad s,t \in \mathcal{T},$$

satisfying $\mathbb{E}[\|Y_1\|^2] = \int_{\mathcal{T}} \mathbb{E}[Y_1^2(t)]dt < \infty$. The unknown integer $k^* \in \{1, \ldots, n\}$ is called the change-point. We assume that

$$k^* = \lfloor \theta n \rfloor \qquad \text{with some fixed } \theta \in (0,1]. \tag{1.2}$$

Model (1.1) describes a sequence of functional observations which suffer from a mean change if $k^* < n$ or, equivalently, if $\theta < 1$. The corresponding hypothesis testing problem

$$H_0 : \ k^* = n \qquad \text{vs} \qquad H_A : \ k^* < n$$

has been addressed in Berkes *et al.* (2007). To explain their results and present our contribution, we must state several consequences of the assumptions made so far. First, Mercer's theorem (see Chapter 4 of Indritz, 1963) implies that, under the null hypothesis, there is a spectral decomposition for the covariance operator $K(s,t)$, namely

$$K(s,t) = \sum_{\ell=1}^{\infty} \lambda_\ell \varphi_\ell(s) \varphi_\ell(t), \qquad s,t \in \mathcal{T},$$

where λ_ℓ and φ_ℓ denote the eigenvalues and eigenfunctions of $K(s,t)$, respectively. These can be obtained as the solutions of the equation system $\int_{\mathcal{T}} K(s,t)\varphi_\ell(t)dt = \lambda_\ell \varphi_l(s)$ with $s, t \in \mathcal{T}$. Since the eigenfunctions form a complete orthonormal basis in $\mathcal{L}^2(\mathcal{T})$ and all eigenvalues of $K(s,t)$ are nonnegative, they lead to the the Karhunen-Loéve representation (in $\mathcal{L}^2(\mathcal{T})$, not pointwise in $t \in \mathcal{T}$)

$$Y_i(t) = \sum_{\ell=1}^{\infty} \sqrt{\lambda_\ell}\rho_{i,\ell}\varphi_\ell(t), \qquad t \in \mathcal{T},\ i = 1, \ldots, n,$$

where $\sqrt{\lambda_\ell}\rho_{i,\ell} = \int_{\mathcal{T}} Y_i(t)\varphi_\ell(t)dt$ is called the ℓth functional principal component score. It is also implied that the sequences $(\rho_{i,\ell})_{\ell\geq 1}$ consist of uncorrelated random variables with zero mean and unit variance and that, for $i \neq j$, $(\rho_{i,\ell})_{\ell\geq 1}$ and $(\rho_{j,\ell})_{\ell\geq 1}$ are independent.

For the statistical analysis, the population eigenvalues and eigenfunctions have to be replaced by their estimated versions. These are based on the estimated covariance operator

$$\hat{K}(s,t) = \frac{1}{n}\sum_{i=1}^{n}[X_i(s) - \bar{X}_n(s)][X_i(t) - \bar{X}_n(t)], \tag{1.3}$$

where $\bar{X}_n = n^{-1}(X_1 + \ldots + X_n)$. From this, estimated eigenvalues $\hat{\lambda}_\ell$ and eigenfunctions $\hat{\varphi}_\ell$ can then be derived as the solutions of the equations

$$\int_{\mathcal{T}} \hat{K}(s,t)\hat{\varphi}_\ell(t)dt = \hat{\lambda}_\ell\hat{\varphi}_\ell(s).$$

We make the assumption

$$\lambda_1 > \lambda_2 > \ldots > \lambda_d > \lambda_{d+1}, \tag{1.4}$$

which together with the assumption of finite fourth moment of the Y_i guarantees that the estimated and population eigenvalues and eigenfunctions are sufficiently close under H_0, see Chapter 4 of Bosq (2000) and Dauxois *et al.* (1982).

The hypothesis test for H_0 versus H_A in Berkes *et al.* (2007) is based on the projection of the functions $\bar{X}_{\lfloor nx\rfloor} - \bar{X}_n$, $x \in (0,1)$, on the space spanned by the first d estimated eigenfunctions $\hat{\varphi}_1, \ldots, \hat{\varphi}_d$. The corresponding estimated scores are

$$\hat{\eta}_{i,\ell} = \int_{\mathcal{T}} [X_i(t) - \bar{X}_n(t)]\hat{\varphi}_\ell(t)dt.$$

Berkes *et al.* (2007) introduced the test statistic

$$S_{n,d} = \frac{1}{n^2}\sum_{\ell=1}^{d}\frac{1}{\hat{\lambda}_\ell}\sum_{k=1}^{n}\left(\sum_{i=1}^{k}\hat{\eta}_{i,\ell} - \frac{k}{n}\sum_{i=1}^{n}\hat{\eta}_{i,\ell}\right)^2$$

and established its limit distribution under the null hypothesis, as well as its consistency under the alternative. For the convenience of the reader, these results are stated as a theorem.

Theorem 1.1 *Let $\mathbb{E}[\|Y_1\|^4] < \infty$. Then, it holds under H_0 that*

$$S_{n,d} \xrightarrow{\mathcal{D}} \sum_{\ell=1}^{d} \int_0^1 B_\ell^2(x)dx \qquad (n \to \infty),$$

where $\xrightarrow{\mathcal{D}}$ indicates convergence in distribution. If Δ is not orthogonal to the subspace spanned by the eigenfunctions $\varphi_1, \ldots, \varphi_d$, then it holds under H_A that $S_{n,d} \xrightarrow{P} \infty$ as $n \to \infty$.

While the theorem guarantees in its second part that $S_{n,d}$ will eventually detect a change given that there are sufficiently many observations, it does not contain information on how to locate the change-point, and what the distributional properties of an appropriate estimator are. The main aim of the present paper is therefore to introduce an estimator $\hat{k}_n^*$ for k^* and to derive its limit distribution under different assumptions on the function Δ which determines the type of change. This will be done in Section 2. In Section 3, we evaluate the finite sample behavior via a small simulation study. All proofs are relegated to Section 4.

2 Change-point estimator and its limit distribution

It is assumed throughout this section that the alternative hypothesis H_A holds true. Letting $\mathbf{x}^T$ denote the transpose of a vector $\mathbf{x}$, define $\hat{\boldsymbol{\eta}}_i = (\hat{\eta}_{i,1}, \ldots, \hat{\eta}_{i,d})^T$ and the diagonal matrix $\hat{\Sigma} = \mathrm{diag}(\hat{\lambda}_\ell : \ell = 1, \ldots, d)$. Introducing the quantities

$$\hat{\boldsymbol{\kappa}}_n(k) = \sum_{i=1}^{k} \hat{\boldsymbol{\eta}}_i - \frac{k}{n} \sum_{i=1}^{n} \hat{\boldsymbol{\eta}}_i$$

and the quadratic forms

$$\hat{Q}_n(k) = \frac{1}{n} \hat{\boldsymbol{\kappa}}_n^T(k) \hat{\Sigma}^{-1} \hat{\boldsymbol{\kappa}}_n(k),$$

a suitable estimator for k^* is given by

$$\hat{k}_n^* = \min\left\{ k : \hat{Q}_n(k) = \max_{1 \le j \le n} \hat{Q}_n(j) \right\}. \tag{2.1}$$

With this procedure, we select as change-point the time k that maximizes the random quadratic form $\hat{Q}_n(k)$ which is directly linked to the test statistic $S_{n,d}$ from the previous section via the equality $S_{n,d} = \int_0^1 \hat{Q}_n(\lfloor nx \rfloor)dx$. Because $\hat{Q}_n(k)$ lives on the subspace spanned by the first d estimated eigenfunctions $\hat{\varphi}_1, \ldots, \hat{\varphi}_d$ of the covariance operator $\hat{K}(s,t)$, we need to determine the behavior of $\hat{K}(s,t)$ under H_A. Due to the additional Δ appearing after the change-point k^*, it cannot be expected that $\hat{K}(s,t)$ provides an estimator for $K(s,t)$ anymore. Indeed, the following holds true instead. If we let

$$K_A(s,t) = K(s,t) + \theta(1-\theta)\Delta(t)\Delta(s), \qquad s, t \in \mathcal{T},$$

then $K_A(s,t)$ is symmetric, square integrable and positive–definite, so it admits a representation

$$K_A(s,t) = \sum_{j=1}^{\infty} \gamma_\ell \psi_\ell(s) \psi_\ell(t)$$

with eigenfunctions ψ_ℓ and eigenvalues γ_ℓ obtained from solving the system $\int_{\mathcal{T}} K_A(s,t)\psi_\ell(t)dt = \gamma_\ell \psi_\ell(s)$. The relation between the pairs (γ_ℓ, ψ_ℓ) and $(\hat{\lambda}_\ell, \hat{\varphi}_\ell)$ is established in Proposition 2.1 whose proof is given in Berkes *et al.* (2007).

Proposition 2.1 *Under H_A it holds that, for all $1 \le \ell \le d$,*
(i) $|\hat{\lambda}_\ell - \gamma_\ell| = o_P(1)$ as $n \to \infty$ and
(ii) $\|\hat{\varphi}_\ell - \hat{c}_\ell \psi_\ell\| = o_P(1)$ as $n \to \infty$,
where $\hat{c}_\ell = \operatorname{sign} \int_{\mathcal{T}} \psi_\ell(t) \hat{\varphi}_\ell(t) dt$.

The proposition identifies γ_ℓ and ψ_ℓ (up to a sign) as the stochastic limits of their estimated versions $\hat{\lambda}_\ell$ and $\hat{\varphi}_\ell$. As a consequence, it implies that the limit distribution of $\hat{k}_n^*$ depends on the behavior of the projection of Δ on the subspace spanned by the eigenfunctions $\psi_1, \ldots, \psi_d$. For $1 \le \ell \le d$, denote by

$$\zeta_{i,\ell} = \sqrt{\gamma_\ell}\,\xi_{i,\ell} = \int_{\mathcal{T}} Y_i(t)\psi_\ell(t)dt \qquad \text{and} \qquad \beta_\ell = \sqrt{\gamma_\ell}\,\delta_\ell = \int_{\mathcal{T}} \Delta(t)\psi_\ell(t)dt$$

the principal component scores and set

$$\boldsymbol{\zeta}_i = (\zeta_{i,1}, \ldots, \zeta_{i,d})^T, \quad \boldsymbol{\xi}_i = (\xi_{i,1}, \ldots, \xi_{i,d})^T, \quad \boldsymbol{\delta} = (\delta_1, \ldots, \delta_d)^T.$$

We distinguish two cases

$$\boldsymbol{\delta} \neq \mathbf{0} \quad \text{is constant} \tag{2.2}$$

and

$$\boldsymbol{\delta} = \boldsymbol{\delta}_n \neq \mathbf{0} \quad \text{such that} \quad \|\boldsymbol{\delta}_n\|_2 \to 0 \quad (n \to \infty), \tag{2.3}$$

where $\|\cdot\|_2$ denotes Euclidean norm on $\mathbb{R}^d$. Assumptions (2.2) and (2.3) reflect two common approaches to deriving an asymptotic distribution of change point estimators, see for example, Csörgő and Horváth (1997) and references therein.

We first state the result for the case (2.2).

Theorem 2.1 *Let $\mathbb{E}[\|Y_1\|^4] < \infty$. If $\boldsymbol{\delta} \neq \mathbf{0}$ is constant, then it holds under H_A that*

$$\hat{k}_n^* - k^* \xrightarrow{\mathcal{D}} \min\left\{k : P(k) = \sup_j P(j)\right\} \qquad (n \to \infty),$$

where

$$P(k) = \begin{cases} (1-\theta)\|\boldsymbol{\delta}\|_2^2 k + \boldsymbol{\delta}^T \boldsymbol{S}_k & \text{if } k < 0, \\ & \text{if } k = 0, \\ -\theta\|\boldsymbol{\delta}\|_2^2 k + \boldsymbol{\delta}^T \boldsymbol{S}_k & \text{if } k > 0, \end{cases}$$

with $\boldsymbol{S}_k$ defined by

$$\boldsymbol{S}_k = \sum_{i=1}^{k} \boldsymbol{\xi}_i + \sum_{i=k}^{-1} \boldsymbol{\xi}_i, \qquad -\infty < k < \infty.$$

Here $(\boldsymbol{\xi}_{-i})$ denotes an independent copy of $(\boldsymbol{\xi}_i)$ for all $i \geq 1$ and, as usual, an empty sum is set to equal zero.

Since $\boldsymbol{\delta}$ does not vary with the number of observations, it appears naturally also in the limit variable, which is given as the argument of the maximum of a two-sided sequence of random variables with drift.

A corresponding result holds true for the case (2.3). It is stated next.

Theorem 2.2 *Let $\mathbb{E}[\|Y_1\|^4] < \infty$. If $\boldsymbol{\delta} = \boldsymbol{\delta}_n \neq \mathbf{0}$ is such that*

$$\|\boldsymbol{\delta}_n\|_2 \to 0, \qquad \text{but} \qquad \frac{n\|\boldsymbol{\delta}_n\|_2^2}{\log\log n} \to \infty \qquad (n \to \infty),$$

then it holds under H_A that

$$\|\boldsymbol{\delta}_n\|_2^2 \big(\hat{k}_n^* - k^*\big) \xrightarrow{\mathcal{D}} \min\big\{t : V(t) = \sup_s V(s)\big\} \qquad (n \to \infty),$$

where

$$V(t) = \begin{cases} (1-\theta)t + W(t) & \text{if } t \leq 0, \\ & \text{if } t = 0, \\ -\theta t + W(t) & \text{if } t > 0, \end{cases}$$

with $(W(t) : -\infty < t < \infty)$ denoting a two-sided standard Brownian motion.

Note that the limit processes $P(k)$ and $V(t)$ contain drift terms which attain their maximum at 0, and whose slope on the negative and positive half line is determined by the location θ of the change-point. If $\theta = 1/2$, then the drift parts are symmetric, while the change-point detection becomes significantly harder if θ is close to 0 (or 1). In these cases, the slope of the drift for positive (or negative) arguments is close to zero. In the case of Theorem 2.1, the constant order of magnitude of $\|\boldsymbol{\delta}\|_2$ also plays a role, with larger changes naturally being more easily identifiable. Theorems 2.1 and 2.2 thus provide clear theoretical justification of the empirical properties discussed in Section 3.

It is possible to develop a feel for the size of the function $\Delta = \Delta_n$ which implies the assumptions of Theorem 2.2. If $\|\Delta_n\| \to 0$, then $\|\hat{K} - K\| \to 0$, so by inequalities (4.38) and (4.44) of Bosq (2000), $\|\hat{\varphi}_\ell - c_\ell \varphi_\ell\| \to 0$ and $\hat{\lambda}_\ell \to \lambda_\ell$ in probability. In view of Proposition 2.1, we have that eigenvalues and eigenfunctions under H_0 and H_A coincide in the limit. This means that $\delta_{\ell,n} \approx c_{n,\ell}\lambda_\ell^{-1} \int_{\mathcal{T}} \Delta_n(t)\varphi_\ell(t)dt$ and so $\|\boldsymbol{\delta}_n\|^2 \approx \sum_{\ell=1}^{d} \lambda_\ell^{-2} \left(\int_{\mathcal{T}} \Delta_n(t)\varphi_\ell(t)dt\right)^2$. Thus, by the Cauchy-Schwartz inequality, $\|\Delta_n\| \to 0$ implies $\|\boldsymbol{\delta}_n\| \to 0$. A sufficient condition for

$n\|\boldsymbol{\delta}_n\|^2/(\log\log n) \to \infty$ cannot be stated as easily, but it is roughly $n\|\Delta_n\|^2/(\log\log n) \to \infty$ because by Parseval's inequality, for sufficiently large d, $\int \Delta_n^2(t)dt \approx \sum_{\ell=1}^d \left(\int_{\mathcal{T}} \Delta_n(t)\varphi_\ell(t)dt\right)^2$. These approximate calculations could be formalized, but our goal is to merely indicate that Theorem 2.2 holds if $||\Delta_n||$ tends to zero at the rate slower than $n^{-1/2}$.

Finally, we discuss the consistency of the estimator. Observe that we have assumed in (2.2) and (2.3) that $\boldsymbol{\delta} \neq \mathbf{0}$. This means that there exists $1 \leq \ell \leq d$ such that $\int_{\mathcal{T}} \Delta(t)\psi_\ell(t)dt \neq 0$. If instead the change function Δ is orthogonal to $\psi_1, \ldots, \psi_d$, that is if

$$\int_0^1 \Delta(t)\psi_\ell(t)dt = 0 \qquad \text{for all} \quad \ell = 1, \ldots, d,$$

then $\hat{k}_n^*$ cannot be a consistent estimator of k^*, since the principal components analysis has been performed in an eigenspace with a too small dimension to capture the change. On the other hand, see e.g. Chapter 8 of Ramsay and Silverman (2005), using large d is not practical because it bears the difficulty of interpreting a multitude of principal components. Moreover, since the eigenvalues λ_ℓ are in decreasing order, the impact of a change occurring for moderate-sized and large ℓ might be small and a detection of it therefore less crucial.

3 Finite sample behavior

We carried out simulations to illustrate our theoretical results in finite samples. We simulated change-point processes under conditions of Theorems 2.1 and 2.2 for different sample sizes, and always used $1,000$ replications. For each replication we estimated the location of a change-point k^*. We generated functional observations according to (1.1). Without loss of generality, μ was chosen to be equal to zero. Two different cases of Y_i were considered, namely the trajectories of the standard Brownian motion (BM), and the Brownian bridge (BB). The number d of the principal components was chosen to be equal to 2 and 3 in order to explain at least 75% of variability. The properties of the sampling distributions of the change-point estimator $\hat{k}_n^*$ are now briefly discussed.

To illustrate the simulation results based on Theorem 2.1 we introduced the quantity $\tau_n^* = k_n^*/n$ and the corresponding estimator $\hat{\tau}_n^* = \hat{k}_n^*/n$. We concentrated on $\hat{\tau}_n^* - \tau^*$ rather than on $\hat{k}_n^* - k^*$ to show the effect of the increase in sample size more clearly. Various functions Δ were analyzed: $\Delta = t$, t^2, $\sqrt{t}$, $\exp(t)$, $\sin(t)$, and $\cos(t)$. To assess the accuracy of the estimator, bias, root mean square error (RMSE), and mean absolute error (MAE) of $\hat{\tau}_n^*$ were computed. To conserve space, we do not display the whole set of tables we obtained, but rather display representative results in Table 1, and discuss general findings. From Table 1 we see that by increasing the sample size we attain a smaller bias, RMSE, and MAE. A similar pattern is observed for the increase in the number of principal components. In all cases we considered, the summary statistics indicate that estimation is more accurate if BB was used, even though the same number of principal components explains more variability for BM. This is easy to understand because the BB is a "smaller" process in the sense that $\mathbb{E}[\|\mathrm{BB}\|^2] = 1/6$ and $\mathbb{E}[\|\mathrm{BM}\|^2] = 1/4$, so the same change function Δ is more pronounced if the Y_i are the BB. As expected from the discussion following Theorem 2.2, the closer the change point is to the middle of the sample, the better the estimator is. For τ^* equal to 0.25 and 0.75 an increased bias is observed.

Next we illustrate Theorem 2.2 which deals with nonconstant Δ. We chose $\Delta = \Delta_n$ satisfying conditions of Theorem 2.2 and carried out the change-point estimation. Several different forms of Δ_n were considered, namely $\sin(t)\frac{n^\alpha}{\sqrt{n}}$, $t\frac{n^\alpha}{\sqrt{n}}$, $\sqrt{t}\frac{n^\alpha}{\sqrt{n}}$, $\cos(t)\frac{n^\alpha}{\sqrt{n}}$, $e^t\frac{n^\alpha}{\sqrt{n}}$, where $\alpha \in (0, 0.5)$. To illustrate Theorem 2.2, we concentrated on the distribution of $\|\boldsymbol{\delta}_n\|_2^2\big(\hat{k}_n^* - k^*\big)$. We computed δ_ℓ from $\sqrt{\gamma_\ell}\delta_\ell = \int_{\mathcal{T}} \Delta(t)\psi_\ell(t)dt$, where for $\ell = 1, \ldots, d$

$$\psi_\ell(t) = \sqrt{2}\sin\left(\frac{2\ell+1}{2}\pi t\right), \quad t \in [0,1], \quad \text{and} \quad \gamma_\ell = \frac{4}{[\pi(2\ell+1)]^2}$$

are the eigenfunctions and eigenvalues of the BM and

$$\psi_\ell(t) = \sqrt{2}\sin\left(\ell\pi t\right), \quad t \in [0,1], \quad \text{and} \quad \gamma_\ell = \frac{1}{[\pi\ell]^2}$$

are the corresponding eigenfunctions and eigenvalues of the BB.

As before, we chose k_n^* to be the lower, middle and upper quartile of the sample size. The graphs of the estimated density of $\|\boldsymbol{\delta}_n\|_2^2\big(\hat{k}_n^* - k^*\big)$ are shown in Figures 1 and 2. The densities are close to each other, as Theorem 2.2 implies that they must be close to the limit distribution. In most cases, a convergence with increasing n is also clearly visible. For example, in the top and middle panels of Figure 2, the densities for $n = 600$ and $n = 900$ almost coincide. These properties hold for all choices of $\alpha \in (0, 0.5)$, Figures 1 and 2 show the extreme cases of $\alpha = 0.05$ and $\alpha = 0.45$.

Table 1: Summary statistics for the change-point estimator. The change-point processes were generated by combining BB and t + BB for three different locations of the change-point τ^*. We used $d = 2$ and $d = 3$ (in parenthesis).

τ^*	Average($\hat{\tau}$)	Bias($\hat{\tau}$)	Median($\hat{\tau}$)	RMSE($\hat{\tau}$)	MAE($\hat{\tau}$)
$n = 60$					
0.25	0.27 (0.26)	0.0152 (0.0107)	0.25 (0.25)	0.0336 (0.0252)	0.0158 (0.0108)
0.50	0.50 (0.50)	0.0002 (-0.0003)	0.50 (0.50)	0.0108 (0.0058)	0.0038 (0.0018)
0.75	0.73 (0.74)	-0.0152 (-0.0087)	0.75 (0.75)	0.0356 (0.0205)	0.0157 (0.0088)
$n = 100$					
0.25	0.26 (0.26)	0.0096 (0.0052)	0.25 (0.25)	0.0220 (0.0122)	0.0101 (0.0053)
0.50	0.50 (0.50)	0.0002 (0.0000)	0.50 (0.50)	0.0063 (0.0039)	0.0024 (0.0011)
0.75	0.74 (0.74)	-0.0096 (-0.0052)	0.75 (0.75)	0.0215 (0.0155)	0.0100 (0.0063)
$n = 140$					
0.25	0.26 (0.25)	0.0062 (0.0039)	0.25 (0.25)	0.0141 (0.0096)	0.0064 (0.0040)
0.50	0.50 (0.50)	-0.0001 (-0.0001)	0.50 (0.50)	0.0043 (0.0027)	0.0017 (0.0007)
0.75	0.74 (0.75)	-0.0071 (-0.0039)	0.75 (0.75)	0.0147 (0.0093)	0.0068 (0.0040)
$n = 200$					
0.25	0.25 (0.25)	0.0046 (0.0030)	0.25 (0.25)	0.0107 (0.0070)	0.0050 (0.0031)
0.50	0.50 (0.50)	0.0001 (0.0000)	0.50 (0.50)	0.0033 (0.0016)	0.0013 (0.0005)
0.75	0.75 (0.75)	-0.0050 (-0.0023)	0.75 (0.75)	0.0110 (0.0062)	0.0052 (0.0024)
$n = 300$					
0.25	0.25 (0.25)	0.0030 (0.0018)	0.25 (0.25)	0.0066 (0.0047)	0.0032 (0.0019)
0.50	0.50 (0.50)	0.0000 (0.0001)	0.50 (0.50)	0.0021 (0.0012)	0.0008 (0.0004)
0.75	0.75 (0.75)	-0.0032 (-0.0018)	0.75 (0.75)	0.0079 (0.0048)	0.0034 (0.0019)
$n = 600$					
0.25	0.25 (0.25)	0.0015 (0.0007)	0.25 (0.25)	0.0036 (0.0019)	0.0016 (0.0008)
0.50	0.50 (0.50)	0.0000 (0.0000)	0.50 (0.50)	0.0010 (0.0006)	0.0004 (0.0002)
0.75	0.75 (0.75)	-0.0015 (-0.0009)	0.75 (0.75)	0.0037 (0.0022)	0.0016 (0.0009)

Figure 1: Estimated density of $\|\boldsymbol{\delta}_n\|_2^2\big(\hat{k}_n^* - k^*\big)$ for the process obtained combining BM and $t\frac{n^{0.05}}{\sqrt{n}} + BM$.

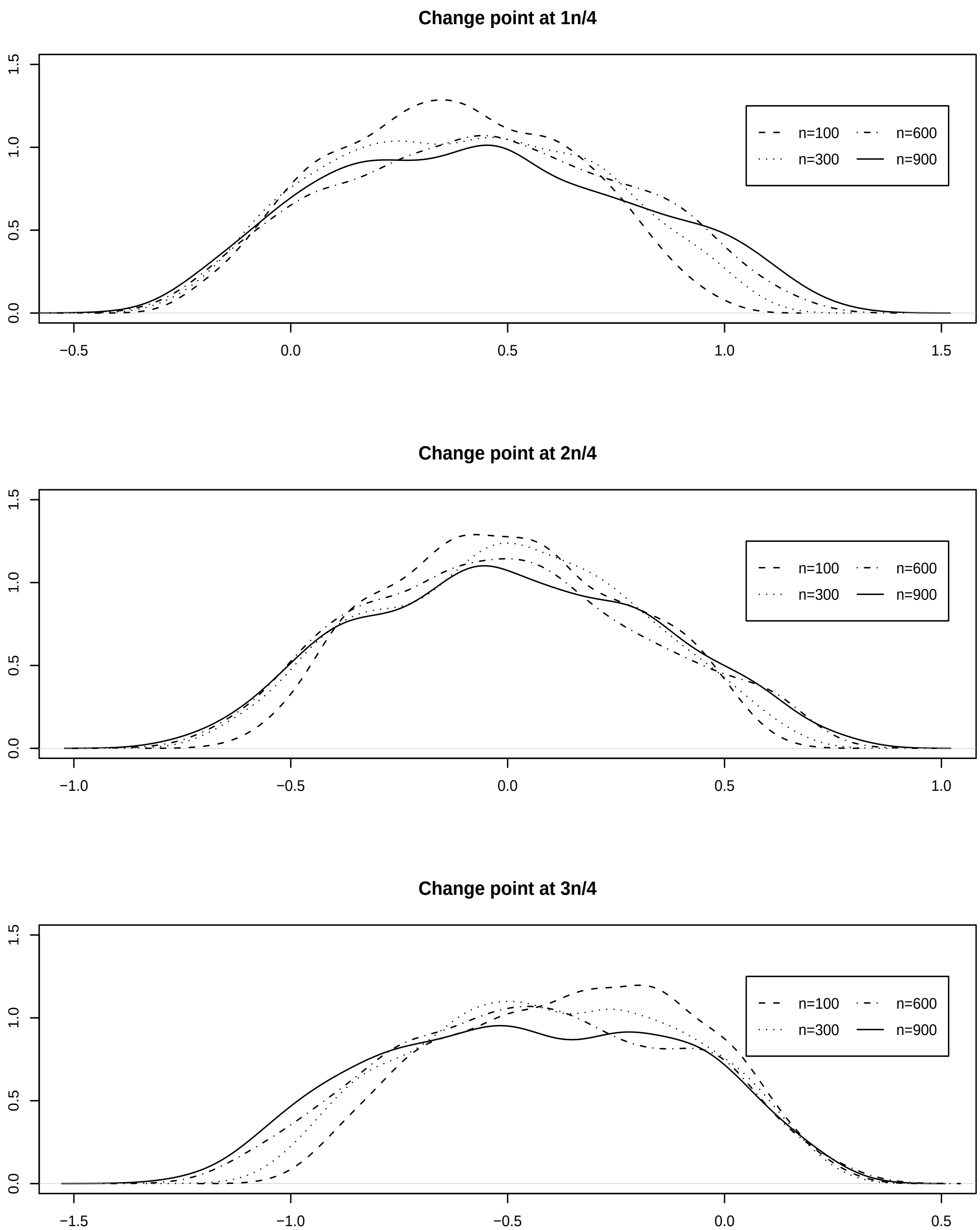
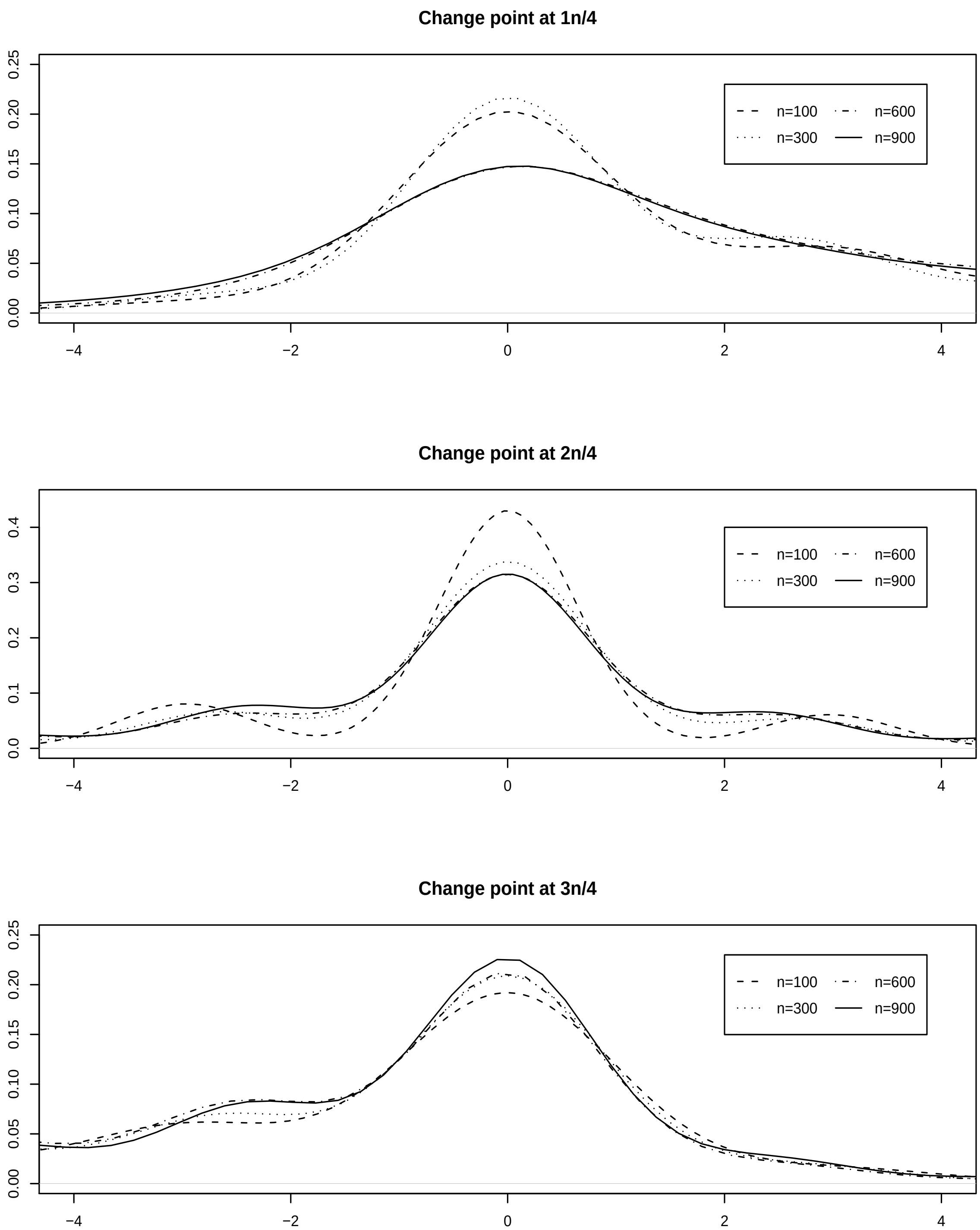


Figure 2: Estimated density of $\|\boldsymbol{\delta}_n\|_2^2\big(\hat{k}_n^* - k^*\big)$ for the process obtained combining BB and $\sin(t)\frac{n^{0.45}}{\sqrt{n}} + BB$.



4 Proofs

The proof section is divided into three parts. In the first subsection, we derive a decomposition that will be used to derive Theorems 2.1 and 2.2, whose proofs will be pursued in Subsections 4.2 and 4.3, respectively.

4.1 Preliminary calculations

Let $\hat{R}_n(k) = \hat{Q}_n(k) - \hat{Q}_n(k^*)$. Since $\hat{R}_n(k)$ and the original $\hat{Q}_n(k)$ differ only by the constant value $\hat{Q}_n(k^*)$, it holds that they attain their maximum for the same value of k. Consequently, we have

$$\hat{k}_n^* = \max\left\{k : \hat{R}_n(k) = \max_{1\le j\le n} \hat{R}_n(j)\right\}.$$

Denote by $\hat{\zeta}_{i,\ell} = \sqrt{\hat{\lambda}_\ell}\hat{\xi}_{i,\ell} = \int_{\mathcal{T}} Y_i(t)\hat{\varphi}_\ell(t)dt$ and $\hat{\beta} = \sqrt{\hat{\lambda}_\ell}\hat{\delta}_\ell = \int_{\mathcal{T}} \Delta(t)\hat{\varphi}_\ell(t)dt$ the counterparts of $\zeta_{i,\ell}$ and β_ℓ which are obtained by replacing the true eigenvalues and eigenfunctions with the estimated versions. Note that the quantities $\hat{\zeta}_{i,\ell}$, $\hat{\xi}_{i,\ell}$, $\hat{\beta}_\ell$ and $\hat{\delta}_\ell$ are unobservable. The proofs to come will fall back on the following decomposition of $\hat{R}_n(k)$. First, we have for $1 \le k < k^*$ that

$$\begin{aligned}
\hat{R}_n(k) =& \frac{1}{n}\sum_{\ell=1}^{d}\left(\sum_{i=1}^{k}\hat{\xi}_{i,\ell} - \frac{k}{n}\sum_{i=1}^{n}\hat{\xi}_{i,\ell} - k\frac{n-k^*}{n}\hat{\delta}_\ell\right)^2 \\
&- \frac{1}{n}\sum_{\ell=1}^{d}\left(\sum_{i=1}^{k^*}\hat{\xi}_{i,\ell} - \frac{k^*}{n}\sum_{i=1}^{n}\hat{\xi}_{i,\ell} - k\frac{n-k^*}{n}\hat{\delta}_\ell\right)^2 \\
=& \frac{1}{n}\sum_{\ell=1}^{d}\left(-\sum_{i=k+1}^{k^*}\hat{\xi}_{i,\ell} - \frac{k-k^*}{n}\sum_{i=1}^{n}\hat{\xi}_{i,\ell} - (k-k^*)\frac{n-k^*}{n}\hat{\delta}_\ell\right) \\
&\times\left(\sum_{i=1}^{k}\hat{\xi}_{i,\ell} + \sum_{i=1}^{k^*}\hat{\xi}_{i,\ell} - \frac{k+k^*}{n}\sum_{i=1}^{n}\hat{\xi}_{i,\ell} - (k+k^*)\frac{n-k^*}{n}\hat{\delta}_\ell\right) \\
=& \frac{1}{n}\sum_{\ell=1}^{d}\left(\hat{E}_{k,\ell}^{(1)} + \hat{D}_{k,\ell}^{(1)}\right)\left(\hat{E}_{k,\ell}^{(2)} + \hat{D}_{k,\ell}^{(2)}\right),
\end{aligned} \tag{4.1}$$

where $\hat{E}_{k,\ell}^{(1)}$ and $\hat{E}_{k,\ell}^{(2)}$ ($\hat{D}_{k,\ell}^{(1)}$ and $\hat{D}_{k,\ell}^{(2)}$) denote the estimated random part (deterministic part) in the first and second bracket of (4.1), respectively. A similar expression can be obtained if $k^* < k \le n$. Here it holds,

$$\hat{R}_n(k) = \frac{1}{n}\sum_{\ell=1}^{d}\left(\sum_{i=1}^{k}\hat{\xi}_{i,\ell} - \frac{k}{n}\sum_{i=1}^{n}\hat{\xi}_{i,\ell} - (n-k)\frac{k^*}{n}\hat{\delta}_\ell\right)^2$$

$$
\begin{aligned}
&-\frac{1}{n}\sum_{\ell=1}^{d}\left(\sum_{i=1}^{k^*}\hat{\xi}_{i,\ell}-\frac{k^*}{n}\sum_{i=1}^{n}\hat{\xi}_{i,\ell}-(n-k^*)\frac{k^*}{n}\hat{\delta}_\ell\right)^2\\
=&\frac{1}{n}\sum_{\ell=1}^{d}\left(-\sum_{k^*+1}^{k}\hat{\xi}_{i,\ell}-\frac{k-k^*}{n}\sum_{i=1}^{n}\hat{\xi}_{i,\ell}+(k-k^*)\frac{k^*}{n}\hat{\delta}_\ell\right)\\
&\times\left(\sum_{i=1}^{k}\hat{\xi}_{i,\ell}+\sum_{i=1}^{k^*}\hat{\xi}_{i,\ell}-\frac{k+k^*}{n}\sum_{i=1}^{n}\hat{\xi}_{i,\ell}-(2n-k-k^*)\frac{k^*}{n}\hat{\delta}_\ell\right)\\
=&\frac{1}{n}\sum_{\ell=1}^{d}\left(\hat{E}_{k,\ell}^{(3)}+\hat{D}_{k,\ell}^{(3)}\right)\left(\hat{E}_{k,\ell}^{(4)}+\hat{D}_{k,\ell}^{(4)}\right),
\end{aligned}
\tag{4.2}
$$

where $\hat{E}_{k,\ell}^{(3)}$, $\hat{E}_{k,\ell}^{(4)}$ and $\hat{D}_{k,\ell}^{(3)}$, $\hat{D}_{k,\ell}^{(4)}$ are the corresponding estimated random and drift parts. Using (4.1) and (4.2), we proceed with the proof of Theorem 2.1 in the next subsection. Since the arguments to be employed are symmetric for time lags before and after the change-point, detailed expositions will only be given for $1 \le k < k^*$.

4.2 Proof of Theorem 2.1

The proof is divided into two parts. At first, we show that the estimator $\hat{k}_n^*$ will be close to k^* by showing that $\hat{R}_n(k)$ will attain its maximum not too far from the change-point. In the second step, we will derive the limit distribution.

Lemma 4.1 *Under the assumptions of Theorem 2.1, it holds that*

$$
\left|\hat{k}_n^*-k^*\right|=\mathcal{O}_P(1)\qquad(n\to\infty).
$$

Proof. To show the assertion of the lemma, we determine the behavior of those k satisfying $1 \le k \le k^*-N$ or $k^*+N \le k \le n$ for some $N \ge 1$. Let $1 \le \ell \le d$. At first, we derive the order of magnitude of the estimated deterministic term in (4.1), that is, of $\frac{1}{n}\hat{D}_{k,\ell}^{(1)}\hat{D}_{k,\ell}^{(2)}$. To this end, note that

$$
\max_{1\le k\le k^*-N}\frac{k+k^*}{n}\left(\frac{n-k^*}{n}\right)^2\to 2\theta(1-\theta)^2\qquad(n\to\infty).
$$

In the next step, we shall replace $\hat{\delta}_\ell$ by δ_ℓ. To do so, observe that $\hat{\beta}_\ell=\sqrt{\hat{\lambda}_\ell}\hat{\delta}_\ell$ and $\beta_\ell=\sqrt{\lambda_\ell}\delta_\ell$ by definition. Moreover, part (ii) of Proposition 2.1 states that $\|\hat{\varphi}_\ell-\hat{c}_\ell\psi_\ell\|\to 0$ in probability. Therefore,

$$
\hat{\beta}_\ell=\int_{\mathcal{T}}\Delta(t)\hat{\varphi}_\ell(t)dt=\hat{c}_\ell\int_{\mathcal{T}}\Delta(t)\psi_\ell(t)dt+o_P(1)=\hat{c}_\ell\beta_\ell+o_P(1)\qquad(n\to\infty),
$$

using that $\Delta(t) \in \mathcal{L}^2(\mathcal{T})$. Consequently, $\hat{\beta}_\ell^2 = \beta_\ell^2 + o_P(1)$. Since the estimated eigenvalues $\hat{\lambda}_\ell$ converge in probability to γ_ℓ (see part (i) of Proposition 2.1), we arrive at

$$\hat{\delta}_\ell^2 = \delta_\ell^2 + o_P(1) \qquad (n \to \infty). \tag{4.3}$$

Combining the above arguments yields

$$\max_{1 \le k \le k^* - N} \frac{1}{n} \hat{D}_{k,\ell}^{(1)} \hat{D}_{k,\ell}^{(2)} = \max_{1 \le k \le k^* - N} (k - k^*) \delta_\ell^2 \frac{k + k^*}{n} \left(\frac{n - k^*}{n} \right)^2 + o_P(1) = -2\theta(1-\theta)^2 N + o_P(1).$$

It is shown in the Appendix that this deterministic part is the dominating term in (4.1). It follows thus that, for all $K > 0$,

$$\lim_{N \to \infty} \limsup_{n \to \infty} P\left(\max_{1 \le k \le k^* - N} \hat{R}_n(k) > -K \right) = 0. \tag{4.4}$$

On the other hand, using (4.2), it can be proved in a similar fashion that

$$\max_{k^* + N \le k \le n} \frac{1}{n} \hat{D}_{k,\ell}^{(3)} \hat{D}_{k,\ell}^{(4)} = -2\theta^2(1-\theta)N + o_P(1),$$

which implies that, for all $K > 0$,

$$\lim_{N \to \infty} \limsup_{n \to \infty} P\left(\max_{k^* + N \le k \le n} \hat{R}_n(k) > -K \right) = 0. \tag{4.5}$$

Equations (4.4) and (4.5) now yield that

$$\lim_{N \to \infty} \limsup_{n \to \infty} P\left(\left\{ \hat{k}_n^* < k^* - N \right\} \cup \left\{ \hat{k}_n^* > k^* + N \right\} \right) = 0,$$

which consequently finishes the proof of the lemma. $\square$

To derive the limit distribution, it suffices to investigate the asymptotic behavior of $\hat{R}_n(k)$ for the range $k^* - N \le k \le k^* + N$ of those time lags close to the change-point. The result is presented as a lemma.

Lemma 4.2 *Under the assumptions of Theorem 2.1, it holds that, for any $N \ge 1$,*

$$\left\{ \hat{R}_n(k + k^*) : -N \le k \le N \right\} \xrightarrow{\mathcal{D}} \left\{ 2\theta(1-\theta)P(k) : -N \le k \le N \right\} \qquad (n \to \infty).$$

Proof. Let $1 \le \ell \le d$. Using (4.3), it is easy to see that, for any fixed $N \ge 1$ and as $n \to \infty$,

$$\begin{aligned} &\max_{k^* - N \le k \le k^*} \left| \frac{1}{n} \hat{D}_{k,\ell}^{(1)} \hat{D}_{k,\ell}^{(2)} - \theta(1-\theta)^2 \delta_\ell^2 (k - k^*) \right| + o_P(1) \\ &\qquad = \delta_\ell^2 N \max_{k^* - N \le k \le k^*} \left| \frac{k + k^*}{n} \left(\frac{n - k^*}{n} \right)^2 - 2\theta(1-\theta)^2 \right| + o_P(1) \end{aligned}$$

$$= o_P(1).$$

By a similar argument,

$$\max_{k^* \le k \le k^*+N} \left| \frac{1}{n} \hat{D}^{(3)}_{k,\ell} \hat{D}^{(4)}_{k,\ell} + 2\theta^2(1-\theta)\delta_\ell^2 (k-k^*) \right| = o_P(1) \qquad (n \to \infty).$$

In the following, we are dealing with the estimated random parts. The functional central limit theorem implies that, for all $x \in [0,1]$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nx \rfloor} \boldsymbol{\zeta}_i \xrightarrow{d} \boldsymbol{\Gamma}(x) \qquad (n \to \infty),$$

where $\xrightarrow{d}$ indicates weak convergence in the Skorohod space $D^d[0,1]$ and $(\boldsymbol{\Gamma}(x) : x \in [0,1])$ is an $\mathbb{R}^d$-valued, zero mean stochastic process with covariance matrix Σ. Then,

$$\begin{aligned}
&\sup_{x\in(0,1)} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^{\lfloor nx \rfloor} \zeta_{i,l} - \sum_{i=1}^{\lfloor nx \rfloor} \hat{\zeta}_{i,\ell} \right| \\
&= \sup_{x\in(0,1)} \left| \int_{\mathcal{T}} \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nx \rfloor} Y_i(t) \left[\hat{c}_\ell \psi_\ell(t) - \hat{\varphi}_\ell(t) \right] \right| \\
&\le \sup_{x\in(0,1)} \left(\int_{\mathcal{T}} \left[\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nx \rfloor} Y_i(t) \right]^2 dt \right)^{1/2} \left(\int_{\mathcal{T}} \left[\hat{c}_\ell \psi_\ell(t) - \hat{\varphi}_\ell(t) \right] dt \right)^{1/2} \\
&= o_P(1) \qquad (4.6)
\end{aligned}$$

by an application of Proposition 2.1. The same statement holds true also if $\xi_{i,l}$ and $\hat{\xi}_{i,l}$ are used in place of $\zeta_{i,\ell}$ and $\hat{\zeta}_{i,\ell}$.

Equations (4.3) and (4.6) imply now that

$$\begin{aligned}
&\max_{k^*-N \le k \le k^*} \left| \left(\frac{k-k^*}{n} \sum_{i=1}^{n} \hat{\xi}_{i,\ell} \right) \left(\frac{k+k^*}{n} \frac{n-k^*}{n} \delta_\ell \right) \right| \\
&= \max_{k^*-N \le k \le k^*} \left| \left(\frac{k-k^*}{n} \sum_{i=1}^{n} \xi_{i,\ell} \right) \left(\frac{k+k^*}{n} \frac{n-k^*}{n} \delta_\ell \right) \right| + o_P(1) \\
&= \mathcal{O}_P(1) \frac{N}{n} \left| \sum_{i=1}^{n} \int_{\mathcal{T}} Y_i(t) \varphi_\ell(t) dt \right| + o_P(1) \\
&= o_P(1).
\end{aligned}$$

Hence,

$$\max_{k^*-k\le k\le k^*}\left|\frac{1}{n}\hat{E}^{(1)}_{k,\ell}\hat{D}^{(2)}_{k,\ell}+2\theta(1-\theta)\delta_\ell\sum_{i=k+1}^{k^*}\xi_{i,\ell}\right|=o_P(1)$$

as $n\to\infty$ for any $N\ge 1$ which follows from (4.3) and (4.6) as well. Similar arguments apply also to $\frac{1}{n}\hat{E}^{(3)}_{k,\ell}\hat{D}^{(4)}_{k,\ell}$ for which $k^*\le k\le k^*+N$ holds. In view of the definition of the limit process $P(k)$ in Theorem 2.1, it suffices to verify that the remaining terms in (4.1) and (4.2) do not contribute asymptotically. To this end, write

$$\begin{aligned}
&\max_{k^*-N\le k\le k^*}\frac{1}{n}\left|\hat{E}^{(1)}_{k,\ell}\hat{E}^{(2)}_{k,\ell}\right|\\
&=\max_{k^*-N\le k\le k^*}\frac{1}{n}\left|\left(\sum_{i=k+1}^{k^*}\hat{\xi}_{i,\ell}+\frac{k-k^*}{n}\sum_{i=1}^{n}\hat{\xi}_{i,\ell}\right)\left(\sum_{i=1}^{k}\hat{\xi}_{i,\ell}+\sum_{i=1}^{k^*}\hat{\xi}_{i,\ell}-\frac{k+k^*}{n}\sum_{i=1}^{n}\hat{\xi}_{i,\ell}\right)\right|\\
&\le\max_{k^*-N\le k\le k^*}\left|\sum_{i=k+1}^{k^*}\hat{\xi}_{i,\ell}+\frac{k-k^*}{n}\sum_{i=1}^{n}\hat{\xi}_{i,\ell}\right|\max_{k^*-N\le k\le k^*}\frac{1}{n}\left|\sum_{i=1}^{k}\hat{\xi}_{i,\ell}+\sum_{i=1}^{k^*}\hat{\xi}_{i,\ell}-\frac{k+k^*}{n}\sum_{i=1}^{n}\hat{\xi}_{i,\ell}\right|\\
&=o_P(1).
\end{aligned}$$

Here, the first maximum is $\mathcal{O}_P(1)$, since the first sum $\sum_{i=k+1}^{k^*}\hat{\xi}_{i,\ell}$ contains at most N terms, while the second sum is $o_P(1)$ because of (4.6). Another application of (4.6) gives that the second maximum is $o_P(1)$. Moreover,

$$\begin{aligned}
&\max_{k^*-N\le k\le k^*}\frac{1}{n}\left|\hat{D}^{(1)}_{k,\ell}\hat{E}^{(2)}_{k,\ell}\right|\\
&=\max_{k^*-N\le k\le k^*}\frac{1}{n}\left|(k-k^*)\frac{n-k^*}{n}\hat{\delta}_\ell\left(\sum_{i=1}^{k}\hat{\xi}_{i,\ell}+\sum_{i=1}^{k^*}\hat{\xi}_{i,\ell}-\frac{k+k^*}{n}\sum_{i=1}^{n}\hat{\xi}_{i,\ell}\right)\right|\\
&\le\max_{k^*-N\le k\le k^*}\left|(k-k^*)\frac{n-k^*}{n}\hat{\delta}_\ell\right|\max_{k^*-N\le k\le k^*}\frac{1}{n}\left|\sum_{i=1}^{k}\hat{\xi}_{i,\ell}+\sum_{i=1}^{k^*}\hat{\xi}_{i,\ell}-\frac{k+k^*}{n}\sum_{i=1}^{n}\hat{\xi}_{i,\ell}\right|\\
&=o_P(1).
\end{aligned}$$

The same arguments apply also to the remaining terms in (4.2) and the proof of the lemma is therefore complete. □

Proof of Theorem 2.1. The assertion follows immediately from Lemmas 4.1 and 4.2. □

4.3 Proof of Theorem 2.2

We follow the proof steps developed in the previous subsection.

Lemma 4.3 *Under the assumptions of Theorem 2.1, it holds that*

$$\|\boldsymbol{\delta}_n\|^2 \left|\hat{k}_n^* - k^*\right| = \mathcal{O}_P(1) \qquad (n \to \infty).$$

Proof. At first, we derive the order of magnitude of $\frac{1}{n}\hat{D}_{k,\ell}^{(1)}\hat{D}_{k,\ell}^{(2)}$ in (4.1). Let $N \geq 1$ and define $N_\delta = N\|\boldsymbol{\delta}_n\|_2^{-2}$. Recognizing that $n^{-1}N_\delta \to 0$, since by assumption $n\|\boldsymbol{\delta}_n\|_2^2 \to \infty$, it follows that

$$\max_{1\leq k\leq k^*-N_\delta} \frac{k+k^*}{n}\left(\frac{n-k^*}{n}\right)^2 = \frac{2k^*-N_\delta}{n}\left(\frac{n-k^*}{n}\right)^2 \to 2\theta(1-\theta)^2 \qquad (n\to\infty).$$

Consequently, (4.3) yields

$$\begin{aligned}
\max_{1\leq k\leq k^*-N_\delta} \frac{1}{n}\sum_{\ell=1}^d \hat{D}_{k,\ell}^{(1)}\hat{D}_{k,\ell}^{(2)} &= \max_{1\leq k\leq k^*-N_\delta}(k-k^*)\frac{k+k^*}{n}\left(\frac{n-k^*}{n}\right)^2\sum_{\ell=1}^d \hat{\delta}_\ell^2 \\
&= \max_{1\leq k\leq k^*-N_\delta}(k-k^*)\frac{k+k^*}{n}\left(\frac{n-k^*}{n}\right)^2\sum_{\ell=1}^d \delta_\ell^2 + o_P(1) \\
&= -2\theta(1-\theta)^2 N + o_P(1).
\end{aligned}$$

It is shown in Appendix B that, under the assumptions of Theorem 2.2, this deterministic part is the dominating contributor in (4.1). It follows thus that, for all $K > 0$,

$$\lim_{N\to\infty}\limsup_{n\to\infty} P\left(\max_{1\leq k\leq k^*-N_\delta}\hat{R}_n(k) > -K\right) = 0 \tag{4.7}$$

Moreover, utilizing the decomposition in display (4.2), it can be proved similarly that

$$\max_{k^*+N_\delta\leq k\leq n}\frac{1}{n}\sum_{\ell=1}^d \hat{D}_{k,\ell}^{(3)}\hat{D}_{k,\ell}^{(4)} = -2\theta^2(1-\theta)N + o_P(1),$$

which implies that, for all $K > 0$,

$$\lim_{N\to\infty}\limsup_{n\to\infty} P\left(\max_{k^*+N_\delta\leq k\leq n}\hat{R}_n(k) > -K\right) = 0. \tag{4.8}$$

Equations (4.7) and (4.8) now yield that

$$\lim_{N\to\infty}\limsup_{n\to\infty} P\left(\left\{\hat{k}_n^* < k^* - N_\delta\right\}\cup\left\{\hat{k}_n^* > k^* + N_\delta\right\}\right) = 0,$$

which, noticing the definition of N_δ, completes the proof of the lemma. □

Lemma 4.4 *Under the assumptions of Theorem 2.2, it holds that, for any* $N \geq 1$,

$$\left\{\hat{R}_n\big(k^* + \left\lfloor t\|\boldsymbol{\delta}_n\|_2^{-2}\right\rfloor\big) : t\in[-N,N]\right\} \xrightarrow{d} \{2\theta(1-\theta)V(t) : t\in[-N,N]\} \qquad (n\to\infty),$$

where $\xrightarrow{d}$ *indicates weak convergence in the Skorohod space* $D[-N,N]$.

Proof. Denote by k the integer part of $t\|\boldsymbol{\delta}_n\|_2^2$. Then, as $n\to\infty$,

$$\sup_{t\in[-N,0]}\left|\frac{1}{n}\sum_{\ell=1}^{d}\hat{D}^{(1)}_{k^*+k,\ell}\hat{D}^{(2)}_{k^*+k,\ell}+2\theta(1-\theta)^2t\right|=\mathcal{O}_P(1)\sup_{t\in[-N,0]}\left(t-\|\boldsymbol{\delta}_n\|_2^2\left\lfloor t\|\boldsymbol{\delta}_n\|_2^{-2}\|\right\rfloor\right)=o_P(1).$$

Similarly,

$$\sup_{t\in[0,N]}\left|\frac{1}{n}\sum_{\ell=1}^{d}D^{(3)}_{k^*+k,\ell}D^{(4)}_{k^*+k,\ell}-2\theta^2(1-\theta)t\right|=o_P(1)\qquad(n\to\infty).$$

Note next that, after an application of (4.6) and the law of the iterated logarithm

$$\sup_{t\in[-N,0]}\frac{|t|}{n\|\boldsymbol{\delta}_n\|_2^2}\left|\sum_{\ell=1}^{d}\boldsymbol{\delta}_\ell\sum_{i=1}^{n}\boldsymbol{\xi}_{i,\ell}\right|=\mathcal{O}_P(1)\frac{1}{n\|\boldsymbol{\delta}_n\|_2}\left|\sum_{\ell=1}^{d}\sum_{i=1}^{n}\xi_{i,\ell}\right|=\mathcal{O}_P(1)\sqrt{\frac{\log\log n}{n\|\boldsymbol{\delta}_n\|_2^2}}=o_P(1)$$

by assumption on $\boldsymbol{\delta}_n$. It follows from the weak convergence of partial sum processes that there exist independent standard Brownian motions $(W_\ell(t):t\geq 0)$, $1\leq\ell\leq d$, such that

$$\|\boldsymbol{\delta}_n\|_2\sum_{i=k^*+k+1}^{k^*}\xi_{i,\ell}\overset{\mathcal{D}}{=}\|\boldsymbol{\delta}_n\|_2\sum_{i=1}^{-k}\xi_{i,\ell}\overset{\mathcal{D}[-N,0]}{\longrightarrow}W_\ell(-t),$$

where k is the integer part of $t\|\boldsymbol{\delta}_n\|_2^{-2}$. Checking the finite-dimensional distributions, it follows that the process

$$\left(\frac{1}{\|\boldsymbol{\delta}\|_2}\sum_{\ell=1}^{d}\delta_\ell W_\ell(t):t\geq 0\right)$$

is a standard Brownian motion. Therefore, there exists a standard Brownian motion $(W^{(1)}(t):t\geq 0)$ such that

$$\begin{aligned}
&\sup_{t\in[-N,0]}\left|\frac{1}{n}\sum_{\ell=1}^{d}\hat{E}^{(1)}_{k,\ell}\hat{D}^{(2)}_{k,\ell}-2\theta(1-\theta)W^{(1)}(t)\right|\\
&=\sup_{t\in[-N,0]}\left|\frac{2k^*+k}{n}\frac{n-k^*}{n}\sum_{\ell=1}^{d}\delta_\ell\left(\sum_{i=k^*+k+1}^{k^*}\xi_{i,\ell}+\frac{t}{n\|\boldsymbol{\delta}_n\|_2^2}\sum_{i=1}^{n}\xi_{i,\ell}\right)-2\theta(1-\theta)W^{(1)}(t)\right|+o_P(1)\\
&=\mathcal{O}(1)\sup_{t\in[-N,0]}\left|\sum_{\ell=1}^{d}\delta_\ell\sum_{i=k^*+k+1}^{k^*}\xi_{i,\ell}-W^{(1)}(t)\right|+o_P(1)\\
&=o_P(1).
\end{aligned}$$

A similar string of arguments yields that there is a standard Brownian motion $(W^{(2)}(t):t\geq 0)$ such that

$$\sup_{t\in[0,N]}\left|\frac{1}{n}\sum_{\ell=1}^{d}\hat{E}^{(3)}_{k,\ell}\hat{D}^{(4)}_{k,\ell}-2\theta(1-\theta)W^{(2)}(t)\right|=o_P(1).$$

It remains to verify that the remaining parts in displays (4.1) and (4.2) do not contribute to the limit distribution. So, consider first

$$
\begin{aligned}
\max_{k^*-N_\delta \le k \le k^*} \frac{1}{n} \big| \hat{D}^{(1)}_{k,\ell} \hat{E}^{(2)}_{k,\ell} \big| \\
&= \max_{k^*-N_\delta \le k \le k^*} \left| \frac{k^*-k}{n} \frac{n-k^*}{n} \delta_\ell \left(\sum_{i=1}^{k} \xi_{i,\ell} + \sum_{i=1}^{k^*} - \frac{k+k^*}{n} \sum_{i=1}^{n} \xi_{i,\ell} \right) \right| + o_P(1) \\
&= \mathcal{O}(1) \max_{k^*-N_\delta \le k \le k^*} \left| \frac{1}{n \|\boldsymbol{\delta}_n\|_2} \left(\sum_{i=1}^{k} \xi_{i,\ell} + \sum_{i=1}^{k^*} - \frac{k+k^*}{n} \sum_{i=1}^{n} \xi_{i,\ell} \right) \right| + o_P(1) \\
&= o_P(1),
\end{aligned}
$$

since, for example,

$$
\max_{k^*-N_\delta \le k \le k^*} \frac{1}{n \|\boldsymbol{\delta}_n\|_2} \left| \sum_{i=1}^{k} \xi_{i,\ell} \right| = \mathcal{O}_P(1) \sqrt{\frac{\log \log n}{n \|\boldsymbol{\delta}_n\|_2^2}} + o_P(1) = o_P(1)
$$

by (4.6), the law of the iterated logarithm and assumption on $\boldsymbol{\delta}_n$. All other terms can be handled in the same way. Next, note that by (4.6) and the law of the iterated logarithm, it holds that

$$
\max_{k^*-N_\delta \le k \le k^*} \sum_{i=k+1}^{k^*} \boldsymbol{\xi}_{i,\ell} = \mathcal{O}_P\left(\sqrt{N_\delta}\right) \qquad \text{and} \qquad \max_{k^*-N_\delta \le k \le k^*} \frac{k^*-k}{n} \left| \sum_{i=1}^{n} \boldsymbol{\xi}_{i,\ell} \right| = o_P(1),
$$

respectively. Hence,

$$
\begin{aligned}
\max_{k^*-N_\delta \le k \le k^*} \frac{1}{n} \big| \hat{E}^{(1)}_{k,\ell} \hat{E}^{(2)}_{k,\ell} \big| \\
&= \max_{k^*-N_\delta \le k \le k^*} \frac{1}{n} \left| \left(\sum_{i=k+1}^{k^*} \hat{\xi}_{i,\ell} + \frac{k-k^*}{n} \sum_{i=1}^{n} \hat{\xi}_{i,\ell} \right) \left(\sum_{i=1}^{k} \hat{\xi}_{i,\ell} + \sum_{i=1}^{k^*} - \frac{k+k^*}{n} \sum_{i=1}^{n} \hat{\xi}_{i,\ell} \right) \right| \\
&\mathcal{O}_P(1) \max_{k^*-N_\delta \le k \le k^*} \frac{\sqrt{N_\delta}}{n} \left| \sum_{i=1}^{k} \hat{\xi}_{i,\ell} + \sum_{i=1}^{k^*} - \frac{k+k^*}{n} \sum_{i=1}^{n} \hat{\xi}_{i,\ell} \right| \\
&= o_P(1),
\end{aligned}
$$

since, by (4.6) and the law of the iterated logarithm,

$$
\max_{k^*-N_\delta \le k \le k^*} \frac{\sqrt{N}}{n \|\hat{\delta}_n\|_2} \left| \sum_{i=1}^{k} \hat{\xi}_{i,\ell} \right| = \mathcal{O}_P(1) \sqrt{\frac{\log \log n}{n \|\delta_n\|_2^2}} + o_P(1) = o_P(1).
$$

Similar for the other two terms and also for the terms coming from (4.2). The proof is complete. □

Appendices

A Verification of equation (4.4)

Lemma A.1 *Under the assumptions of Theorem 2.1 it holds that, for all* $1 \le \ell \le d$ *and* $\varepsilon > 0$*,*

$$\lim_{N\to\infty} \limsup_{n\to\infty} P\left(\max_{1\le k\le k^*-N} \frac{|\hat{E}^{(1)}_{k,\ell}\hat{E}^{(2)}_{k,\ell}|}{|\hat{D}^{(1)}_{k,\ell}\hat{D}^{(2)}_{k,\ell}|} \ge \varepsilon \right) = 0.$$

Proof. Let $1 \le \ell \le d$ and $1 \le k \le k^* - N$ for some $N \ge 1$. From the definition in (4.1) and the argument leading to display (4.3) it follows that the absolute value of the estimated deterministic term $|\hat{D}^{(1)}_{k,\ell}\hat{D}^{(2)}_{k,\ell}|$ has precise stochastic order $n(k^* - k)$. Hence,

$$\max_{1\le k\le k^*-N} \frac{|\hat{E}^{(1)}_{k,\ell}\hat{E}^{(2)}_{k,\ell}|}{|\hat{D}^{(1)}_{k,\ell}\hat{D}^{(2)}_{k,\ell}|} = \mathcal{O}(1) \max_{1\le k\le k^*-N} \frac{|\hat{E}^{(1)}_{k,\ell}|}{k^*-k} \max_{1\le k\le k^*-N} \frac{|\hat{E}^{(2)}_{k,\ell}|}{n} = \mathcal{O}_P(1) M_1(N,n) M_2(N,n).$$

We start by examining $M_1(N,n)$. The law of the iterated logarithm in combination with (4.6) imply that

$$\max_{1\le k\le k^*-N} \frac{1}{k^*-k} \left| \sum_{i=k+1}^{k^*} \hat{\xi}_{i,\ell} \right| = \mathcal{O}_P\left(\max_{1\le k\le k^*-N} \frac{1}{(k^*-k)^{1-\alpha}} \right) = \mathcal{O}_P\left(\frac{1}{N^{1-\alpha}} \right),$$

for any $1/2 < \alpha < 1$. Moreover, on account of (4.6),

$$\max_{1\le k\le k^*-N} \frac{1}{n} \left| \sum_{i=1}^{n} \hat{\xi}_{i,\ell} \right| = \frac{1}{n} \left| \sum_{i=1}^{n} \hat{\xi}_{i,\ell} \right| = o_P(1) \qquad (n\to\infty).$$

Three further applications of (4.6) to $M_2(N,n)$ yield that

$$\lim_{N\to\infty} \limsup_{n\to\infty} P\big(M_1(N,n) M_2(N,n) \ge \varepsilon \big) = 0$$

and the lemma is proved. □

Lemma A.2 *Under the assumptions of Theorem 2.1 it holds that, for all* $1 \le \ell \le d$ *and* $\varepsilon > 0$*,*

$$\lim_{N\to\infty} \limsup_{n\to\infty} P\left(\max_{1\le k\le k^*-N} \frac{|\hat{E}^{(1)}_{k,\ell}\hat{D}^{(2)}_{k,\ell}|}{|\hat{D}^{(1)}_{k,\ell}\hat{D}^{(2)}_{k,\ell}|} \ge \varepsilon \right) = 0.$$

Proof. Write

$$\max_{1\le k\le k^*-N}\frac{|\hat{E}^{(1)}_{k,\ell}\hat{D}^{(2)}_{k,\ell}|}{|\hat{D}^{(1)}_{k,\ell}\hat{D}^{(2)}_{k,\ell}|}=\mathcal{O}_P(1)\max_{1\le k\le k^*-N}\frac{|\hat{E}^{(1)}_{k,\ell}|}{k^*-k}\max_{1\le k\le k^*-N}\frac{|\hat{D}^{(2)}_{k,\ell}|}{n}=\mathcal{O}_(1)M_1(N,n)M_3(N,n),$$

where $M_1(N,n)$ has already been dealt with in Lemma A.1. Noticing that

$$M_3(N,n)=\max_{1\le k\le k^*-N}\frac{k+k^*}{n}\frac{n-k^*}{n}\delta_\ell+o_P(1)=\mathcal{O}_P(1)$$

hence yields the assertion. □

Lemma A.3 *Under the assumptions of Theorem 2.1 it holds that, for all $1\le\ell\le d$ and $\varepsilon>0$,*

$$\lim_{N\to\infty}\limsup_{n\to\infty}P\left(\max_{1\le k\le k^*-N}\frac{|\hat{D}^{(1)}_{k,\ell}\hat{E}^{(2)}_{k,\ell}|}{|\hat{D}^{(1)}_{k,\ell}\hat{D}^{(2)}_{k,\ell}|}\ge\varepsilon\right)=0.$$

Proof. In an analogous fashion, we obtain

$$\max_{1\le k\le k^*-N}\frac{|\hat{D}^{(1)}_{k,\ell}\hat{E}^{(2)}_{k,\ell}|}{|\hat{D}^{(1)}_{k,\ell}\hat{D}^{(2)}_{k,\ell}|}=\mathcal{O}(1)\max_{1\le k\le k^*-N}\frac{|\hat{D}^{(1)}_{k,\ell}|}{k^*-k}\max_{1\le k\le k^*-N}\frac{|\hat{E}^{(2)}_{k,\ell}|}{n}=\mathcal{O}_P(1)M_4(N,n)M_2(N,n)$$

with $M_2(N,n)$ from Lemma A.1. Therefore

$$M_4(N,n)=\max_{1\le k\le k^*-N}\frac{n-k^*}{n}\frac{k^*-k}{k^*-k}\delta_\ell+o_P(1)=\mathcal{O}_P(1)$$

gives the result. □

Similar calculations can be be performed for the terms appearing in display (4.2). Details are omitted.

B Verification of equation (4.7)

Lemma B.1 *Under the assumptions of Theorem 2.2 it holds that, for all $\varepsilon>0$,*

$$\lim_{N\to\infty}\limsup_{n\to\infty}P\left(\max_{1\le k\le k^*-N_\delta}\frac{\sum_{\ell=1}^d|\hat{E}^{(1)}_{k,\ell}\hat{E}^{(2)}_{k,\ell}|}{\sum_{\ell=1}^d|\hat{D}^{(1)}_{k,\ell}\hat{D}^{(2)}_{k,\ell}|}\ge\varepsilon\right)=0.$$

Proof. Observe that, uniformly in k,

$$\sum_{\ell=1}^d|\hat{D}^{(1)}_{k,\ell}\hat{D}^{(2)}_{k,\ell}|\sim_P n(k^*-k)\|\boldsymbol{\delta}_n\|_2^2.$$

Therefore, for any $1\le\ell\le d$,

$$\max_{1\le k\le k^*-N_\delta}\frac{|\hat{E}^{(1)}_{k,\ell}\hat{E}^{(2)}_{k,\ell}|}{\sum_{\ell=1}^d|\hat{D}^{(1)}_{k,\ell}\hat{D}^{(2)}_{k,\ell}|}=\mathcal{O}_P(1)\max_{1\le k\le k^*-N_\delta}\frac{|\hat{E}^{(1)}_{k,\ell}|}{(k^*-k)\|\boldsymbol{\delta}_n\|_2}\max_{1\le k\le k^*-N_\delta}\frac{|\hat{E}^{(2)}_{k,\ell}|}{n\|\boldsymbol{\delta}_n\|_2}$$

$$= \mathcal{O}_P(1) M_1^\delta(N,n) M_2^\delta(N,n).$$

We first study the asymptotics of $M_1^\delta(N,n)$. To this end note that

$$\max_{1\le k\le k^*-N_\delta} \frac{|\sum_{i=k+1}^{k^*} \xi_{i,\ell}|}{(k^*-k)\|\boldsymbol{\delta}_n\|_2} \overset{\mathcal{D}}{=} \max_{1\le k\le k^*-N_\delta} \frac{|\sum_{i=1}^{k^*-k} \xi_{i,\ell}|}{(k^*-k)\|\boldsymbol{\delta}_n\|_2}$$

$$= \mathcal{O}_P(1) \max_{1\le k\le k^*-N_\delta} \frac{1}{(k^*-k)\|\boldsymbol{\delta}_n\|_2} = \mathcal{O}_P\left(\frac{1}{\sqrt{N}}\right).$$

Furthermore, from the law of the iterated logarithm,

$$\max_{1\le k\le k^*-N_\delta} \frac{|\sum_{i=1}^{n} \xi_{i,\ell}|}{n\|\boldsymbol{\delta}_n\|_2} = o_P(1) \qquad (n\to\infty).$$

Since the same arguments apply also to the term $M_2^\delta(N,n)$, it follows from (4.6) that

$$\lim_{N\to\infty} \limsup_{n\to\infty} P\left(M_1^\delta(N,n) M_2^\delta(N,n) \ge \varepsilon\right) = 0.$$

This proves the assertion. □

Lemma B.2 *Under the assumptions of Theorem 2.2 it holds that, for all $\varepsilon > 0$,*

$$\lim_{N\to\infty} \limsup_{n\to\infty} P\left(\max_{1\le k\le k^*-N_\delta} \frac{\sum_{\ell=1}^d |\hat{E}_{k,\ell}^{(1)} \hat{D}_{k,\ell}^{(2)}|}{\sum_{\ell=1}^d |\hat{D}_{k,\ell}^{(1)} \hat{D}_{k,\ell}^{(2)}|} \ge \varepsilon\right) = 0.$$

Proof. Along the lines of the previous proof, we may write

$$\max_{1\le k\le k^*-N_\delta} \frac{|\hat{E}_{k,\ell}^{(1)} \hat{D}_{k,\ell}^{(2)}|}{\sum_{\ell=1}^d |\hat{D}_{k,\ell}^{(1)} \hat{D}_{k,\ell}^{(2)}|} = \mathcal{O}_P(1) \max_{1\le k\le k^*-N_\delta} \frac{|\hat{E}_{k,\ell}^{(1)}|}{(k^*-k)\|\boldsymbol{\delta}_n\|_2} \max_{1\le k\le k^*-N_\delta} \frac{|\hat{D}_{k,\ell}^{(2)}|}{n\|\boldsymbol{\delta}_n\|_2}$$

$$= \mathcal{O}_P(1) M_1^\delta(N,n) M_3^\delta(N,n),$$

where

$$M_3^\delta(N,n) = \max_{1\le k\le k^*-N_\delta} \frac{k+k^*}{n} \frac{n-k^*}{n} \frac{\delta_\ell}{\|\boldsymbol{\delta}_n\|_2} + o_P(1) = \mathcal{O}_P(1).$$

Since $M_1^\delta(N,n)$ has already been estimated in Lemma B.1, the proof is complete. □

Lemma B.3 *Under the assumptions of Theorem 2.2 it holds that, for all $\varepsilon > 0$,*

$$\lim_{N\to\infty} \limsup_{n\to\infty} P\left(\max_{1\le k\le k^*-N_\delta} \frac{\sum_{\ell=1}^d |\hat{D}_{k,\ell}^{(1)} \hat{E}_{k,\ell}^{(2)}|}{\sum_{\ell=1}^d |\hat{D}_{k,\ell}^{(1)} \hat{D}_{k,\ell}^{(2)}|} \ge \varepsilon\right) = 0.$$

Proof. Write

$$\max_{1\le k\le k^*-N_\delta} \frac{|\hat{D}_{k,\ell}^{(1)} \hat{E}_{k,\ell}^{(2)}|}{\sum_{\ell=1}^d |\hat{D}_{k,\ell}^{(1)} \hat{D}_{k,\ell}^{(2)}|} = \mathcal{O}_P(1) M_4^\delta(N,n) M_2^\delta(N,n)$$

with

$$M_4^\delta(N,n) = \max_{1\le k\le k^*-N_\delta} \frac{k-k^*}{n} \frac{n-k^*}{n} \frac{\delta_\ell}{\|\boldsymbol{\delta}_n\|_2} + o_P(1) = \mathcal{O}_P(1)$$

and the lemma is proved. □

Again, the same arguments give the corresponding results for the terms in (4.2).

References

Antoch, J., Husková, M. and Prášková, Z. (1997). Effect of dependence on statistics for determination of change. *Journal of Statistical Planning and Inference*, **60,** 291–310.

Antoniadis, A. and Sapatinas, T. (2003). Wavelet methods for continuous time prediction using Hilbert–valued autoregressive processes. *Journal of Multivariate Analysis*, **87,** 133–158.

Berkes, I., Gabrys, R., Horváth, L. and kokoszka, P. (2007). Detecting changes in the mean of functional observations. Technical Report. Utah State University.

Bosq, D. (2000). *Linear Processes in Function Spaces.* Springer, New York.

Chiou, J-M., Müller, H-G. and Wang, J-L. (2004). Functional response models. *Statistica Sinica*, **14,** 675–693.

Cobb, G. W. (1978). The problem of the Nile: conditional solution to a change-point problem. *Biometrika*, **65,** 243–251.

Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis.* Wiley, New York.

Dauxois, J., Pousse, A. and Romain, Y. (1982). Asymptotic theory for principal component analysis of a vector random function. *Journal of Multivariate Analysis*, **12,** 136–154.

Davis, R. D., Huang, D. and Yao, Y-C. (1995). Testing for a change in the parameter values and order of an autoregressive model. *The Annals of Statistics*, **23,** 282–304.

Fernández de Castro, B., Guillas, S. and Gonzáles Manteiga, W. (2005). Functional samples and bootstrap for predicting sulfur dioxide levels. *Technometrics*, **47,** 212–222.

Garcia, R. and Ghysels, E. (1998). Structural change and asset pricing in emerging markets. *Journal of International Money and Finance,*, **17,** 455–473.

Glendinning, R. H. and Fleet, S. L. (2007). Classifying functional time series. *Signal Processing*, **87,** 79–100.

Horváth, L., Kokoszka, P. S. and Steinebach, J. (1999). Testing for changes in multivariate dependent observations with applications to temperature changes. *Journal of Multivariate Analysis*, **68,** 96–119.

Inclán, C. and Tiao, G. C. (1994). Use of cummulative sums of squares for retrospective detection of change of variance. *Journal of American Statistical Association*, **89,** 913–923.

Kokoszka, P., Maslova, I., Sojka, J. and Zhu, L. (2008a). Testing for lack of dependence in functional linear model. *Canadian Journal of Statistics*, **00,** 000–000; Forthcoming.

Kokoszka, P., Maslova, I., Sojka, J. and Zhu, L. (2008b). Effect of substorms on the magnetic field variability at mid- and low-latitudes. Technical Report. Utah State University.

Kokoszka, P. S. and Leipus, R. (2000). Change-point estimation in ARCH models. *Bernoulli*, **6,** 513–539.

Laukaitis, A. and Račkauskas, A. (2005). Functional data analysis for clients segmentation tasks. *European Journal of Operational Research*, **163,** 210–216.

Müller, H-G. and Stadtmüller, U. (2005). Generalized functional linear models. *The Annals of Statistics*, **33,** 774–805.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis.* Springer.

Yao, F., Müller, H-G. and Wang., J-L. (2005). Functional linear regression analysis for longitudinal data. *The Annals of Statistics*, **33,** 2873–2903.