

Online supplementary material to “Detecting and dating structural breaks in functional data without dimension reduction”^{*}

Alexander Aue[†]

Gregory Rice[‡]

Ozan Sönmez[†]

August 28, 2017

Abstract

This online supplement contains proofs of the theorems of the main paper Aue et al. (2017+). The proofs are given in Section A. Several helpful auxiliary results are collected in Section B. Section C contains implementation details that did not fit into the main paper. The outcomes of additional simulation experiments are reported in Section D. Further evidence from the temperature data example is provided in Section E. The results of a second data analysis to cumulative intra-day returns of Microsoft stock are reported in Section F.

Keywords: Change-point analysis; Functional data; Functional principal components; Functional time series; Intra-day financial data; Structural breaks

MSC 2010: Primary: 62G99, 62H99, Secondary: 62M10, 91B84

A Proofs

A.1 Proof of Theorems 2.1 and 2.2

The proof of Theorem 2.1 is based on the following result of Jirak (2013), which is provided here for ease of reference.

Theorem A.1 (Theorem 1.2 of Jirak, 2013). *Let*

$$S_n(x, t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nx \rfloor} \varepsilon_i(t). \quad (\text{A.1})$$

^{*}This research was partially supported by NSF grants DMS 1209226, DMS 1305858 and DMS 1407530.

[†]Department of Statistics, University of California, Davis, CA 95616, USA, email: [aaue, osonmez]@ucdavis.edu

[‡]Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON, Canada, email: grice@uwaterloo.ca

Then, under Assumption 2.1, there exists a sequence of Gaussian processes, $(\Gamma_n(x, t) : n \in \mathbb{N}, x, t \in [0, 1])$, such that $E[\Gamma_n(x, t)] = 0$, $E[\Gamma_n(x, t)\Gamma_n(x', t')] = \min\{x, x'\}C_\varepsilon(t, t')$ and

$$\sup_{0 \leq x \leq 1} \int (S_n(x, t) - \Gamma_n(x, t))^2 dt = o_P(1).$$

Proof of Theorem 2.1. Under H_0 it can be assumed without loss of generality that $\mu = 0$ in model (2.1). In this case, $S_{n,k}^0(t) = S_n(\lfloor k/n \rfloor, t) - k/n S_n(1, t)$, with $S_n(x, t)$ defined in (A.1). Moreover, it can be shown using moment arguments that

$$T_n = \sup_{0 \leq x \leq 1} \|S_n(x, \cdot) - x S_n(1, \cdot)\|^2 + o_P(1). \quad (\text{A.2})$$

Letting $\Gamma_n^0(x, t) = \Gamma_n(x, t) - x\Gamma_n(1, t)$, it follows that

$$\begin{aligned} & \sup_{0 \leq x \leq 1} \|S_n(x, \cdot) - x S_n(1, \cdot)\|^2 \\ &= \sup_{0 \leq x \leq 1} \|\Gamma_n^0(x, \cdot)\|^2 + \sup_{0 \leq x \leq 1} \|S_n(x, \cdot) - x S_n(1, \cdot)\|^2 - \sup_{0 \leq x \leq 1} \|\Gamma_n^0(x, \cdot)\|^2. \end{aligned} \quad (\text{A.3})$$

The triangle inequality in combination with Theorem A.1 imply that

$$\sup_{0 \leq x \leq 1} \|S_n(x, \cdot) - x S_n(1, \cdot)\|^2 - \sup_{0 \leq x \leq 1} \|\Gamma_n^0(x, \cdot)\|^2 = o_P(1),$$

and this along with (A.2) and (A.3) yield

$$T_n = \sup_{0 \leq x \leq 1} \|\Gamma_n^0(x, \cdot)\|^2 + o_P(1). \quad (\text{A.4})$$

Direct calculations using the definition of $\Gamma_n(x, t)$ show that $E[\Gamma_n^0(x, t)\Gamma_n^0(x, t')] = (\min\{x, x'\} - xx')C_\varepsilon(t, t')$, and hence, for each n , the Gaussian process $\Gamma_n^0(x, t)$ has the same distribution as

$$\sum_{\ell=1}^{\infty} \sqrt{\lambda_\ell} B_\ell(x) \phi_\ell(t),$$

where $(\lambda_\ell : \ell \in \mathbb{N})$ and $(\phi_\ell : \ell \in \mathbb{N})$ are defined in (2.5), and $(B_\ell : \ell \in \mathbb{N})$ are standard Brownian bridges on $[0, 1]$. It follows that, for all n ,

$$\sup_{0 \leq x \leq 1} \|\Gamma_n^0(x, \cdot)\|^2 \stackrel{\mathcal{D}}{=} \sup_{0 \leq x \leq 1} \sum_{\ell=1}^{\infty} \lambda_\ell B_\ell^2(x),$$

which, in light of (A.4), implies the theorem. $\square$

Proof of Theorem 2.2. Letting

$$S_{n,k,\varepsilon}^0 = \frac{1}{\sqrt{n}} \left(\sum_{i=1}^k \varepsilon_i - \frac{k}{n} \sum_{i=1}^n \varepsilon_i \right),$$

one has that, under H_A ,

$$T_n = \max_{1 \leq k \leq n} \|S_{n,k}^0\|^2 \geq \|S_{n,k^*}^0\|^2 = \left\| S_{n,k^*,\varepsilon}^0 - \frac{k^*}{\sqrt{n}} \frac{n - k^*}{n} \delta \right\|^2.$$

Moreover, by the triangle inequality,

$$\left\| S_{n,k^*,\varepsilon}^0 - \frac{k^*}{\sqrt{n}} \frac{n - k^*}{n} \delta \right\| \geq \left\| \frac{k^*}{\sqrt{n}} \frac{n - k^*}{n} \delta \right\| - \|S_{n,k^*,\varepsilon}^0\|.$$

Theorem A.1 implies that $\|S_{n,k^*,\varepsilon}^0\| = O_P(1)$, and, since $k^* = \lfloor n\theta \rfloor$ for some $\theta \in (0, 1)$,

$$\left\| \frac{k^*}{\sqrt{n}} \frac{n - k^*}{n} \delta \right\| \geq c_{1,n} \sqrt{n} \theta (1 - \theta) \|\delta\| \rightarrow \infty, \quad (n \rightarrow \infty),$$

as $c_{1,n}$ may be taken so that $c_{1,n} \rightarrow 1$, which implies the result. $\square$

A.2 Proof of Theorems 2.3 and 2.4

The proofs of Theorems 2.3 and 2.4 have a similar starting point as the proofs of the main results in Aue et al. (2009), with the primary differences being that in this paper weakly dependent functional time series are considered and that the statistic is defined without a dimension reduction step. The latter difference is a simplification on one hand, since the rate of approximation of empirical eigenfunctions need not be accounted for, and a complication on the other, since throughout infinite dimensional objects are to be studied.

The proofs will be carried out via a sequence of lemmas that rely on the following observations. Notice that, after squaring $\|S_{n,k}^0\|$ and then centering the resulting quantity with $\|S_{n,k^*}^0\|^2$, it follows that

$$\hat{k}_n^* = \min \left\{ k : R_{n,k} = \max_{1 \leq k' \leq n} R_{n,k'} \right\},$$

where

$$R_{n,k} = \|S_{n,k}^0\|^2 - \|S_{n,k^*}^0\|^2.$$

A standard calculation common in the analysis of structural breaks, using the definitions of $S_{n,k}^0$ and the model equation (2.1), yields that, for $1 \leq k < k^*$,

$$R_{n,k} = \frac{1}{n} \left\langle E_k^{(1)} + D_k^{(1)}, E_k^{(2)} + D_k^{(2)} \right\rangle, \quad (\text{A.5})$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product in $L^2[0, 1]$,

$$D_k^{(1)} = -(k - k^*) \frac{n - k^*}{n} \delta \quad \text{and} \quad D_k^{(2)} = -(k + k^*) \frac{n - k^*}{n} \delta$$

are deterministic drift functions and

$$E_k^{(1)} = - \sum_{i=k+1}^{k^*} \varepsilon_i - \frac{k - k^*}{n} \sum_{i=1}^n \varepsilon_i, \quad \text{and} \quad E_k^{(2)} = \sum_{i=1}^k \varepsilon_i + \sum_{i=1}^{k^*} \varepsilon_i - \frac{k + k^*}{n} \sum_{i=1}^n \varepsilon_i$$

are random functions. In a similar way, when $k^* < k \leq n$, one obtains the decomposition

$$R_{n,k} = \frac{1}{n} \left\langle E_k^{(3)} + D_k^{(3)}, E_k^{(4)} + D_k^{(4)} \right\rangle, \quad (\text{A.6})$$

into respective drift and random functions

$$D_k^{(3)} = -(k - k^*) \frac{k^*}{n} \delta \quad \text{and} \quad D_k^{(4)} = -(2n - k - k^*) \frac{k^*}{n} \delta,$$

$$E_k^{(3)} = - \sum_{i=k^*+1}^k \varepsilon_i - \frac{k - k^*}{n} \sum_{i=1}^n \varepsilon_i \quad \text{and} \quad E_k^{(4)} = \sum_{i=1}^k \varepsilon_i + \sum_{i=1}^{k^*} \varepsilon_i - \frac{k + k^*}{n} \sum_{i=1}^n \varepsilon_i.$$

In the proofs below, the focus is on the asymptotics for $R_{n,k}$ when $1 \leq k < k^*$ using (A.5), noting that similar arguments may be applied when $k^* < k \leq n$ using (A.6). Throughout the proofs, for $j \in \mathbb{N}_0$, c_j denote positive absolute constants.

A.2.1 Proof of Theorem 2.3

Lemma A.1. *If the assumptions of Theorem 2.3 are satisfied, $|\hat{k}_n^* - k^*|$ is bounded in probability.*

Proof. Let $N \geq 1$. Because $k^* = \lfloor \theta n \rfloor$, it follows that

$$\max_{1 \leq k \leq k^* - N} \frac{k + k^*}{n} \left(\frac{n - k^*}{n} \right)^2 = \frac{2k^* - N}{n} \left(\frac{n - k^*}{n} \right)^2 \rightarrow 2\theta(1 - \theta)^2,$$

as $n \rightarrow \infty$. Using the preceding together with the definitions of $D_k^{(1)}$ and $D_k^{(2)}$ implies that

$$\begin{aligned} \max_{1 \leq k \leq k^* - N} \frac{1}{n} \left\langle D_k^{(1)}, D_k^{(2)} \right\rangle &= \max_{1 \leq k \leq k^* - N} (k - k^*) \frac{k + k^*}{n} \left(\frac{n - k^*}{n} \right)^2 \|\delta\|^2 \\ &= -2N\theta(1 - \theta)^2 \|\delta\|^2 + o(1). \end{aligned}$$

Therefore,

$$\lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} \max_{1 \leq k \leq k^* - N} \frac{1}{n} \left\langle D_k^{(1)}, D_k^{(2)} \right\rangle = -\infty. \quad (\text{A.7})$$

In the following, it will be shown that the deterministic component whose asymptotics is established in (A.7) is the dominant term on the right hand side of (A.5). In this direction, it is first established that, for all $x > 0$,

$$\lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} P \left(\max_{1 \leq k \leq k^* - N} \frac{\left\langle E_k^{(1)}, E_k^{(2)} \right\rangle}{\left\langle D_k^{(1)}, D_k^{(2)} \right\rangle} > x \right) = 0. \quad (\text{A.8})$$

Note that, for all $n \geq 1$ and $1 \leq k < k^*$,

$$\left\langle D_k^{(1)}, D_k^{(2)} \right\rangle = n(k^* - k) \frac{k + k^*}{n} \frac{n - k^*}{n} \|\delta\|^2 \geq c_0 n(k^* - k)$$

for some constant $c_0 > 0$. Therefore, an application of the Cauchy–Schwarz inequality yields that

$$\frac{\langle E_k^{(1)}, E_k^{(2)} \rangle}{\langle D_k^{(1)}, D_k^{(2)} \rangle} \leq c_1 \frac{\langle E_k^{(1)}, E_k^{(2)} \rangle}{n(k^* - k)} \leq c_1 \frac{\|E_k^{(1)}\| \|E_k^{(2)}\|}{n(k^* - k)}.$$

Taking the maximum on right- and left-hand side leads to

$$\max_{1 \leq k \leq k^* - N} \frac{\langle E_k^{(1)}, E_k^{(2)} \rangle}{\langle D_k^{(1)}, D_k^{(2)} \rangle} \leq c_2 \max_{1 \leq k \leq k^* - N} \frac{\|E_k^{(1)}\|}{k^* - k} \max_{1 \leq k \leq k^* - N} \frac{\|E_k^{(2)}\|}{n}$$

The definition of $E_k^{(1)}$ and the triangle inequality give

$$\max_{1 \leq k \leq k^* - N} \frac{\|E_k^{(1)}\|}{k^* - k} \leq \max_{1 \leq k \leq k^* - N} \frac{1}{k^* - k} \left\| \sum_{i=k+1}^{k^*} \varepsilon_i \right\| + \frac{1}{n} \left\| \sum_{i=1}^n \varepsilon_i \right\|. \quad (\text{A.9})$$

Now, for all $\alpha \in (1/2, 1)$,

$$\max_{1 \leq k \leq k^* - N} \frac{1}{k^* - k} \left\| \sum_{i=k+1}^{k^*} \varepsilon_i \right\| \stackrel{\mathcal{D}}{=} \max_{N \leq k \leq k^*} \frac{1}{k} \left\| \sum_{i=1}^k \varepsilon_i \right\| \leq \frac{1}{N^{1-\alpha}} \sup_{k \geq 1} \frac{1}{k^\alpha} \left\| \sum_{i=1}^k \varepsilon_i \right\|.$$

By Lemma B.1, $\sup_{k \geq 1} (1/k^\alpha) \left\| \sum_{i=1}^k \varepsilon_i \right\| = O_P(1)$, and thus, applying the Ergodic Theorem in $L^2[0, 1]$ to handle the second term on the right-hand side of (A.9), it follows that

$$\lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} P \left(\max_{1 \leq k \leq k^* - N} \frac{\|E_k^{(1)}\|}{k^* - k} > x \right) = 0.$$

One may apply similar arguments to show that

$$\lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} P \left(\max_{1 \leq k \leq k^* - N} \frac{\|E_k^{(2)}\|}{n} > x \right) = 0,$$

which yields (A.8). Similar arguments also imply that

$$\lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} P \left(\max_{1 \leq k \leq k^* - N} \frac{\langle E_k^{(1)}, D_k^{(2)} \rangle}{\langle D_k^{(1)}, D_k^{(2)} \rangle} > x \right) = 0, \quad (\text{A.10})$$

and

$$\lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} P \left(\max_{1 \leq k \leq k^* - N} \frac{\langle D_k^{(1)}, E_k^{(2)} \rangle}{\langle D_k^{(1)}, D_k^{(2)} \rangle} > x \right) = 0. \quad (\text{A.11})$$

Combining (A.8), (A.10) and (A.11) with (A.5) yields that, for all $M < 0$,

$$\lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} P \left(\max_{1 \leq k \leq k^* - N} R_{n,k} > M \right) = 0. \quad (\text{A.12})$$

It may be proven analogously using $D_k^{(3)}, D_k^{(4)}, E_k^{(3)}, E_k^{(4)}$ and (A.6) that

$$\lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} P \left(\max_{k^* + N \leq k \leq n} R_{n,k} > M \right) = 0. \quad (\text{A.13})$$

Since $\hat{k}_n^*$ is equivalent with the maximum argument of $R_{n,k}$, and $R_{n,k^*} = 0$, for all $M < 0$,

$$P\left(|\hat{k}_n^* - k^*| > N\right) \leq P\left(\max_{1 \leq k \leq k^* - N} R_{n,k} > M\right) + P\left(\max_{k^* + N \leq k \leq n} R_{n,k} > M\right).$$

Therefore, (A.12) and (A.13) imply that

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

which is equivalent with the statement of the lemma. $\square$

Lemma A.2. *If the assumptions of Theorem 2.3 are satisfied, then*

$$(R_{n,k^*+k} : k \in [-N, N]) \xrightarrow{\mathcal{D}} (2\theta(1-\theta)P(k) : k \in [-N, N])$$

for any $N \geq 1$ as $n \rightarrow \infty$.

Proof. According to the definitions of $D_k^{(1)}$, $D_k^{(2)}$ and k^* ,

$$\begin{aligned} & \max_{k^* - N \leq k \leq k^*} \left| \frac{1}{n} \langle D_k^{(1)}, D_k^{(2)} \rangle - 2\theta(1-\theta)^2 \|\delta\|^2 (k - k^*) \right| \\ &= \max_{k^* - N \leq k \leq k^*} \left| \|\delta\|^2 (k - k^*) \left[\frac{k + k^*}{n} \left(\frac{n - k^*}{n} \right)^2 - 2\theta(1-\theta)^2 \right] \right| \\ &\leq \|\delta\|^2 N \max_{k^* - N \leq k \leq k^*} \left| \frac{k + k^*}{n} \left(\frac{n - k^*}{n} \right)^2 - 2\theta(1-\theta)^2 \right| \\ &= o(1), \end{aligned} \tag{A.14}$$

as $n \rightarrow \infty$. One may obtain similarly that

$$\max_{k^* \leq k \leq k^* + N} \left| \frac{1}{n} \langle D_k^{(3)}, D_k^{(4)} \rangle - [-2\theta^2(1-\theta)\|\delta\|^2(k - k^*)] \right| = o(1). \tag{A.15}$$

This establishes the limits of the drift terms. To analyze the stochastic part, note first that

$$\frac{1}{n} \langle E_k^{(1)}, D_k^{(2)} \rangle = \frac{n - k^*}{n} \frac{k + k^*}{n} \sum_{i=k+1}^{k^*} \langle \varepsilon_i, \delta \rangle + \frac{n - k^*}{n} \frac{k + k^*}{n} \frac{k - k^*}{n} \sum_{i=1}^n \langle \varepsilon_i, \delta \rangle. \tag{A.16}$$

The first term on the right-hand side of (A.16) can be estimated as

$$\begin{aligned} & \max_{k^* - N \leq k \leq k^*} \left| \frac{n - k^*}{n} \frac{k + k^*}{n} \sum_{i=k+1}^{k^*} \langle \varepsilon_i, \delta \rangle - 2(1-\theta)\theta \sum_{i=k+1}^{k^*} \langle \varepsilon_i, \delta \rangle \right| \\ &\leq \max_{k^* - N \leq k \leq k^*} \left| \frac{n - k^*}{n} \frac{k + k^*}{n} - 2(1-\theta)\theta \right| \max_{k^* - N \leq k \leq k^*} \left| \sum_{i=k+1}^{k^*} \langle \varepsilon_i, \delta \rangle \right| \\ &= o_P(1). \end{aligned} \tag{A.17}$$

For the second term on the right-hand side of (A.16), Fubini's theorem shows that $\mathbb{E}[\langle \varepsilon_i, \delta \rangle] = 0$, and hence by the Ergodic Theorem, $(1/n) \sum_{i=1}^n \langle \varepsilon_i, \delta \rangle \rightarrow 0$ as $n \rightarrow \infty$ with probability one. Consequently,

$$\max_{k^*-N \leq k \leq k^*} \left| \frac{n - k^*}{n} \frac{k + k^*}{n} (k - k^*) \frac{1}{n} \sum_{i=1}^n \langle \varepsilon_i, \delta \rangle \right| = o_P(1). \quad (\text{A.18})$$

Taken together, (A.16)–(A.18) imply that

$$\max_{k^*-N \leq k \leq k^*} \left| \frac{1}{n} \langle E_k^{(1)}, D_k^{(2)} \rangle - 2(1 - \theta)\theta \sum_{i=k+1}^{k^*} \langle \varepsilon_i, \delta \rangle \right| = o_P(1). \quad (\text{A.19})$$

Similar arguments lead to

$$\max_{k^* \leq k \leq k^*+N} \left| \frac{1}{n} \langle E_k^{(1)}, D_k^{(2)} \rangle - 2(1 - \theta)\theta \sum_{i=k^*+1}^k \langle \varepsilon_i, \delta \rangle \right| = o_P(1). \quad (\text{A.20})$$

It remains to be show that the other four terms on the concluding lines of (A.5) and (A.6) do not contribute asymptotically. To this end, by the Cauchy–Schwarz inequality,

$$\max_{k^*-N \leq k \leq k^*} \left| \frac{1}{n} \langle E_k^{(1)}, E_k^{(2)} \rangle \right| \leq \max_{k^*-N \leq k \leq k^*} \frac{1}{n} \|E_k^{(1)}\| \|E_k^{(2)}\| \quad (\text{A.21})$$

$$\leq \max_{k^*-N \leq k \leq k^*} \|E_k^{(1)}\| \max_{k^*-N \leq k \leq k^*} \frac{1}{n} \|E_k^{(2)}\|. \quad (\text{A.22})$$

The triangle inequality yields that

$$\max_{k^*-N \leq k \leq k^*} \|E_k^{(1)}\| \leq \max_{k^*-N \leq k \leq k^*} \left\| \sum_{i=k+1}^{k^*} \varepsilon_i \right\| + \frac{N}{n} \left\| \sum_{i=1}^n \varepsilon_i \right\| = O_P(1) + o_P(1). \quad (\text{A.23})$$

The last equality follows since the first term contains at most N terms, and the second term is subject to the Ergodic Theorem in Hilbert spaces. Furthermore, again by the triangle inequality,

$$\begin{aligned} & \max_{k^*-N \leq k \leq k^*} \frac{1}{n} \|E_k^{(2)}\| \\ & \leq \max_{k^*-N \leq k \leq k^*} \frac{1}{n} \left\| \sum_{i=1}^k \varepsilon_i \right\| + \frac{1}{n} \left\| \sum_{i=1}^{k^*} \varepsilon_i \right\| + \max_{k^*-N \leq k \leq k^*} \frac{k + k^*}{n^2} \left\| \sum_{i=1}^k \varepsilon_i \right\| \\ & = o_P(1), \end{aligned} \quad (\text{A.24})$$

since the first term on the right-hand side is $o_P(1)$ by Lemma B.1, and the second and third terms are $o_P(1)$ by the Ergodic Theorem. Equations (A.21)–(A.24) imply that

$$\max_{k^*-N \leq k \leq k^*} \left| \frac{1}{n} \langle E_k^{(1)}, E_k^{(2)} \rangle \right| = o_P(1).$$

Additionally, in a similar fashion to (A.21),

$$\max_{k^*-N \leq k \leq k^*} \left| \frac{1}{n} \langle D_k^{(1)}, E_k^{(2)} \rangle \right| \leq \max_{k^*-N \leq k \leq k^*} \|D_k^{(1)}\| \max_{k^*-N \leq k \leq k^*} \frac{1}{n} \|E_k^{(2)}\| \quad (\text{A.25})$$

$$\begin{aligned}
&\leq \|\delta\|^2 N \max_{k^*-N \leq k \leq k^*} \frac{n-k^*}{n} \max_{k^*-N \leq k \leq k^*} \frac{1}{n} \|E_k^{(2)}\| \\
&= o_P(1),
\end{aligned}$$

according to (A.24). Equations (A.14), (A.19), (A.21), and (A.25) imply the lemma when $-N \leq k - k^* \leq 0$, and (A.15), (A.20) plus similar arguments applied to the remaining two terms in (A.6) imply the result when $0 \leq k - k^* \leq N$. $\square$

Proof of Theorem 2.3. Lemma A.1 implies that it suffices to consider the weak convergence of $\hat{k}_n^* - k^*$ on a bounded subset of the integers, and Lemma A.2 along with the Continuous Mapping Theorem gives this the weak convergence to the limit on all bounded subsets and therefore proves the theorem. $\square$

A.2.2 Proof of Theorem 2.4

The proof of Theorem 2.4 is carried out analogously to Theorem 2.3, with two lemmas establishing that the sequence of random variables of interest is bounded in probability, and also converges in distribution to the limit on every bounded set.

Lemma A.3. *If the assumptions of Theorem 2.4 are satisfied, $\|\delta_n\|^2 |\hat{k}_n^* - k^*|$ is bounded in probability.*

Proof. For $N \geq 1$ define $N_\delta = \|\delta_n\|^{-2} N$. Since $\|\delta_n\|^2 n \rightarrow \infty$, $N_\delta/n \rightarrow 0$, and hence

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

as $n \rightarrow \infty$. It follows that

$$\begin{aligned}
\max_{1 \leq k \leq k^* - N_\delta} \frac{1}{n} \langle D_k^{(1)}, D_k^{(2)} \rangle &= \max_{1 \leq k \leq k^* - N_\delta} \|\delta_n\|^2 (k - k^*) \left(\frac{k + k^*}{n} \right) \left(\frac{n - k^*}{n} \right)^2 \\
&= \|\delta_n\|^2 (-N_\delta) \left(\frac{2k^* - N_\delta}{n} \right) \left(\frac{n - k^*}{n} \right)^2 \\
&\rightarrow -2N\theta(1 - \theta)^2
\end{aligned}$$

as $n \rightarrow \infty$. Using similar arguments as those used to establish (A.8), (A.10), and (A.11), one can show that this term is the asymptotically dominant term in equation (A.5). Thus, for all $M < 0$,

$$\lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} P \left(\max_{1 \leq k \leq k^* - N_\delta} R_n(k) > M \right) = 0.$$

Moreover,

$$\lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} P \left(\max_{k^* + N_\delta \leq k \leq n} R_n(k) > M \right) = 0$$

by applying the same reasoning to the terms in (A.6). The preceding two equations imply the lemma. $\square$

Lemma A.4. *If the assumptions of Theorem 2.4 are satisfied, then*

$$(R_{n,k^*+k(x)}: x \in [-N, N]) \xrightarrow{\mathcal{D}[-N, N]} (2\theta(1-\theta)V(x): x \in [-N, N]),$$

for any $N \geq 1$ as $n \rightarrow \infty$, where $k(x) = \lfloor \|\delta_n\|^{-2}x \rfloor$.

Proof. According to the definitions of $D_k^{(1)}$ and $D_k^{(2)}$,

$$\begin{aligned} & \sup_{x \in [-N, 0]} \left| \frac{1}{n} \left\langle D_{k^*+k(x)}^{(1)}, D_{k^*+k(x)}^{(2)} \right\rangle - (-2\theta(1-\theta)^2x) \right| \\ &= \sup_{x \in [-N, 0]} \left| k(x) \frac{2k^* + k(x)}{n} \left(\frac{n - k^*}{n} \right)^2 \|\delta_n\|^2 - (-2\theta(1-\theta)^2x) \right| \\ &= o(1), \end{aligned}$$

using that $k^* = \lfloor \theta n \rfloor$ and that $|k(x)\|\delta_n\|^2 - x| = o(1)$ uniformly in $x \in [-N, 0]$. It follows similarly that

$$\sup_{x \in [0, N]} \left| \frac{1}{n} \left\langle D_{k^*+k(x)}^{(3)}, D_{k^*+k(x)}^{(4)} \right\rangle - (-2\theta^2(1-\theta)x) \right| = o(1),$$

which establishes the limit of the drift terms in $R_n(k^* + k(x))$ as those of $2\theta(1-\theta)V(x)$. An application of Lemma B.2 yields that there exist two independent sequences of two-parameter Gaussian processes $(\Gamma_n^{(1)}(\cdot, \cdot): n \in \mathbb{N})$ and $(\Gamma_n^{(2)}(\cdot, \cdot): n \in \mathbb{N})$ satisfying

$$\sup_{x \in [-N, 0]} \int \left(\|\delta_n\| \sum_{i=k^*+k(x)+1}^{k^*} \varepsilon_i(t) - \Gamma_n^{(1)}(-x, t) \right)^2 dt = o_P(1) \quad (\text{A.26})$$

and

$$\sup_{x \in [0, N]} \int \left(\|\delta_n\| \sum_{i=k^*}^{k^*+k(x)+1} \varepsilon_i(t) - \Gamma_n^{(2)}(x, t) \right)^2 dt = o_P(1), \quad (\text{A.27})$$

such that $\mathbb{E}[\Gamma_n^{(j)}(x, t)] = 0$, and $\text{Cov}(\Gamma_n^{(j)}(x, t), \Gamma_n^{(j)}(x', t')) = \min(x, x')C_\varepsilon(t, t')$, for all $n \in \mathbb{N}$, $x, x' \in [0, N]$, $t, t' \in [0, 1]$ and $j = 1, 2$. In the following it is shown that

$$\sup_{x \in [-N, 0]} \left| \frac{1}{n} \left\langle E_{k^*+k(x)}^{(1)}, D_{k^*+k(x)}^{(2)} \right\rangle - 2\theta(1-\theta) \int \Gamma_n^{(1)}(-x, t) \frac{\delta_n(t)}{\|\delta_n\|} dt \right| = o_P(1). \quad (\text{A.28})$$

The definitions of $E_k^{(1)}$ and $D_k^{(2)}$ give that

$$\begin{aligned} & \sup_{x \in [-N, 0]} \left| \frac{1}{n} \left\langle E_{k^*+k(x)}^{(1)}, D_{k^*+k(x)}^{(2)} \right\rangle - 2\theta(1-\theta) \int \Gamma_n^{(1)}(-x, t) \frac{\delta_n(t)}{\|\delta_n\|} dt \right| \\ &= \sup_{x \in [-N, 0]} \left| \frac{(2k^* + k(x))(n - k^*)}{n^2} \int \|\delta_n\| \sum_{i=k^*+k(x)+1}^{k^*} \varepsilon_i(t) \frac{\delta_n(t)}{\|\delta_n\|} dt \right. \\ & \quad \left. - 2\theta(1-\theta) \int \Gamma_n^{(1)}(-x, t) \frac{\delta_n(t)}{\|\delta_n\|} dt + \frac{k(x)(2k^* + k(x))(n - k^*)}{n^3} \int \sum_{i=1}^n \varepsilon_i(t) \delta_n(t) dt \right|. \end{aligned} \quad (\text{A.29})$$

$$\leq \sup_{x \in [-N, 0]} H_1(x) + \sup_{x \in [-N, 0]} H_2(x),$$

where

$$H_1(x) = \left| \frac{(2k^* + k(x))(n - k^*)}{n^2} \int \sum_{i=k^*+k(x)+1}^{k^*} \varepsilon_i(t) \delta_n(t) - 2\theta(1 - \theta) \Gamma_n^{(1)}(-x, t) \frac{\delta_n(t)}{\|\delta_n\|} dt \right|,$$

$$H_2(x) = \left| \frac{k(x)}{n} \frac{2k^* + k(x)}{n} \frac{n - k^*}{n} \int \sum_{i=1}^n \varepsilon_i(t) \delta_n(t) ds \right|.$$

According (A.26) and since $k(x)/n = O(1/n\|\delta_n\|^2)$ uniformly in $x \in [-N, 0]$, it follows that

$$\sup_{x \in [-N, 0]} H_2(x) = O(1) \frac{1}{\sqrt{n}\|\delta_n\|} \int \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i(t) \frac{\delta_n(t)}{\|\delta_n\|} dt = o_P(1). \quad (\text{A.30})$$

Furthermore, $(2k^* + k(x))(n - k^*)/n^2 \rightarrow 2\theta(1 - \theta)$ as $n \rightarrow \infty$. Hence,

$$\sup_{x \in [-N, 0]} H_1(x) = O(1) \sup_{x \in [-N, 0]} \left| \int \left(\|\delta_n\| \sum_{i=k^*+k(x)+1}^{k^*} \varepsilon_i(t) - \Gamma_n^{(1)}(-x, t) \right) \frac{\delta_n(t)}{\|\delta_n\|} dt \right|. \quad (\text{A.31})$$

It follows from the Cauchy–Schwarz inequality that

$$\left| \int \left(\|\delta_n\| \sum_{i=k^*+k(x)+1}^{k^*} \varepsilon_i(t) - \Gamma_n^{(1)}(-x, t) \right) \frac{\delta_n(t)}{\|\delta_n\|} dt \right| \leq \left\| \|\delta_n\| \sum_{i=k^*+k(x)+1}^{k^*} \varepsilon_i(\cdot) - \Gamma_n^{(1)}(-x, \cdot) \right\|.$$

This combined with (A.30) and (A.26) implies that

$$\sup_{x \in [-N, 0]} H_1(x) = o_P(1),$$

which with (A.30) and (A.29) establishes (A.28). A parallel argument shows that

$$\sup_{x \in [0, N]} \left| \frac{1}{n} \left\langle E_{k^*+k(x)}^{(3)}, D_{k^*+k(x)}^{(4)} \right\rangle - 2\theta(1 - \theta) \int \Gamma_n^{(2)}(x, t) \frac{\delta_n(t)}{\|\delta_n\|} dt \right| = o_P(1). \quad (\text{A.32})$$

For $x \in [0, N]$, let

$$\Psi_n(x) = \int \Gamma_n^{(1)}(-x, t) \frac{\delta_n(t)}{\|\delta_n\|} dt.$$

Then $(\Psi(x) : x \in [0, N])$ defined by $\Psi(x) = \lim_{n \rightarrow \infty} \Psi_n(x)$ for $x \in [0, N]$, is a Gaussian process. It follows from Fubini's theorem and the definition of σ^2 that $E[\Psi(x)] = 0$ and $\text{Cov}(\Psi(x), \Psi(x')) = \sigma^2 \min(x, x')$, which implies that $\Psi = \sigma B$, where B is a Brownian motion and equality is in $\mathcal{D}[0, N]$. The same argument applies when Ψ_n is defined using $\Gamma_n^{(2)}$. Therefore, (A.28) and (A.32) imply, upon showing that the remaining terms in (A.5) and (A.6) do not asymptotically contribute, that the stochastic part of the process $(R_{n, k^*+k(x)} : x \in [-N, N])$ converges to that of $(2\theta(1 - \theta)V(x), x \in [-N, N])$. To see that the other terms do not contribute, note first that, by the triangle inequality,

$$\max_{k^*-N_\delta \leq k \leq k^*} \left| \frac{1}{n} \left\langle D_k^{(1)}, E_k^{(2)} \right\rangle \right| \leq T_1 + T_2 + T_3,$$

where

$$\begin{aligned} T_1 &= \max_{k^*-N_\delta \leq k \leq k^*} \frac{k-k^*}{n} \frac{n-k^*}{n} \int \sum_{i=1}^k \varepsilon_i(t) \delta_n(t) dt, \\ T_2 &= \max_{k^*-N_\delta \leq k \leq k^*} \frac{k-k^*}{n} \frac{n-k^*}{n} \int \sum_{i=1}^{k^*} \varepsilon_i(t) \delta_n(t) dt, \\ T_3 &= \max_{k^*-N_\delta \leq k \leq k^*} \frac{k+k^*}{n} \frac{k-k^*}{n} \frac{n-k^*}{n} \int \sum_{i=1}^n \varepsilon_i(t) \delta_n(t) dt. \end{aligned}$$

It follows from the definition N_δ that for all k satisfying $k^* - N_\delta \leq k \leq k^*$,

$$\left| \frac{k-k^*}{n} \right| \leq \frac{N}{n \|\delta_n\|^2}.$$

This implies that

$$\begin{aligned} T_1 &\leq O(1) \frac{N}{n \|\delta_n\|^2} \max_{k^*-N_\delta \leq k \leq k^*} \int \sum_{i=1}^k \varepsilon_i(t) \delta_n(t) dt \\ &= O(1) \frac{N}{\sqrt{n \|\delta_n\|^2}} \max_{k^*-N_\delta \leq k \leq k^*} \int \frac{1}{\sqrt{n}} \sum_{i=1}^k \varepsilon_i(t) \frac{\delta_n(t)}{\|\delta_n\|} dt = o_P(1), \end{aligned}$$

where in the last line Lemma B.2 and the continuous mapping theorem were applied. The same argument with small modifications shows that $T_2 = o_P(1)$ and $T_3 = o_P(1)$. Therefore,

$$\max_{k^*-N_\delta \leq k \leq k^*} \left| \frac{1}{n} \langle D_k^{(1)}, E_k^{(2)} \rangle \right| = o_P(1).$$

Additionally, by the Cauchy–Schwarz and triangle inequalities,

$$\begin{aligned} &\left| \frac{1}{n} \langle E_k^{(1)}, E_k^{(2)} \rangle \right| \\ &= \left| \frac{1}{n} \int \left(\sum_{i=k+1}^{k^*} \varepsilon_i(t) + \frac{k-k^*}{n} \sum_{i=1}^n \varepsilon_i(t) \right) \left(\sum_{i=1}^k \varepsilon_i(t) + \sum_{i=1}^{k^*} \varepsilon_i(t) - \frac{k+k^*}{n} \sum_{i=1}^n \varepsilon_i(t) \right) dt \right| \\ &\leq \frac{1}{n} \left\| \sum_{i=k+1}^{k^*} \varepsilon_i + \frac{k-k^*}{n} \sum_{i=1}^n \varepsilon_i \right\| \left\| \sum_{i=1}^k \varepsilon_i + \sum_{i=1}^{k^*} \varepsilon_i - \frac{k+k^*}{n} \sum_{i=1}^n \varepsilon_i \right\| \\ &\leq \frac{1}{n} \left(\left\| \sum_{i=k+1}^{k^*} \varepsilon_i \right\| + \left\| \frac{k-k^*}{n} \sum_{i=1}^n \varepsilon_i \right\| \right) \left(\left\| \sum_{i=1}^k \varepsilon_i \right\| + \left\| \sum_{i=1}^{k^*} \varepsilon_i \right\| + \left\| \frac{k+k^*}{n} \sum_{i=1}^n \varepsilon_i \right\| \right). \end{aligned}$$

Therefore,

$$\begin{aligned} \max_{k^*-N_\delta \leq k \leq k^*} \left| \frac{1}{n} \langle E_k^{(1)}, E_k^{(2)} \rangle \right| &\leq \frac{1}{n} \left(\max_{k^*-N_\delta \leq k \leq k^*} \left\| \sum_{i=k+1}^{k^*} \varepsilon_i \right\| + \max_{k^*-N_\delta \leq k \leq k^*} \left\| \frac{k-k^*}{n} \sum_{i=1}^n \varepsilon_i \right\| \right) \\ &\quad \times \left(\max_{k^*-N_\delta \leq k \leq k^*} \left\| \sum_{i=1}^k \varepsilon_i \right\| + \left\| \sum_{i=1}^{k^*} \varepsilon_i \right\| + \max_{k^*-N_\delta \leq k \leq k^*} \left\| \frac{k+k^*}{n} \sum_{i=1}^n \varepsilon_i \right\| \right). \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n} O_P \left(\sqrt{N_\delta} + \frac{N_\delta}{\sqrt{n}} \right) O_P(\sqrt{n}) = O_P \left(\sqrt{\frac{N_\delta}{n}} + \frac{N_\delta}{n} \right) \\
&= o_P(1),
\end{aligned}$$

where Lemma B.2 and the Continuous Mapping Theorem were again used to obtain upper bounds on the maximum norm terms. The remaining terms coming from (A.6) can be shown to be $o_P(1)$ in a similar way. $\square$

Proof of Theorem 2.4. The proof of the theorem follows from the lemmas in this subsection as the proof of Theorem 2.3 from the lemmas in the previous subsection. $\square$

A.3 Proof of Theorem 2.5

Proof of Theorem 2.5. This result follows from Theorem A.1 and the Continuous Mapping Theorem for argmax functionals, see Theorem 2.7 of Kim and Pollard (1990). $\square$

A.4 Proof of Theorem 2.6

Lemma A.5. *Under the assumptions of Theorem 2.4,*

$$|\hat{\theta} - \theta| = o_P(1), \quad (\text{A.33})$$

$$\|\hat{\delta}_n - \delta_n\| = O_P \left(\frac{1}{n\|\delta_n\|} + \frac{1}{\sqrt{n}} \right), \quad (\text{A.34})$$

$$\iint \left(\hat{C}_\varepsilon(t, t') - C_\varepsilon(t, t') \right)^2 dt dt' = o_P(1), \quad (\text{A.35})$$

and

$$|\hat{\sigma}^2 - \sigma^2| = o_P(1). \quad (\text{A.36})$$

Proof. It follows directly from Theorem 2.4 that

$$|\hat{\theta} - \theta| = \left| \frac{\hat{k}_n^* - k^*}{n} \right| = \left| \frac{\|\delta_n\|^2 (\hat{k}_n^* - k^*)}{\|\delta_n\|^2 n} \right| = O_P(1) \frac{1}{\|\delta_n\|^2 n} = o_P(1),$$

which gives (A.33).

Evidently,

$$\|\hat{\delta}_n - \delta\| = \|\hat{\delta}_n - \hat{\delta}\| \mathbf{1}\{\hat{k}_n^* > k^*\} + \|\hat{\delta}_n - \hat{\delta}\| \mathbf{1}\{\hat{k}_n^* \leq k^*\}. \quad (\text{A.37})$$

When $\hat{k}_n^* > k^*$, according to the definition of $\hat{\delta}_n$,

$$\hat{\delta}_n(t) = \frac{1}{n - \hat{k}_n^*} \sum_{i=\hat{k}_n^*+1}^n \varepsilon_i(t) - \frac{1}{\hat{k}_n^*} \sum_{i=1}^{\hat{k}_n^*} \varepsilon_i(t) + \delta_n(t) \left(1 - \frac{\hat{k}_n^* - k^*}{\hat{k}_n^*} \right).$$

From this it follows from the triangle inequality that

$$\|\hat{\delta}_n - \delta_n\| \leq \left\| \frac{1}{n - \hat{k}_n^*} \sum_{i=\hat{k}_n^*+1}^n \varepsilon_i(t) - \frac{1}{\hat{k}_n^*} \sum_{i=1}^{\hat{k}_n^*} \varepsilon_i(t) \right\| + \|\delta_n\| \frac{\hat{k}_n^* - k^*}{\hat{k}_n^*}. \quad (\text{A.38})$$

By Theorem 2.4, $\hat{k}_n^* - k^* / \hat{k}_n^* = O_P(1/(\|\delta_n\|^2 n))$. Hence the second term on the right-hand side of (A.38) is $O_P(1/(\|\delta_n\| n))$. Theorem 2.4 and Lemma B.1 imply that

$$\begin{aligned} \left\| \frac{1}{\hat{k}_n^*} \sum_{i=1}^{\hat{k}_n^*} \varepsilon_i(t) \right\| &= O_P(1) \frac{1}{k^*} \left\| \sum_{i=1}^{\hat{k}_n^*} \varepsilon_i(t) \right\| \\ &\leq O_P(1) \frac{1}{k^*} \max_{1 \leq k \leq n} \left\| \sum_{i=1}^k \varepsilon_i(t) \right\| = O_P\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

The same bound can be obtained for the remaining term on the right hand side of (A.38), thus showing

$$\|\hat{\delta}_n - \delta\| \mathbb{1}\{\hat{k}_n^* > k^*\} = O_P\left(\frac{1}{n\|\delta_n\|} + \frac{1}{\sqrt{n}}\right).$$

A parallel argument can be used to establish the same bound for the second term on the right-hand side of (A.37), from which (A.34) follows.

Towards establishing (A.35), let

$$\mu_i = \begin{cases} \mu, & 1 \leq i \leq k^*. \\ \mu + \delta_n, & k^* + 1 \leq i \leq n. \end{cases}$$

The same arguments used to establish (A.34) also imply that

$$\sup_{1 \leq i \leq n} \|\bar{X}_i^* - \mu_i\| = O_P\left(\frac{1}{n\|\delta_n\|} + \frac{1}{\sqrt{n}}\right).$$

Since under the assumptions of Theorem 2.4, $n\|\delta_n\|^2 \rightarrow \infty$ as $n \rightarrow \infty$, it also follows that $\sqrt{n}\|\delta_n\| \rightarrow \infty$.

Thus,

$$\sup_{1 \leq i \leq n} \|\bar{X}_i^* - \mu_i\| = O_P\left(\frac{1}{\sqrt{n}}\right). \quad (\text{A.39})$$

Let

$$\tilde{C}_\varepsilon(t, t') = \sum_{\ell=-\infty}^{\infty} w_\tau\left(\frac{\ell}{h}\right) \tilde{\gamma}_\ell(t, t'), \quad (\text{A.40})$$

where

$$\tilde{\gamma}_\ell(t, t') = \begin{cases} \frac{1}{n} \sum_{i=1}^{T-\ell} [X_i(t) - \mu_i(t)] [X_{i+\ell}(t') - \mu_i(t')], & \ell \geq 0. \\ \frac{1}{n} \sum_{j=1-\ell}^n [X_i(t) - \mu_i(t)] [X_{i+\ell}(t') - \mu_i(t')], & \ell < 0. \end{cases}$$

It follows along the lines of the calculations on page 17 of Horváth et al. (2013) that (A.39) implies

$$\iint [\hat{C}_\varepsilon(t, t') - \tilde{C}_\varepsilon(t, t')]^2 dt dt' = o_P(1). \quad (\text{A.41})$$

The conditions of Theorem 2 of Horváth et al. (2013) hold under Assumptions 2.1 and 3.1, which yields

$$\iint [\tilde{C}_\varepsilon(t, t') - C_\varepsilon(t, t')]^2 dt dt' = o_P(1). \quad (\text{A.42})$$

The triangle inequality along with (A.41) and (A.42) imply (A.35).

To verify the last claim of the lemma, let

$$\hat{\sigma}_*^2 = \iint C_\varepsilon(t, t') \frac{\hat{\delta}_n(t) \hat{\delta}_n(t')}{\|\hat{\delta}_n\|^2} dt dt'.$$

Use the triangle inequality to obtain

$$|\hat{\sigma}^2 - \sigma^2| \leq |\hat{\sigma}^2 - \hat{\sigma}_*^2| + |\hat{\sigma}_*^2 - \sigma^2|. \quad (\text{A.43})$$

By the Cauchy–Schwarz inequality and (A.35),

$$|\hat{\sigma}^2 - \hat{\sigma}_*^2| \leq \left(\iint [\hat{C}_\varepsilon(t, t') - C_\varepsilon(t, t')]^2 dt dt' \right)^{1/2} = o_P(1).$$

Another application of the Cauchy–Schwarz inequality shows that

$$|\hat{\sigma}_*^2 - \sigma^2| \leq \left(\iint C_\varepsilon^2(t, t') dt dt' \right)^{1/2} \left(\iint \left(\frac{\hat{\delta}_n(t) \hat{\delta}_n(t')}{\|\hat{\delta}_n\|^2} - \delta_u(t) \delta_u(t') \right)^2 dt dt' \right)^{1/2}, \quad (\text{A.44})$$

where $\delta_u = \delta_n / \|\delta_n\|$ is the normalized direction of the break. It follows from the triangle inequality, (A.34) and the definition of δ_u that

$$\begin{aligned} \left(\iint \left(\frac{\hat{\delta}_n(t) \hat{\delta}_n(t')}{\|\hat{\delta}_n\|^2} - \delta_u(t) \delta_u(t') \right)^2 dt dt' \right)^{1/2} &\leq 2 \left\| \frac{\hat{\delta}_n}{\|\hat{\delta}_n\|} - \delta_u \right\| \\ &= O_P \left(\frac{1}{n \|\delta_n\|^2} + \frac{1}{\sqrt{n} \|\delta_n\|} \right) = o_P(1). \end{aligned}$$

This along with (A.44) and (A.43) imply (A.34). $\square$

Lemma A.6. *If Ξ_0 has the same distribution as Ξ with $Q(x)$ defined with the two-sided Brownian motion W from the definition of $\hat{\Xi}$, then*

$$|\hat{\Xi} - \Xi_0| = o_P(1).$$

Proof. Throughout this proof the notation $\{\cdot\}$ will be used to denote the set of all ω in the underlying probability space Ω satisfying the condition $\cdot$; for example, $\{|\hat{\Xi} - \Xi_0| > \epsilon\}$ denotes $\{\omega \in \Omega: |\hat{\Xi}(\omega) - \Xi_0(\omega)| > \epsilon\}$. Let $\epsilon > 0$. According to the definitions of $\hat{\Xi}$ and Ξ_0 ,

$$\{|\hat{\Xi} - \Xi_0| > \epsilon\} = A_1 \cup A_2, \quad (\text{A.45})$$

where $A_1 = \{\hat{Q}(\hat{\Xi}) = \hat{Q}(\Xi_0)\} \cap \{\hat{\Xi} < \Xi_0 - \epsilon\}$ and $A_2 = \{\hat{Q}(\hat{\Xi}) > \hat{Q}(\Xi_0)\} \cap \{|\hat{\Xi} - \Xi_0| > \epsilon\}$. Define the sets

$$S_1 = \{\hat{\Xi} \geq 0, \Xi_0 \geq 0\}, \quad S_2 = \{\hat{\Xi} < 0, \Xi_0 \geq 0\} \quad S_3 = \{\hat{\Xi} \geq 0, \Xi_0 < 0\} \quad \text{and} \quad S_4 = \{\hat{\Xi} < 0, \Xi_0 < 0\}.$$

Then,

$$\begin{aligned} P\left(\{\hat{Q}(\hat{\Xi}) = \hat{Q}(\Xi_0)\} \cap \{\hat{\Xi} < \Xi_0 - \epsilon\} \cap S_1\right) \\ = P\left(\{-\hat{\theta}\hat{\Xi} + \hat{\sigma}W(\hat{\Xi}) = -\hat{\theta}\Xi_0 + \hat{\sigma}W(\Xi_0)\} \cap \{\epsilon < \Xi_0 - \hat{\Xi}\}\right) \\ = P\left(\{W(\Xi_0) - W(\hat{\Xi}) = \frac{\hat{\theta}}{\hat{\sigma}}(\Xi_0 - \hat{\Xi})\} \cap \{\epsilon < \Xi_0 - \hat{\Xi}\}\right) = 0. \end{aligned} \quad (\text{A.46})$$

It follows similarly that $P(\{\hat{Q}(\hat{\Xi}) = \hat{Q}(\Xi_0)\} \cap \{\hat{\Xi} < \Xi_0 - \epsilon\} \cap S_i) = 0$, for $i = 2, 3$, and 4. Therefore, for all $n \geq 1$,

$$P(A_1) = 0. \quad (\text{A.47})$$

In light of this and (A.45), we now seek to show that $\lim_{n \rightarrow \infty} P(A_2) = 0$. For $N > 0$ let

$$B_N = \{|\Xi_0| > N \cup |\hat{\Xi}| > N\}, \quad B_N^c \text{ be the compliment of } B_N,$$

as well as

$$R_1(\kappa) = \{Q(\Xi_0) - Q(\hat{\Xi}) > \kappa\} \quad \text{and} \quad R_2(\kappa) = \{0 \leq Q(\Xi_0) - Q(\hat{\Xi}) \leq \kappa\}.$$

From the definition of Ξ_0 it follows that $R_1(\kappa) \cup R_2(\kappa) = \Omega$. It is then evident that

$$P(A_2) = P(A_2 \cap B_N) + P(A_2 \cap R_1(\kappa) \cap B_N^c) + P(A_2 \cap R_2(\kappa) \cap B_N^c). \quad (\text{A.48})$$

According to the definition of $\hat{\Xi}$, for all $N > 0$,

$$P(\hat{\Xi} > N) \leq P(\hat{\sigma}W(0) < -\hat{\theta}N + \hat{\sigma}W(N)) = 1 - \Phi\left(\frac{\hat{\theta}N^{1/2}}{\hat{\sigma}}\right),$$

where Φ denotes the distribution function of a standard normal random variable. Therefore, Lemma A.5 implies that

$$\lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} P(\hat{\Xi} > N) = 0.$$

It follows similarly that $\lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} P(\hat{\Xi} < -N) = 0$ and $\lim_{N \rightarrow \infty} P(|\Xi_0| > N) = 0$. Now it follows that $\lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} P(B_N) = 0$. This implies that the first term on the right-hand side of (A.48) can be made arbitrarily small as $n \rightarrow \infty$ by taking N sufficiently large.

It follows from Buffet (2003) (see also Chapter 2 of Karatzas and Shreve, 1988), that the random variable $\Xi_0 - \hat{\Xi}$ is absolutely continuous with respect to Lebesgue measure, and so let f_n denote its density. Then,

$$\begin{aligned} & P(\{|\Xi_0 - \hat{\Xi}| > \epsilon\} \cap R_2(\kappa) \cap B_N^c \cap S_1) \\ &= P\left(\left\{\frac{\theta}{\sigma}(\Xi_0 - \hat{\Xi}) \leq W(\Xi_0) - W(\hat{\Xi}) \leq \frac{\kappa}{\sigma} + \frac{\theta}{\sigma}(\Xi_0 - \hat{\Xi})\right\} \cap \{|\Xi_0 - \hat{\Xi}| > \epsilon\} \cap B_N^c\right) \\ &= \int_{\epsilon < |x| < 2N} P\left(\frac{\theta}{\sigma}(\Xi_0 - \hat{\Xi}) \leq W(\Xi_0) - W(\hat{\Xi}) \leq \frac{\kappa}{\sigma} + \frac{\theta}{\sigma}(\Xi_0 - \hat{\Xi}) \mid \Xi_0 - \hat{\Xi} = x\right) f_n(x) dx \\ &= \int_{\epsilon < |x| < 2N} \left[\Phi\left(\frac{\kappa}{|x|^{1/2}} + \frac{\theta}{\sigma}|x|^{1/2}\right) - \Phi\left(\frac{\theta}{\sigma}|x|^{1/2}\right)\right] f_n(x) dx \leq \psi_0 \kappa, \end{aligned} \tag{A.49}$$

for some constant $\psi_0 > 0$, where in the last line it was used that the standard normal distribution function is Lipschitz, and the integral is taken over a subset of $\mathbb{R}$ bounded away from the origin. Hence,

$$\lim_{\kappa \rightarrow 0} P(\{|\Xi_0 - \hat{\Xi}| > \epsilon\} \cap R_2(\kappa) \cap B_N^c \cap S_1) = 0,$$

for all $n \geq 0$. A parallel argument gives the same result when S_1 is replaced with S_i , $i = 2, 3$, and 4. This implies that

$$\lim_{\kappa \rightarrow 0} P(A_2 \cap R_2(\kappa) \cap B_N^c) = 0$$

and the second term on the right hand-side of (A.48) is controlled by taking κ to be small. According to Lemma A.5, for all $N > 0$,

$$D_n = \sup_{x \in [-N, N]} |\hat{Q}(x) - Q(x)| = o_P(1).$$

Consequently, for all $\kappa > 0$,

$$P(A_2 \cap R_1(\kappa) \cap B_N^c) \leq P(\{Q(\hat{\Xi}) > Q(\Xi_0) + D_n\} \cap \{Q(\Xi_0) - Q(\hat{\Xi}) > \kappa\}) \rightarrow 0$$

as $n \rightarrow \infty$. This shows that for κ and N used to control the size of the first two terms on the right-hand side of (A.48), the third term may be made arbitrarily small by taking n sufficiently large. Therefore, $\lim_{n \rightarrow \infty} P(A_2) = 0$. This along with (A.47) completes the proof. $\square$

B Technical Lemmas

Lemma B.1. *If $(\varepsilon_i : i \in \mathbb{Z})$ is a centered functional time series satisfying Assumption 2.1, then*

$$\max_{1 \leq k \leq n} \frac{1}{\sqrt{k}} \left\| \sum_{i=1}^k \varepsilon_i \right\| = O_P(\log^{1/p}(n)) \quad (n \rightarrow \infty).$$

Proof. Let $\rho > 1$. Then, with $c = \lfloor 1/\log(\rho) \rfloor + 1$, it follows that

$$\begin{aligned} \max_{1 \leq k \leq n} \frac{1}{\sqrt{k}} \left\| \sum_{i=1}^k \varepsilon_i \right\| &\leq \max_{1 \leq j \leq c \log(n)} \max_{\rho^{j-1} < k \leq \rho^j} \frac{1}{\sqrt{k}} \left\| \sum_{i=1}^k \varepsilon_i \right\| \\ &\leq \max_{1 \leq j \leq c \log(n)} \frac{1}{\rho^{(j-1)/2}} \max_{1 \leq k \leq \rho^j} \left\| \sum_{i=1}^k \varepsilon_i \right\|. \end{aligned}$$

This implies, by the fact that for arbitrary random variables $(X_i : i \in A)$,

$$P\left(\max_{i \in A} X_i > x\right) \leq P\left(\cup_{i \in A} \{X_i > x\}\right) \leq \sum_{i \in A} P(X_i > x),$$

and Chebyshev's inequality that

$$P\left(\max_{1 \leq k \leq n} \frac{1}{\sqrt{k}} \left\| \sum_{i=1}^k \varepsilon_i \right\| > x\right) \leq \sum_{j=1}^{c \log(n)} \rho^{-(j-1)p/2} x^{-p} \mathbb{E}\left[\left(\max_{1 \leq k \leq \rho^j} \left\| \sum_{i=1}^k \varepsilon_i \right\|\right)^p\right]. \quad (\text{B.1})$$

Corollary 1 to Proposition 4 of Berkes et al. (2011) may be easily adapted to the Hilbert space case, from which

$$\mathbb{E}\left[\left(\max_{1 \leq k \leq \rho^j} \left\| \sum_{i=1}^k \varepsilon_i \right\|\right)^p\right] \leq c_0 \rho^{jp/2},$$

and hence, with (B.1),

$$P\left(\max_{1 \leq k \leq n} \frac{1}{\sqrt{k}} \left\| \sum_{i=1}^k \varepsilon_i \right\| > x\right) \leq c_1 \log(n) \rho^{p/2} x^{-p}.$$

Taking $x = c_2 \log^{1/p}(n)$ with a suitably large constant c_2 completes the proof. $\square$

Lemma B.2. Let $(\varepsilon_i : i \in \mathbb{Z})$ be a centered functional time series satisfying Assumption 2.1 for some $p \geq 2$ (instead of $p > 2$) and, for $N \in \mathbb{N}$, $x \in [0, N]$ and $t \in [0, 1]$, define

$$S_n^{(1)}(x, t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nx \rfloor} \varepsilon_i(t) \quad \text{and} \quad S_n^{(2)}(x, t) = \frac{1}{\sqrt{n}} \sum_{i=-\lfloor nx \rfloor}^{-1} \varepsilon_i(t).$$

Then, there exist two independent sequences of Gaussian processes $(\Gamma_n^{(1)} : n \in \mathbb{N})$ and $(\Gamma_n^{(2)} : n \in \mathbb{N})$ such that

$$\sup_{0 \leq x \leq N} \int \left(S_n^{(1)}(x, t) - \Gamma_n^{(1)}(x, t) \right)^2 dt = o_P(1), \quad (\text{B.2})$$

and

$$\sup_{0 \leq x \leq N} \int \left(S_n^{(2)}(x, t) - \Gamma_n^{(2)}(x, t) \right)^2 dt = o_P(1), \quad (\text{B.3})$$

where $\mathbb{E}[\Gamma^{(j)}(x, t)] = 0$ and $\text{Cov}(\Gamma_n^{(j)}(x, t), \Gamma_n^{(j)}(x', t')) = \min\{x, x'\} C_\varepsilon(t, t')$, $j = 1, 2$, with $C_\varepsilon(t, t')$ defined in (2.4).

Proof. According to Theorem 1.2 in Jirak (2013), one can define $\Gamma_n^{(2)}$ such that it is measurable with respect to $\sigma(\epsilon_i: i \leq 0)$ and satisfies (B.3). Let

$$S_n^{(1,*)}(x, t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nx \rfloor} \epsilon_i^{(i)}(t),$$

where $\epsilon_i^{(i)}$ is defined in Assumption 2.1. Note that $S_n^{(1,*)}$ and $S_n^{(2)}$ are independent. It then follows from the triangle inequality and Lyapounov's inequality that

$$\begin{aligned} \mathbb{E} \left[\sup_{0 \leq x \leq N} \|S_n^{(1)}(x, \cdot) - S_n^{(1,*)}(x, \cdot)\| \right] &\leq \frac{1}{\sqrt{n}} \mathbb{E} \left[\sup_{0 \leq x \leq N} \sum_{i=1}^{\lfloor nx \rfloor} \|\epsilon_i - \epsilon_i^{(i)}\| \right] \\ &\leq \frac{1}{\sqrt{n}} \sum_{i=1}^{\infty} (\mathbb{E}[\|\epsilon_i - \epsilon_i^{(i)}\|^2])^{1/2} \\ &= o(1). \end{aligned}$$

Now Markov's inequality implies that

$$\sup_{0 \leq x \leq N} \int \left(S_n^{(1)}(x, t) - S_n^{(1,*)}(x, t) \right)^2 dt = o_P(1). \quad (\text{B.4})$$

Furthermore, again using Theorem 1.2 of Jirak (2013), one can define a sequence $\Gamma_n^{(1)}$, measurable with respect to $\sigma(\epsilon_i: i > 0; \epsilon_{0,0,i}^*: i \leq 0)$, that satisfies

$$\sup_{0 \leq x \leq N} \int \left(S_n^{(1,*)}(x, t) - \Gamma_n^{(1)}(x, t) \right)^2 dt = o_P(1).$$

This and (B.4) imply (B.2), and the sequences $\Gamma_n^{(1)}$ and $\Gamma_n^{(2)}$ have been constructed to be independent. $\square$

C Additional implementation details

C.1 Selection of weight function and bandwidth

This section lends support to some of the statements on the choice of weight function and the selection of bandwidth needed in order to estimate the long-run covariance operator C_ε . To highlight the robustness of the estimation procedure with respect to the weight function and the relative importance of a good choice of the bandwidth a small simulation study was conducted.

As outlined in Section 4 of the main paper, FAR(1) processes with $\kappa = 0.5, 0.75$ and 0.9 of length $n = 100$ and 200 were generated under Settings 2 (fast) and 3 (slow), and under both H_0 and H_A , where in the latter case a structural break in the mean was inserted at $k^* = n/2$ through the function δ_m as in (4.1) of the main paper with $m = 5$, setting the constant c such that the signal-to-noise ratio equaled 0.0 (H_0) and $0.25, 0.5$ (H_A). For all situations, the proposed detection procedure was applied with the estimator $\hat{C}_\varepsilon$

specified with three versions of weight functions w and four choices of bandwidths h . Namely, the *Bartlett kernel*, the *flat-top kernel* and the *Parzen kernel* were considered, respectively given by

$$\begin{aligned} w_b(x) &= (1 - |x|)\mathbf{1}_{[0,1]}(|x|), \\ w_f(x) &= \mathbf{1}_{[0,.1)}(|x|) + (1.1 - |x|)\mathbf{1}_{[.1,1.1)}, \\ w_p(x) &= (1 - 6x^2 + 6|x|^3)\mathbf{1}_{[0,1/2)}(|x|) + 2(1 - |x|)^3\mathbf{1}_{[1/2,1]}(|x|). \end{aligned}$$

This paragraph discusses the estimation of the bandwidth $h = \hat{M}n^{1/(1+2\tau)}$ introduced in Section 3 of the main paper, which is referred to as $h = h_{\text{opt}}$ below. This bandwidth is only defined when the final weight function is of finite order, i.e., in case of Bartlett and Parzen weight functions. The details of how to estimate $\hat{M}$ are given in Rice and Shang (2017), and require the choice of initial bandwidths and weight functions used to estimate C_ε and $C_\varepsilon^{(\tau)}$. For this the initial weight functions are always taken to be as Bartlett kernel in Step 1 of the algorithm in Rice and Shang (2017), and an initial bandwidth $h_{\text{initial}} = n^{1/3}$ was chosen. The following four bandwidth choices were compared:

$$h_1 = n^{1/3}, \quad h_2 = n^{1/4}, \quad h_3 = n^{1/5} \quad \text{and} \quad h_{\text{opt}}.$$

The simulation results are displayed in Figures C.1 to C.3. It can be seen that the performance across the kernel choices is very similar for all three choices of SNR, and both eigenvalue decays specified through Settings 2 and 3. This supports the statement that the choice of weight function w plays a less significant role in the estimation of $\hat{C}_\varepsilon$. This is an expected behavior known from traditional time series analysis. On the other hand, there are some performance discrepancies for the different bandwidth choices under consideration. These are mostly visible in the levels. For example, it can be seen that only h_{opt} keeps levels for the strong dependence $\kappa = 0.95$, while the more ad hoc choices h_1 to h_3 often reject too often under H_0 . Power appears to be less compromised, even for the small signal-to-noise ratios in this simulation.

C.2 Exact quantiles for confidence intervals

Note that the quantile calculation for $\hat{\Xi}$ can be done exactly using the results of Bhattacharya and Brockwell (1976) and their extension in Stryhn (1996). Let $\hat{\sigma}$ and $\hat{\theta}$ be as in Section 3.3 of the main paper. Then, the probability density function of $\hat{\Xi}$ is given by

$$p_{\hat{\Xi}}(t|\hat{\theta}, \hat{\sigma}) = \begin{cases} f\left(-t \left| \frac{1-\hat{\theta}}{\hat{\sigma}}, \frac{\hat{\theta}}{\hat{\sigma}}\right.\right), & t < 0, \\ f\left(t \left| \frac{\hat{\theta}}{\hat{\sigma}}, \frac{1-\hat{\theta}}{\hat{\sigma}}\right.\right), & t > 0, \end{cases}$$

where, with $\Phi(\cdot)$ denoting the standard normal cdf,

$$f(t|\vartheta_1, \vartheta_2) = 2\vartheta_1(\vartheta_1 + 2\vartheta_2) \exp(2\vartheta_2(\vartheta_1 + \vartheta_2)t) \Phi(-(\vartheta_1 + 2\vartheta_2)\sqrt{t}) - 2\vartheta_1^2 \Phi(-\vartheta_1\sqrt{t}), \quad t \geq 0.$$

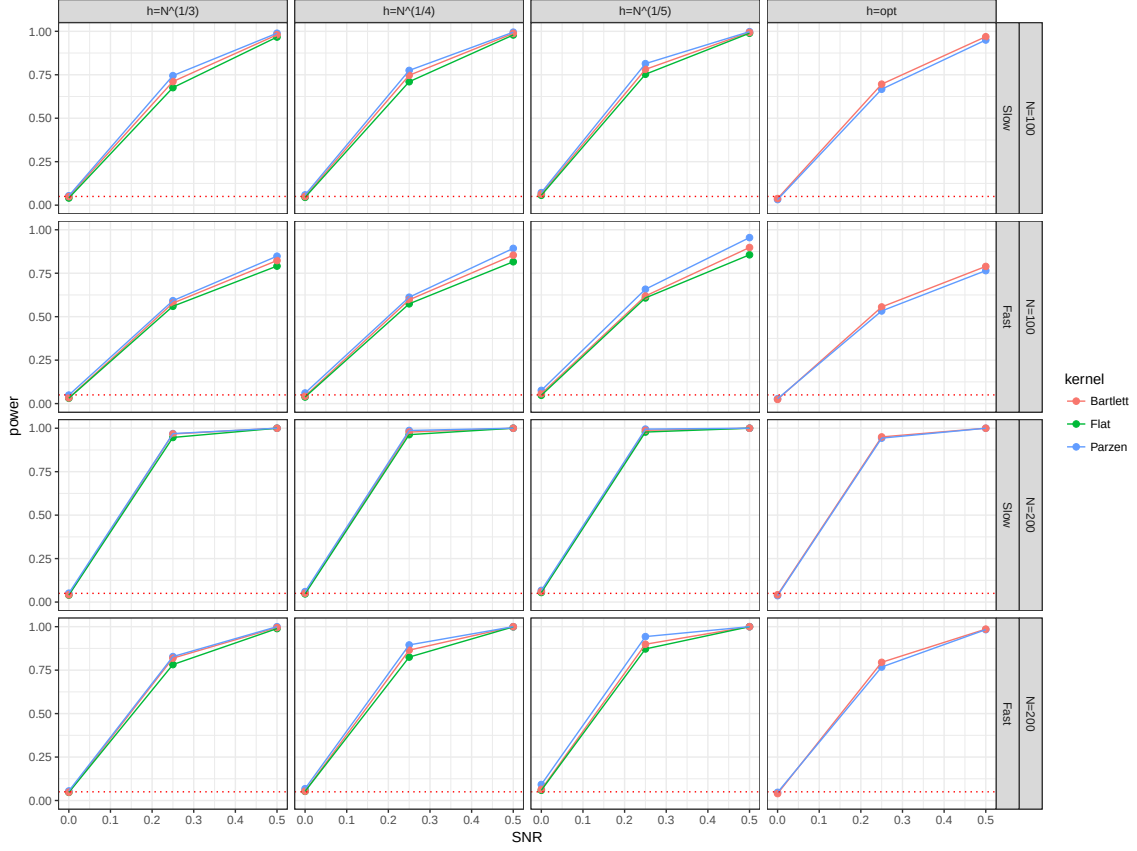


Figure C.1: Simulated power curves for an FAR(1) process with $\kappa = 0.5$ for various kernels, bandwidths, eigenvalue decays and sample sizes. The x -axis displays SNR, the y -axis empirical power. The red horizontal line indicates the nominal level 0.05.

Furthermore, integrals of the function $f(\cdot|\vartheta_1, \vartheta_2)$, with $\vartheta_1, \vartheta_2 > 0$, may easily be computed from

$$\begin{aligned}
 F^+(t|\vartheta_1, \vartheta_2) &= \int_0^t f(t'|\vartheta_1, \vartheta_2) dt' \\
 &= \frac{\vartheta_2}{\vartheta_1 + \vartheta_2} + \frac{2\vartheta_1\sqrt{x}}{\sqrt{2\pi}} \exp(\vartheta_1^2 x/2) \\
 &\quad + \frac{\vartheta_1(\vartheta_1 + 2\vartheta_2)}{\vartheta_2(\vartheta_1 + \vartheta_2)} \exp(2\vartheta_2(\vartheta_1 + \vartheta_2)x) \Phi(-(\vartheta_1 + 2\vartheta_2)\sqrt{x}) \\
 &\quad - \left(2\vartheta_1^2 x + \frac{\vartheta_1^2 + 2\vartheta_2^2 + 2\vartheta_1\vartheta_2}{\vartheta_2(\vartheta_1 + \vartheta_2)} \right) \Phi(-\vartheta_1\sqrt{x}).
 \end{aligned}$$

Taking limits for $t \rightarrow \infty$ yields

$$F(\vartheta_1, \vartheta_2) = \lim_{t \rightarrow \infty} F^+(t|\vartheta_1, \vartheta_2) = \frac{\vartheta_2}{\vartheta_1 + \vartheta_2}.$$

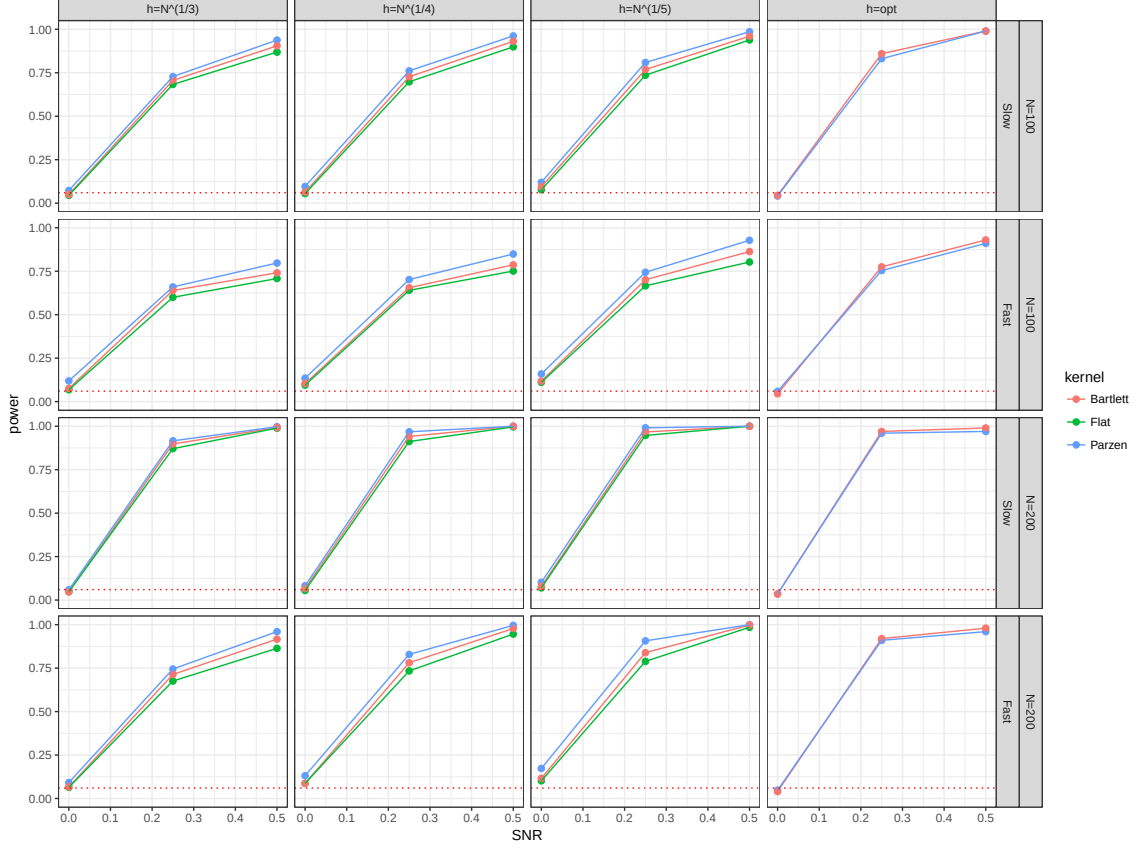


Figure C.2: As in Figure C.1 but with $\kappa = 0.75$.

Thus standard integration techniques give

$$F_{\hat{\Xi}}(t|\hat{\theta}, \hat{\sigma}) = \begin{cases} \frac{\vartheta_2^-}{\vartheta_1^- + \vartheta_2^-} - F^+(-t|\vartheta_1^-, \vartheta_2^-), & t < 0, \\ \frac{\vartheta_2^-}{\vartheta_1^- + \vartheta_2^-} + F^+(t|\vartheta_1^+, \vartheta_2^+), & t > 0, \end{cases}$$

where $\vartheta_1^- = \vartheta_2^+ = (1 - \hat{\theta})/\hat{\sigma}$ and $\vartheta_2^- = \vartheta_1^+ = \hat{\theta}/\hat{\sigma}$. Now quantiles of the distribution $\hat{\Xi}$ may be found through the use of an iterative method. Let $\hat{\Xi}_q$ be the q th quantile of $F_{\hat{\Xi}}$, that is, $F_{\hat{\Xi}}(\hat{\Xi}_q|\hat{\theta}, \hat{\sigma}) = q$. Then, $\hat{\Xi}_q$ can be computed by finding the roots of $F_{\hat{\Xi}}(\hat{\Xi}_q|\hat{\theta}, \hat{\sigma}) - q = 0$ with Newton–Raphson iterations

$$\hat{\Xi}_{q,n+1} = \hat{\Xi}_{q,n} - \frac{F_{\hat{\Xi}}(\hat{\Xi}_{q,n}|\hat{\theta}, \hat{\sigma}) - q}{p_{\hat{\Xi}}(\hat{\Xi}_{q,n}|\hat{\theta}, \hat{\sigma})}$$

with starting value $\hat{\Xi}_{q,0} = 0$. The iterations are stopped once $|F_{\hat{\Xi}}(\hat{\Xi}^*|\hat{\theta}, \hat{\sigma}) - q| < \epsilon$ for some predetermined $\epsilon > 0$. Then, set $\hat{\Xi}_q = \hat{\Xi}^*$. For the practical purposes ϵ was set to equal 10^{-7} .

D Additional simulation evidence

This section provides some additional simulation results that complement Section 4 of the main paper. The set-up is as described in Section 4.1 there. In particular Setting 1 refers to processes designed to highlight the

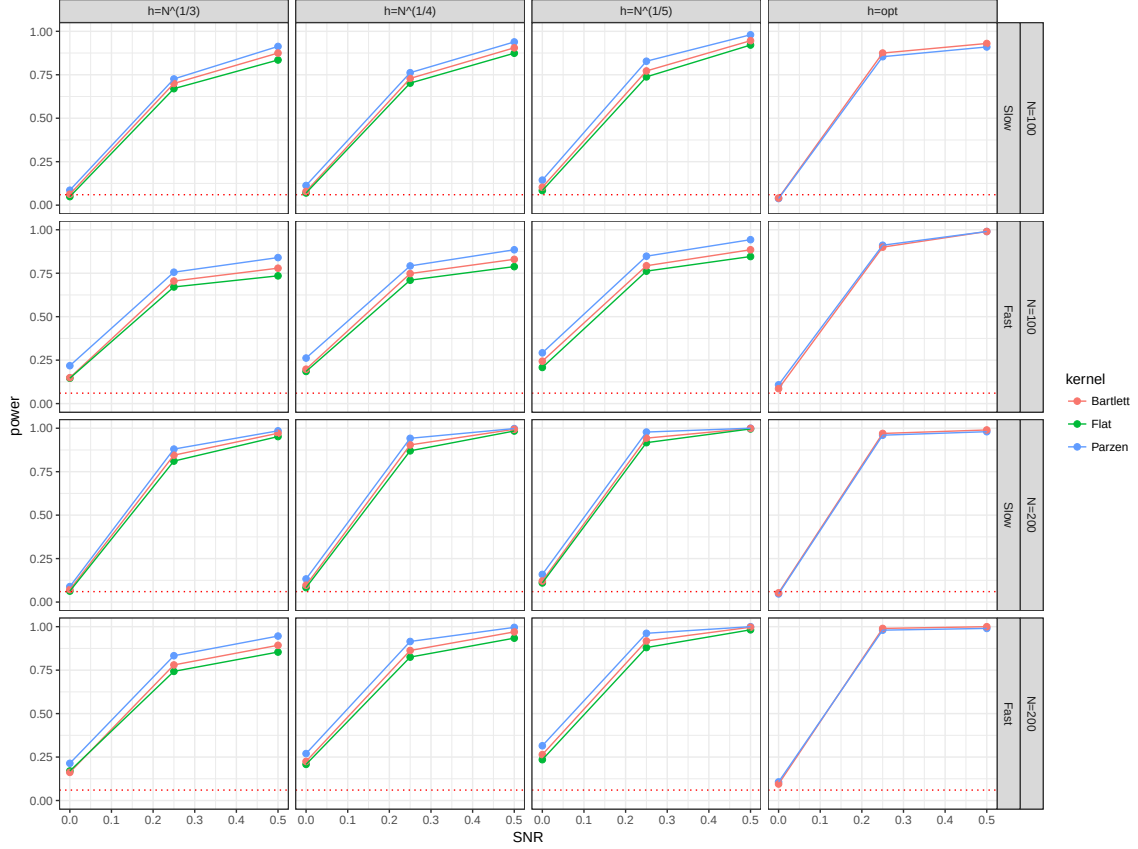


Figure C.3: As in Figure C.1 but with $\kappa = 0.95$.

asymptotic features in finite samples, Setting 2 to the case of fast decay of eigenvalues, and Setting 3 to the case of slow decay of eigenvalues.

Figure D.1 shows the power curves corresponding and the boxplots corresponding to Figures 4.1 and 4.2 of the main paper, the difference being that here FAR(1) processes as described in Section 4.1 were used instead of independent, identically distributed functions. The plots indicate that under dependence the same conclusions remain generally valid as for the independent case.

Table D.1 shows empirical coverages for the confidence intervals for Setting 2 for two locations of the break ($\theta = 0.25$ and 0.5), three choices of nominal level ($\alpha = 0.05, 0.10$ and 0.15), three values of SNR ($0.25, 0.5$ and 1.0), and various forms of the break function δ_m . Table D.2 displays the corresponding results for Setting 3. It can be seen that the coverage rates are above nominal levels for breaks occurring in the middle of the sample ($\theta = 0.5$), pointing to the conservativeness of the intervals. As expected, coverage rates decrease for the breaks occurring away from the middle of the sample ($\theta = 0.25$). Coverage rates appear to be robust against the specification of the eigenvalue decays. The differences between the numbers in Tables D.1 and D.2 are seen to be minor.

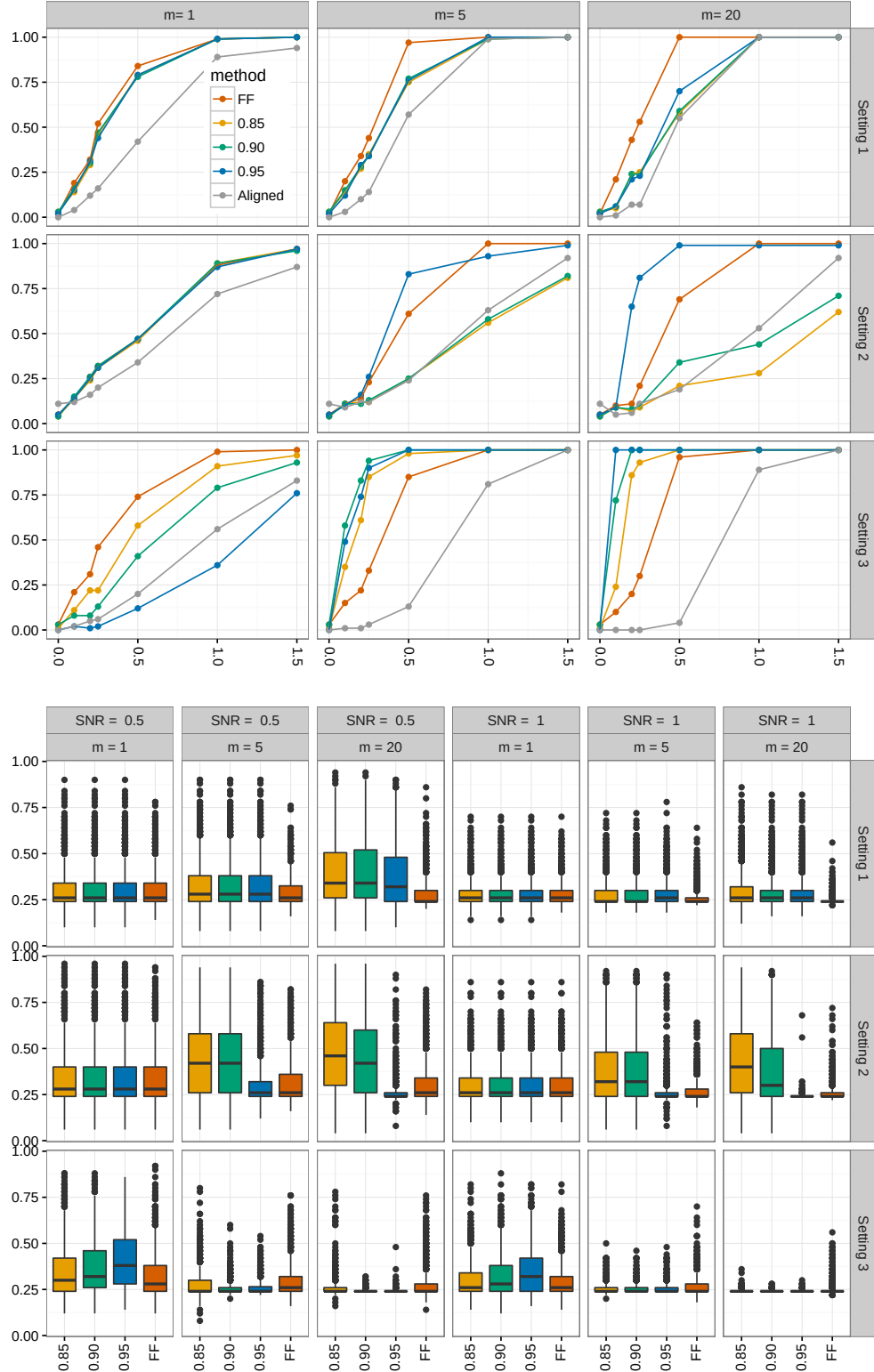


Figure D.1: Upper panel: Power curves for the various break detection procedures for FAR(1) errors. The x -axis gives different choices of SNR. Observe that “FF” refers to the proposed fully functional method, “0.85”, “0.90” and “0.95” correspond to the three levels of TVE in the fPCA procedures, and Aligned to the method of Torgovitski (2016). Lower panel: Box plots for the various break dating procedures.

θ	α	SNR	$m = 1$	3	5	10	20
0.25	0.05	0.25	0.96	0.93	0.94	0.94	0.95
		0.50	0.88	0.87	0.90	0.86	0.84
		1.00	0.92	0.86	0.94	0.95	0.97
	0.1	0.25	0.85	0.74	0.77	0.79	0.70
		0.50	0.75	0.72	0.85	0.78	0.82
		1.00	0.76	0.76	0.80	0.86	0.92
	0.15	0.25	0.72	0.72	0.64	0.55	0.54
		0.50	0.76	0.64	0.70	0.76	0.84
		1.00	0.68	0.76	0.78	0.90	0.90
0.5	0.05	0.25	1.00	1.00	1.00	1.00	1.00
		0.50	1.00	1.00	1.00	1.00	1.00
		1.00	1.00	1.00	1.00	1.00	1.00
	0.1	0.25	1.00	1.00	1.00	1.00	1.00
		0.50	1.00	1.00	1.00	1.00	1.00
		1.00	1.00	1.00	1.00	1.00	1.00
	0.15	0.25	1.00	1.00	1.00	1.00	1.00
		0.50	0.99	1.00	1.00	1.00	0.99
		1.00	0.99	0.99	1.00	1.00	1.00

Table D.1: Empirical coverages (in %) for fully functional confidence intervals constructed from Theorem 3.1 for $n = 100$ and Setting 2.

θ	α	SNR	$m = 1$	3	5	10	20
0.25	0.05	0.25	0.96	0.95	0.81	0.81	0.79
		0.50	0.89	0.92	0.83	0.91	0.91
		1.00	0.74	0.87	0.88	0.95	0.98
	0.1	0.25	0.76	0.69	0.69	0.71	0.76
		0.50	0.71	0.76	0.78	0.72	0.81
		1.00	0.75	0.81	0.88	0.86	0.97
	0.15	0.25	0.70	0.54	0.59	0.61	0.60
		0.50	0.66	0.64	0.65	0.76	0.87
		1.00	0.72	0.81	0.85	0.85	0.90
0.5	0.05	0.25	1.00	1.00	1.00	1.00	1.00
		0.50	1.00	1.00	1.00	0.99	1.00
		1.00	1.00	1.00	1.00	1.00	1.00
	0.1	0.25	1.00	1.00	1.00	1.00	1.00
		0.50	1.00	1.00	1.00	1.00	1.00
		1.00	1.00	1.00	1.00	1.00	1.00
	0.15	0.25	0.99	1.00	1.00	1.00	0.99
		0.50	0.98	0.99	0.99	0.99	1.00
		1.00	0.99	0.99	1.00	1.00	1.00

Table D.2: Empirical coverages (in %) for fully functional confidence intervals constructed from Theorem 3.1 for $n = 100$ and Setting 3.

E Additional temperature data analysis

Figures E.1 and E.2 contain the time series plots of annual temperature curves for all eight stations together with the respective plots of how the estimated break functions load on the leading eigendirections.

F Intra-day log-returns of Microsoft stock

In this section, the proposed methodology is applied to one-minute log-returns of Microsoft stock and contrasted from the fPCA based competitor methods. The observations span the time period starting on 06/13/2001 and ending on 11/07/2001. During each day, 390 stock price values were recorded in one-minute intervals from 9:30 AM to 4:00 PM EST. Rescaling intra-day time to the interval $[0, 1]$ by a linear transformation, let $P_i(t)$ be the Microsoft stock price at intra-day time $t \in [0, 1]$ on day $i = 1, \dots, 100$. The (scaled) cumulative intra-day returns were then computed as

$$R_i(t) = 100[\ln P_i(t) - \ln P_i(0)], \quad t \in [0, 1], \quad i = 1, \dots, 100.$$

The underlying discrete data was converted to functional objects using $D = 31$ B -spline functions. The results reported below are robust against the specification of D , as virtually the same conclusions were reached for a range of other D values. The resulting 100 curves are plotted in Figure F.1. An application of fPCA to this data revealed that the first component explains about 90% of the variation in the log-return data. Both fully functional and fPCA based break point dating procedures were applied with both methods selecting $\hat{k}_{100}^* = \tilde{k}_{100}^* = 64$, corresponding to the calendar date 09/18/2001, as the estimated break date. This date coincides with the second day after the re-opening of the stock markets after the September 11 terrorist attacks. Figure F.2 displays both the first empirical eigenfunction and the sample mean curves prior to and post the estimated break date. The first eigenfunction accounts for the general tendency of the log-returns to increase or decrease (depending on the sign) during a trading day. It can be seen that prior to 9/21/2001, this tendency was negative, while it was positive thereafter.

A natural follow-up question is if the eigenfunctions associated with smaller sample eigenvalues suffer from a break as well. If that was the case, then there was also a change in deviations from the “general tendency” implied by $\hat{\varphi}_1$. These deviations might be interpreted as the risk incurred when trusting the log-return behavior predicted by the main direction of variation. Risk was assessed in the following way. First the impact of the first empirical eigenfunction was removed by constructing the new curves

$$\tilde{P}_i = P_i - \langle P_i, \hat{\varphi}_1 \rangle \hat{\varphi}_1, \quad i = 1, \dots, 100.$$

Applying the proposed methodology to the transformed data leads to selecting the same break date $\hat{k}^* = 64$. However, the break date selection is highly variable for the fPCA methodology. The results for d varying be-

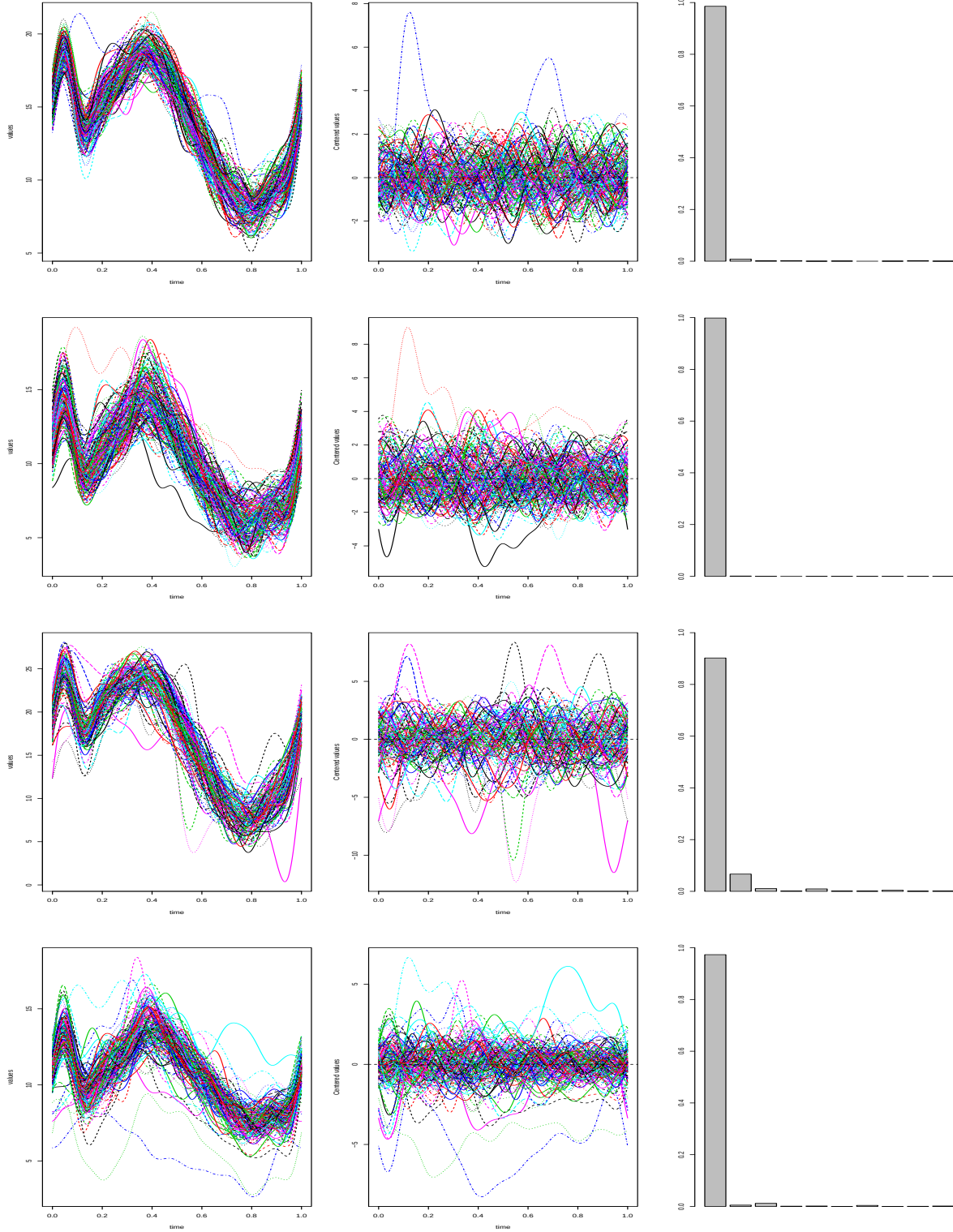


Figure E.1: Time series plots of annual temperature profiles (left), centered profiles (center) and proportion of variation in the estimated break function explained by the leading sample eigenfunctions (right) at Sydney (Observatory Hill), Melbourne (Regional Office), Boulia Airport, and Cape Otway Lighthouse (top row to bottom row).

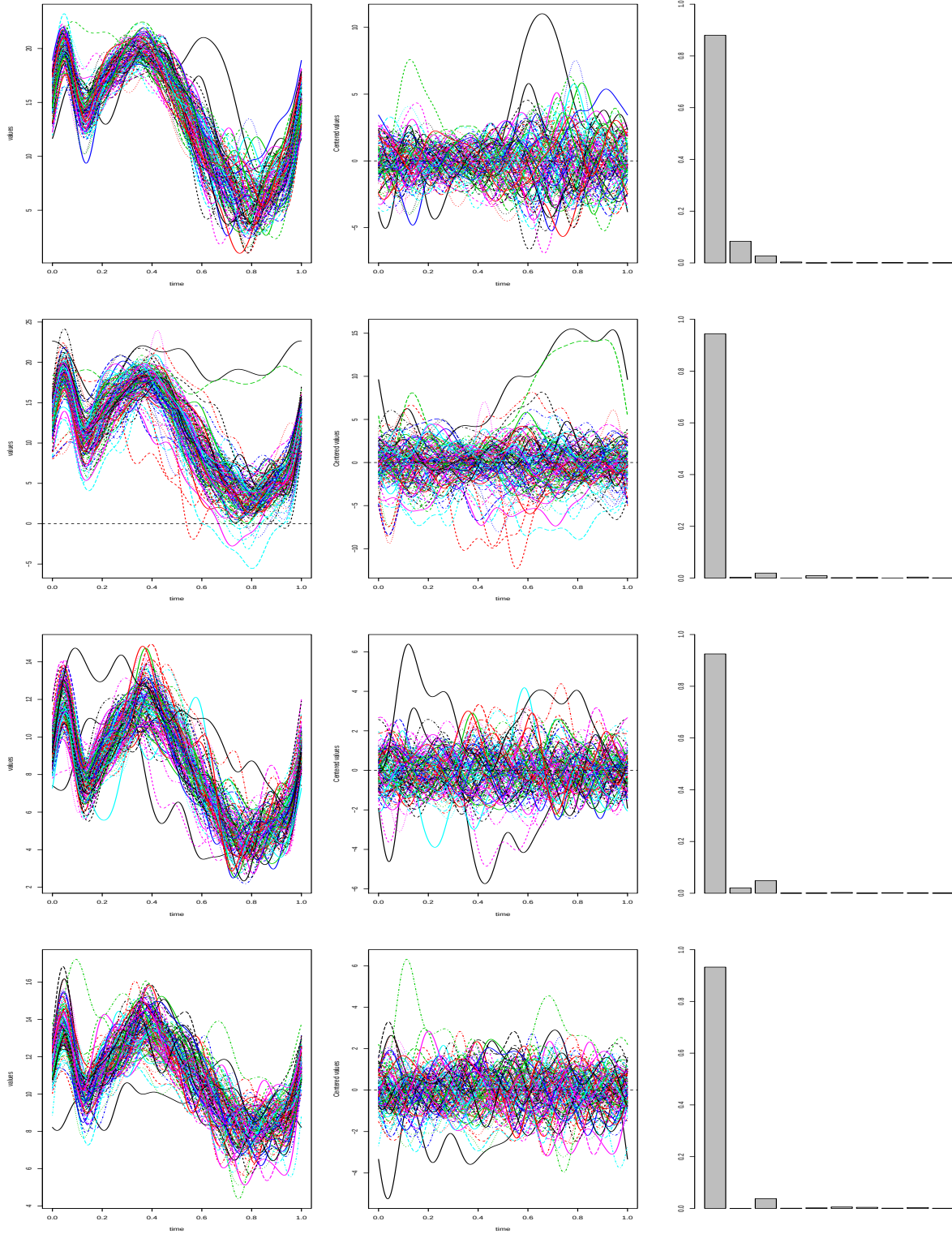


Figure E.2: Time series plots of annual temperature profiles (left), centered profiles (center) and proportion of variation in the estimated break function explained by the leading sample eigenfunctions (right) at Gayndah Post Office, Gunnedah Pool, Hobart (Ellerslie Road), and Robe Comparison (top row to bottom row).

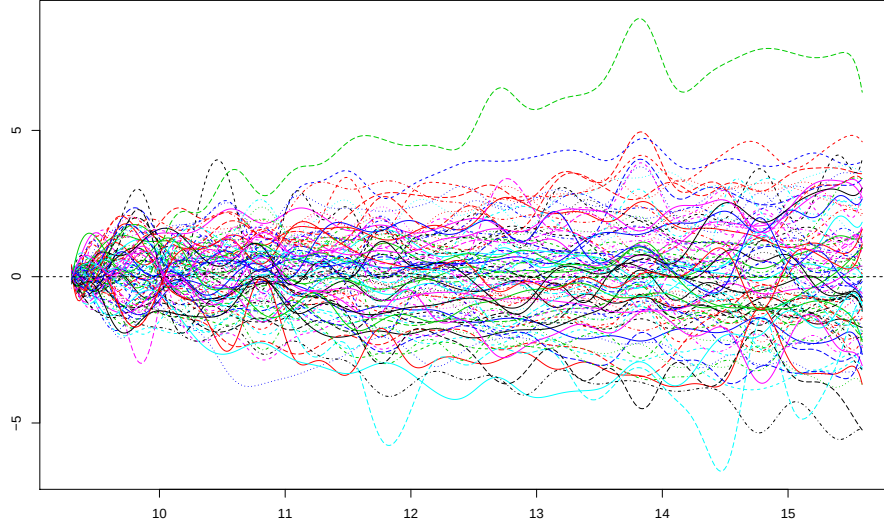


Figure F.1: Daily cumulative log-return curves for Microsoft stock from 6/13/2001 to 11/07/2001. The x -axis gives clock time, the y -axis is proportional to percentage change.

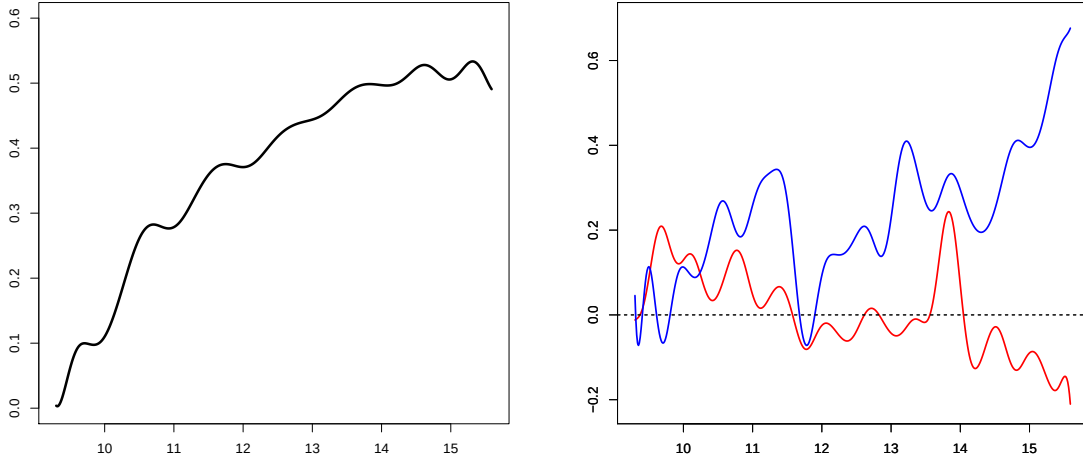


Figure F.2: Left: the (smoothed) first eigenfunction obtained from fPCA. Right: mean curves prior to (red) and post (blue) the estimated break date 09/18/2001.

tween 1 and 5 (corresponding to the second to fifth sample eigenfunctions of the original data) are summarized in Table F.1. For any $d > 5$, the estimated change is 09/18/2001.

Using $\hat{k}^* = 64$ for the computations to follow, the reason for this phenomenon can be found in how the estimated break function $\hat{\delta} = \hat{\mu}_{\text{post}} - \hat{\mu}_{\text{prior}}$ distributes among the sample eigenfunctions, where $\hat{\mu}_{\text{prior}}$ and $\hat{\mu}_{\text{post}}$ are the sample mean curves on the pre-break and the post-break sample, respectively. Figure F.4 shows

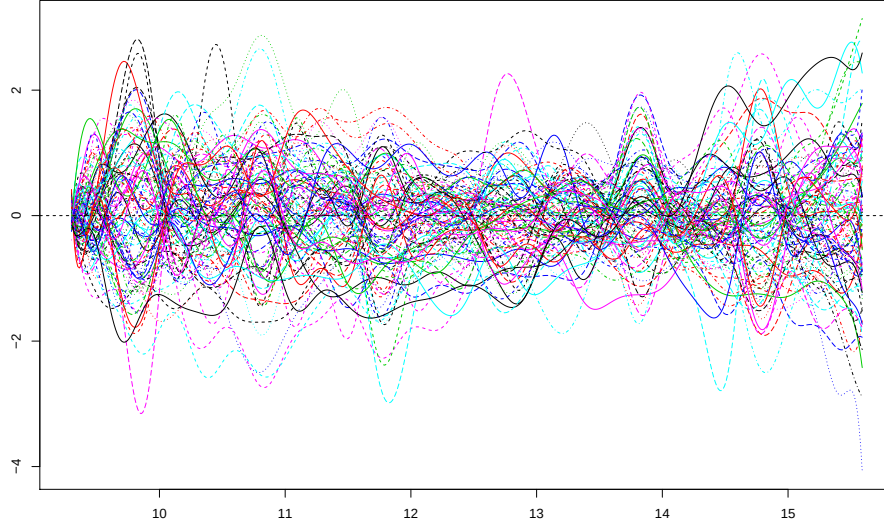


Figure F.3: Daily transformed cumulative log-return curves for Microsoft stock from 6/13/2001 to 11/07/2001. The x -axis gives clock time, the y -axis is proportional to percentage change.

d	TVE	$\tilde{k}_{100}^*$	calendar date
1	0.47	40	08/08/2001
2	0.68	58	09/04/2001
3	0.75	65	09/19/2001
4	0.81	61	09/07/2001
5	0.85	64	09/18/2001

Table F.1: Performance of the fPCA based method on the transformed Microsoft log-returns, where d denotes the number of fPCs used, TVE stands for total variation explained by these fPCs.

both a plot of $\hat{\delta}$ and a plot of

$$\pi_\ell = \frac{|\langle \hat{\delta}, \hat{\varphi}_\ell \rangle|^2}{\|\hat{\delta}\|^2}$$

against ℓ , for the latter plot noting that, by Parseval's identity, $\|\delta\|^2 = \sum_\ell |\langle \delta, \hat{\varphi}_\ell \rangle|^2$. Therefore the π_ℓ measure the proportion of the squared norm of $\hat{\delta}$ explained by the ℓ th sample eigenfunction. The plot clearly shows that the break is not captured by only a few eigen-directions, but that it is rather spread out. The situation is hence akin to the settings of the simulation study, where it was shown that the fully functional method has better accuracy for dating the break. The plot of the estimated break curve also reveals that the different risk behaviors before and after 09/18/2001 led to additional gains (for a positive sign of the corresponding score) in the last, say, 90 minutes of trading, thereby reverting the tendency for smaller additional losses observed earlier in the day.

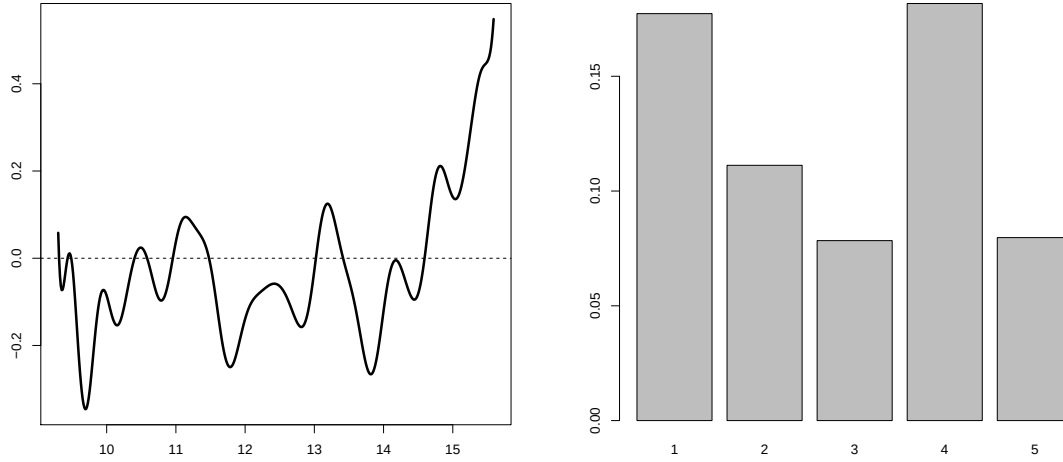


Figure F.4: Estimated break function $\hat{\delta}$ (left) and proportion of variation in $\hat{\delta}$ explained by the ℓ th sample eigenfunction (right) for the transformed cumulative Microsoft log-return curves.

References

- [1] Aue, A., Gabrys, R., Horváth, L. & P. Kokoszka (2009). Estimation of a change-point in the mean function of functional data. *Journal of Multivariate Analysis* **100**, 2254–2269.
- [2] Aue, A., Rice, G. & O. Sönmez (2017+). Detecting and dating structural breaks in functional data without dimension reduction. Preprint, University of California, Davis and University of Waterloo.
- [3] Berkes, I., Hörmann, S. & J. Schauer (2011). Split invariance principles for stationary processes. *The Annals of Probability* **39**, 2441–2473.
- [4] Bhattacharya, P. & P. Brockwell (1976). The minimum of an additive process with applications to signal estimation and storage theory. *Probability Theory and Related Fields* **37**, 51–75.
- [5] Buffet, E. (2003). On the time of the maximum of a Brownian motion with drift. *Journal of Applied Mathematics and Stochastic Analysis*, **16**, 201–207.
- [6] Horváth, L. Kokoszka, P. & R. Reeder (2013). Estimation of the mean of of functional time series and a two sample problem. *Journal of the Royal Statistical Society, Series B* **75**, 103–122.
- [7] Jirak, M. (2013). On weak invariance principles for sums of dependent random functionals. *Statistics & Probability Letters* **83**, 2291–2296.
- [8] Karatzas, I. & S.E. Shreve (1988). *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York.

- [9] Kim, J. & D. Pollard (1990). Cube root asymptotics. *The Annals of Statistics* **18**, 191–219.
- [10] Rice, G. & H.L. Shang (2017). A plug-in bandwidth selection procedure for long run covariance estimation with stationary functional time series. *Journal of Time Series Analysis* **38**, 591–609.
- [11] Stryhn, H. (1996). The location of the maximum of asymmetric two-sided Brownian motion with triangular drift. *Statistics & Probability Letters* **29**, 279–284.