Dating structural breaks in functional data
without dimension reduction

Alexander Aue† Gregory Rice‡ Ozan Sönmez†

November 12, 2015

Abstract

An estimator for the time of a break in the mean of stationary functional data is proposed that is fully functional in the sense that it does not rely on dimension reduction techniques such as functional principal component analysis (fPCA). A thorough asymptotic theory is developed for the estimator of the break date for fixed break size and shrinking break size. The main results highlight that the fully functional procedure performs best under conditions when analogous fPCA based estimators are at their worst, namely when the feature of interest is orthogonal to the leading principal components of the data. The theoretical findings are confirmed by means of a Monte Carlo simulation study in finite samples. An application to one-minute intra-day cumulative log-returns of Microsoft stock data highlights the practical relevance of the proposed fully functional procedure.

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


1 Introduction

This paper operates at the intersection of functional data analysis and structural breaks analysis for dependent observations. Functional data analysis (FDA) has seen an upsurge in research contributions in the past decade. These are documented, for example, in the comprehensive books by Ramsay and Silverman (2005) and Ferraty and Vieu (2010). Research concerned with structural breaks has a longstanding tradition in both the statistics and econometrics communities. Two recent reviews by Aue and Horváth (2013) and Horváth and Rice (2014) highlight newer developments, the first with a particular focus on time series.

†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

*This research was partially supported by NSF grants DMS 1209226, DMS 1305858 and DMS 1407530
Early work in functional structural break analysis dealt primarily with random samples of independent curves, the question of interest being whether all curves have a common mean function or whether there are two or more segments of the data that are homogeneous within but heterogeneous without. For example, Berkes et al. (2009) developed statistical methodology to test the null hypothesis of no structural break against the alternative of a (single) mean function change assuming that the error terms are independent and identically distributed curves. Aue et al. (2009) quantified the large-sample behavior of a break date estimator under a similar set of assumptions. The work in these two papers was generalized by Aston and Kirch (2012a, b) to include functional time series into the modeling framework. The methods developed in these papers were applied to temperature curves (Berkes et al., 2009) and functional magnetic resonance imaging (Aston and Kirch, 2012b). Other papers dealing with functional time series include Antoniadis and Sapatinas (2003) and Antoniadis et al. (2006), who predicted daily electricity demand curves, Aue et al. (2014, 2015), who studied daily particulate matter curves, and Besse et al. (2000), who dealt with functional climate variations.

Most of the procedures in the aforementioned papers are based on dimension reduction techniques, for example, using the widely popular functional principal components analysis (fPCA), by which the functional variation in the data is projected onto the directions of a small number of principal curves, and multivariate techniques are then applied on the resulting sequence of score vectors. This is also the case in functional structural break detection, in which after an initial fPCA step multivariate structural break theory is utilized. Despite the fact that functional data are, at least in principle, infinite dimensional, the state of the art in FDA remains to start the analysis with an initial dimension reduction procedure.

Dimension reduction approaches, however, automatically incur a loss of information, namely all information about the functional data that is orthogonal to the basis onto which it is projected. This weakness is easily illustrated in the context of detecting and dating structural breaks in the mean function: if the function representing the mean break is orthogonal to the basis used for dimension reduction, there cannot be a consistent test or estimator for the break date in that basis. This point will be further argued with a simple simulation example described in detail in Section 2 and in more complex numerical studies in Section 3.

The main purpose of this paper is then to develop methodology for the dating of structural breaks in functional data without the application of dimension reduction techniques, an idea that was touched upon in Horváth et al. (2013) in the context of stationarity tests for functional time series. Here, a fully functional estimator for the break date is proposed, and its asymptotic theory is comprehensively developed under the assumption that the model errors satisfy a general weak dependence condition. This theory illuminates a number of potential advantages of the fully functional estimator. As described before, when the direction of the break is orthogonal to the leading principal components of the data, the estimation of the mean break is asymptotically improved over fPCA based techniques. In addition, the assumptions required for the fully
functional theory are weaker than the ones used in Aue et al. (2009) and Aston and Kirch (2012a, b), as
convergence of the eigenvalues of the empirical covariance operator to the eigenvalues of the population
covariance operator do not have to be accounted for. These assumptions are typically formulated as finiteness
of fourth moment conditions. The relaxation obtained here may be particularly useful for applications to
intra-day financial data such as the one-minute log-returns on Microsoft stock discussed in Section 4. There,
it will be seen that the case of breaks appearing in eigenfunctions corresponding to smaller eigenvalues, that
is, non-obvious breaks, has practical relevance in assessing the risk associated with intra-day log-returns.

The paper is organized as follows. The estimator for the break date is introduced in Section 2, along with
the main asymptotic results of the paper. The asymptotic properties developed in this section are investigated
by means of a comprehensive simulation study in Section 3. Section 5 concludes. Proofs of the main results
are given in Section 6, utilizing some technical lemmas established in Appendix A.

2 Main results

In this paper, a simple functional data model allowing for a mean function break is considered. It is assumed
that the observations $X_1, \ldots, X_n$ are generated from the model

$$X_i = \mu + \delta 1\{i > k^*\} + \varepsilon_i, \quad i \in \mathbb{Z},$$

where $k^* = \lfloor \theta n \rfloor$ with $\theta \in (0, 1)$ labels the unknown time of the mean break parameterized in terms of the
sample size $n$, $\mu$ is the baseline mean function that is distorted by the addition of $\delta$ after the break time $k^*$, $1_A$
denotes the indicator function of the set $A$ and $\mathbb{Z}$ the set of integers. Each $X_i$ is a function defined without loss
of generality on the unit interval $[0, 1]$. The argument $t \in [0, 1]$ will be used to refer to a particular value $X_i(t)$
of the function $X_i$. Correspondingly, the quantities $\mu$, $\delta$ and $\varepsilon_i$ on the right-hand side of 2.1 are functions on
$[0, 1]$ as well. The innovations $(\varepsilon_i: i \in \mathbb{Z})$ are assumed to be centered stationary time series satisfying the
following conditions.

Assumption 2.1. It is assumed that

(a) there is a measurable function $g: S^\infty \rightarrow L^2[0, 1]$, where $S$ is a measurable space and independent,
identically distributed (iid) functional innovations $(\varepsilon_i: i \in \mathbb{Z})$ taking values in $S$ such that $\varepsilon_i = g(\varepsilon_i, \varepsilon_{i-1}, \ldots)$
for $i \in \mathbb{Z}$;

(b) there are $\ell$-dependent sequences $(\varepsilon_{i,\ell}: i \in \mathbb{Z})$ such that, for some $p > 2$,

$$\sum_{\ell=0}^\infty (\mathbb{E}[||\varepsilon_i - \varepsilon_{i,\ell}||^p])^{1/p} < \infty,$$

where $\varepsilon_{i,\ell} = g(\varepsilon_i, \ldots, \varepsilon_{i-\ell+1}, \varepsilon_{i-\ell,\ell-1}, \varepsilon_{i-\ell,\ell-1,\ell-1}, \ldots)$ with $\varepsilon_{i,\ell,j}^*$ being independent copies of $\varepsilon_{i,0}$ independent
of $(\varepsilon_i: i \in \mathbb{Z})$.  

Processes satisfying Assumption 2.1 were termed $L^p$-m-approximable processes by Hörmann and Kokoszka (2010), and cover most stationary functional time series models of interest, including functional ARMA processes (see Bosq, 2000). It is assumed that the underlying error innovations $(\epsilon_i : i \in \mathbb{Z})$ are elements of an arbitrary measurable space $S$. However, in many examples $S$ is itself a function space, and the evaluation of $g(\epsilon_i, \epsilon_{i-1}, \ldots)$ is a functional of $(\epsilon_j : j \leq i)$.

The main task when dealing with model (2.1) is to date the break time $k^*$. The proposed estimator is determined by the (scaled) functional cumulative sum (CUSUM) statistic

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

The superscript 0 indicates the tied-down nature of the CUSUM statistic, since $S_{n,0}^0 = S_{n,n}^0 = 0$ (interpreting an empty sum as zero). Noting that (2.2) tends to be large at the true break date, the estimator for $k^*$ is defined as

$$\hat{k}_n^* = \min \left\{ k : \| S_{n,k}^0 \| = \max_{1 \leq k' \leq n} \| S_{n,k'}^0 \| \right\},$$

where $\| \cdot \|$ denotes $L^2$-norm. The main results of this paper concern the large-sample behavior of $k_n^*$. Two cases are studied: the fixed break situation for which the break size is independent of the sample size, and the shrinking break situation for which the break size converges to zero at a certain rate. Details are given next, starting with the result for the fixed break case.

**Theorem 2.1.** If model (2.1) holds with $0 \neq \delta \in L^2[0, 1]$ and if Assumption 2.1 is satisfied, then

$$\hat{k}_n^* - k^* \overset{D}{\to} \min \left\{ k : P(k) = \sup_{k' \in \mathbb{Z}} P(k') \right\} \quad (n \to \infty),$$

where

$$P(k) = \left\{ \begin{array}{ll} (1 - \theta)\| \delta \|^2 k + \langle \delta, S_{\epsilon,k} \rangle, & k < 0, \\ -\theta\| \delta \|^2 k + \langle \delta, S_{\epsilon,k} \rangle, & k \geq 0, \end{array} \right.$$ (2.5)

with

$$S_{\epsilon,k} = \sum_{i=1}^{k} \epsilon_i + \sum_{i=-k}^{-1} \epsilon_i.$$

**Theorem 2.2.** If model (2.1) holds with $0 \neq \delta = \delta_n \in L^2[0, 1]$ such that $\| \delta_n \| \to 0$ but $n\| \delta_n \|^2 \to \infty$ and if Assumption 2.1 is satisfied, then

$$\| \delta_n \|^2 (\hat{k}_n^* - k^*) \overset{D}{\to} \inf \left\{ x : Q(x) = \sup_{x' \in \mathbb{R}} Q(x') \right\} \quad (n \to \infty),$$

where $\mathbb{R}$ denotes the real numbers and

$$Q(x) = \left\{ \begin{array}{ll} (1 - \theta)x + \sigma B(x), & x < 0, \\ -\theta x + \sigma B(x), & x \geq 0, \end{array} \right.$$ (2.6)
with \((B(x) : x \in \mathbb{R})\) a two-sided Brownian motion and

\[
\sigma^2 = \lim_{n \to \infty} \int \int C_\varepsilon(t, t') \delta_n(t) \delta_n(t') \frac{ddt'dt}{\|\delta_n\|^2},
\]

where \(C_\varepsilon(t, t')\) is the long-run covariance kernel of \((\varepsilon_i : i \in \mathbb{Z})\), namely

\[
C_\varepsilon(t, t') = \sum_{\ell = -\infty}^{\infty} \text{Cov}(\varepsilon_0(t), \varepsilon_{\ell}(t')).
\] (2.7)

An interesting consequence of Theorems 2.1 and 2.2 is that mean changes \(\delta\) that are orthogonal to the primary modes of variation in the data are asymptotically easier to detect. For example, if, under the conditions of Theorem 2.1, \(\delta\) is orthogonal to the error functions, then the stochastic term in the limit distribution vanishes. Moreover, if the functions \(\delta_n\) in Theorem 2.2 tend to align with eigenfunctions corresponding to smaller and smaller eigenvalues of the integral operator with kernel \(C_\varepsilon\), then \(\sigma^2\) tends to zero in the definition of \(Q(x)\). The proofs of Theorems 2.1 and 2.2 are given in Section 6.

The next result concerns the large-sample behavior of \(\hat{k}_n^*\) if no break is present in the data, that is if \(\delta = 0\) in (2.1). The asymptotics of the break date estimator for this case are quantified in the next theorem.

**Theorem 2.3.** If model (2.1) holds with \(\delta = 0\), so that \(X_i = \mu_i + \varepsilon_i\) for all \(i = 1, \ldots, n\), and if Assumption 2.1 is satisfied, then

\[
\frac{\hat{k}_n^*}{n} \overset{\mathbb{D}}{\to} \arg \max_{0 \leq x \leq 1} \|\Gamma^0(x, \cdot)\| \quad (n \to \infty),
\]

where \(\Gamma^0\) is a bivariate Gaussian process with mean zero and covariance function \(\mathbb{E}[\Gamma^0(x, t)\Gamma^0(x', t')] = \left(\min\{x, x'\} - xx'\right)C_\varepsilon(t, t')\).

The proof of Theorem 2.3 is provided in Section 6. Observe that the limiting distribution in Theorem 2.3 is non-pivotal, but it can be approximated via Monte Carlo simulations using an estimator of \(C_\varepsilon\). To see this note that, because of the Karhunen–Loeve representation, \(\Gamma^0(x, t)\) can be written in the form

\[
\sum_{\ell=1}^{\infty} \sqrt{\lambda_{\ell}} \varphi_{\ell}(t) B_\ell(x),
\]

where \((\lambda_{\ell}, \ell \in \mathbb{N})\) and \((\varphi_{\ell}, \ell \in \mathbb{N})\) are the eigenvalues and eigenfunctions of \(C_\varepsilon\) and \((B_\ell : \ell \in \mathbb{N})\) are independent standard Brownian bridges. Computing the norm as required for the limit in Theorem 2.3 yields that

\[
\arg \max_{x \in [0, 1]} \|\Gamma^0(x, \cdot)\| \overset{\mathbb{D}}{=} \arg \max_{x \in [0, 1]} \left(\sum_{\ell=1}^{\infty} \lambda_{\ell} B_\ell^2(x)\right)^{1/2}.
\]

Truncation of the sum under the square-root on the right-hand side gives then approximations to the theoretical limit. For practical purposes population eigenvalues have to be estimated from data.

In the remainder of this section, the fully functional results put forward in this paper are compared to their fPCA counterparts in Aue et al. (2009) and Aston and Kirch (2012a, b). This procedure utilizes the
eigenvalues, say \( \hat{\lambda}_1, \ldots, \hat{\lambda}_n \), and eigenfunctions, say \( \hat{\varphi}_1, \ldots, \hat{\varphi}_n \), of the sample covariance operator \( \hat{K} \) whose kernel is given by \( \hat{K}(t, t') = n^{-1} \sum_{i=1}^{n} [X_i(t) - \bar{X}_n(t)][X_i(t') - \bar{X}_n(t')] \). In the presence of a mean break as in (2.1), \( \hat{K}(t, t') \) is the empirical counterpart of the population covariance kernel

\[
 K(t, t') = K_0(t, t') + \theta(1 - \theta)\delta(t)\delta(t'),
\]

where \( K_0(t, t') = \mathbb{E}[\varepsilon_1(t)\varepsilon_1(t')] \) is the covariance kernel of the innovations \( (\varepsilon_i : i \in \mathbb{Z}) \). In particular, eigenvalues and eigenfunctions of \( \hat{K}(t, t') \) converge to those of \( K(t, t') \) under suitable assumptions that include the finiteness of the fourth moment \( \mathbb{E}[||\varepsilon_1||^4] \). Choosing a suitable \( d \in \{1, \ldots, n\} \) allows one to define the fPCA break point estimator

\[
 \hat{k}_n^* = \min \left\{ k : \hat{R}_{n,k} = \min_{1 \leq k' \leq n} R_{n,k'} \right\},
\]

where \( R_{n,k} = n^{-1} \hat{S}_{n,k}^{\top} \hat{S}_{n,k}, \hat{S}_{n,k} = \sum_{i=1}^{k} \hat{\xi}_i - kn^{-1} \sum_{i=1}^{n} \hat{\xi}_i \) and \( \hat{\xi}_i = (\hat{\xi}_{i,1}, \ldots, \hat{\xi}_{i,d})^T \) with fPCA scores \( \hat{\xi}_{i,t} = \langle X_i - \bar{X}_n, \hat{\varphi}_t \rangle \), and \( \hat{\Sigma}_n = \text{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_d) \). For the independent case, the counterparts of Theorem 2.1 and 2.2 were obtained in Aue et al. (2009). Aston and Kirch (2012a) showed the consistency of \( \hat{k}_n^* \) in the time series case. The performance of \( \hat{k}_n^* \) depends crucially on the selection of \( d \).

To highlight the differences in the performance of \( \hat{k}_n^* \) and \( \hat{k}_n^* \), consider the following simple example (see Remark 4.1 of Aston and Kirch, 2012a). Let \( b_1 \) and \( b_2 \) be two orthogonal \( B \)-spline functions (see the upper left panel in Figure 2.1) and let \( \xi_1, \ldots, \xi_{100} \) and \( \eta_1, \ldots, \eta_{100} \) be two independent sequences of independent standard normal random variables. Define then the functional innovations

\[
 \varepsilon_i = 2\xi_ib_1 + \eta_ib_2, \quad i = 1, \ldots, 100,
\]

and insert the break function \( \delta_c = cb_2 \) indexed by a sensitivity parameter \( c > 0 \) at the break date \( k^* \). This gives rise to the functional observations \( Y_i = \delta \mathbb{I}\{i > k^*\} + \varepsilon_i, i = 1, \ldots, 100 \), following (2.1) with \( \mu = 0 \). A straightforward calculation shows that, for this example, \( K(t, t') = 4b_1(t)b_1(t') + [1 + c^2\theta(1 - \theta)]b_2(t)b_2(t') \), where \( \theta = k^*/100 \). Choose further \( k^* = 50 \), so that \( \theta = .5 \), and \( c = 2, 4 \). Suppose simulations are run 1000 times and both methods are applied with the fPCA method using \( d = 1 \). For the case \( c = 2 \), the leading eigenfunction is \( b_1 \), but the break is in the direction of \( b_2 \). Hence, the fPCA based method is expected to have difficulty dating the break. For the case \( c = 4, b_2 \) becomes the leading eigenfunction and fPCA is expected to perform better. These theoretical findings are corroborated in Figure 2.1, where the superiority of the fully functional methodology for the \( c = 2 \) case becomes apparent, while the performance of both methods becomes more comparable for \( c = 4 \).

The use of the fully functional approach to dating break points is therefore especially advantageous in the interesting case of breaks that are sizable but not obvious in the sense that their influence does not show up in the directions of the leading principal components of the data. The next section follows up on these observations in more complex settings.
3 Simulation Study

3.1 Setting

Following the set-up in Aue et al. (2015), functional data of size $n = 100$ were generated using $D = 21$ B-spline basis functions $v_1, \ldots, v_D$ on the unit interval $[0, 1]$. Without loss of generality, the initial mean curve $\mu$ in 2.1 is assumed to be the zero function. Independent curves were then generated according to

$$\zeta_i = \sum_{\ell=1}^{D} N_{i,\ell} v_{\ell},$$

where the $N_{i,\ell}$ are independent normal random variables with standard deviations $\sigma_{\ell}$ chosen from the following two options:

- Fast setting: $\sigma_{\text{fast}} = (1.2^{-\ell}; \ell = 1, \ldots, D)$;
- Slow setting: $\sigma_{\text{slow}} = (\ell^{-1}; \ell = 1, \ldots, D)$.

The fast setting describes functional data for which the eigenvalues of the covariance operator decay quickly, at a geometric rate, while in the slow setting the eigenvalues decay at only a polynomial rate.

To explore the effect of temporal dependence on the break point estimators, second-order functional autoregressive curves $\varepsilon_i = \Psi_1 \varepsilon_{i-1} + \Psi_2 \varepsilon_{i-2} + \zeta_i$, $i = 1, \ldots, 100$, were generated (with a burn-in period of 100 curves). The operators were set up as $\Psi_j = \kappa_j \Psi$, $j = 1, 2$, where the random operator $\Psi$ is represented by a $D \times D$ matrix whose entries consist of independent, centered normal random variables with standard deviations given by $\sigma_{\text{fast}}' \sigma_{\text{fast}}$ or $\sigma_{\text{slow}}' \sigma_{\text{slow}}$. A scaling was applied to achieve $||\Psi|| = 1$. The constants $\kappa_1$ and $\kappa_2$ can then be used to adjust the strength of the temporal dependence. To ensure stationarity of the time series, $|\kappa_1| + |\kappa_2| < 1$ is used.

Two types of breaks are considered, namely a prominent break affecting directions specified by larger eigenvalues of the data covariance operator, and a moderate break affecting directions corresponding to smaller eigenvalues. More specifically,

- Prominent break: $\delta = c \sum_{\ell=2}^{4} v_{\ell}$;
- Moderate break: $\delta = c \sum_{\ell=4}^{6} v_{\ell}$.

The positive constant $c$ can be used to adjust the break impact. The prominent break case should be more favorable to the fPCA based procedure as the break occurs in the directions used for dimension reduction. If the magnitude of $c$ is small enough, the moderate break date should be more favorable to the proposed fully functional procedure.

Combining the previous paragraphs, functional curves $y_i = \delta \mathbb{I}\{i > k^*\} + \varepsilon_i$ according to (2.1) were generated for $k^* = 25, 50$ and 75. Both the fully functional procedure and its fPCA counterpart were applied...
to a variety of settings, with outcomes reported in the subsequent sections. The fPCA based method was applied choosing $d$ such that the total variation explained exceeded 75%. All results are based on 1000 runs of the simulation experiments.

### 3.2 Functional autoregressions

In the following, FAR(2) processes $\varepsilon_i = \Psi_1 \varepsilon_{i-1} + \Psi_2 \varepsilon_{i-2} + \zeta_i$, $\Psi_j = \kappa_j \Psi$ with $\Psi$ as outlined in the previous section are considered, setting $\kappa_1 = 0.5$ and $\kappa_2 = 0.3$ to ensure stationarity. These processes are then subjected to prominent and moderate breaks.

The results for the prominent break case are summarized in Table 3.1. Combining the prominent break with $\sigma_{\text{fast}}$ should be the most favorable situation for the fPCA based method. The table shows that here both methods perform roughly similar in terms of measures of centrality (mean and median). Measures of dispersion (MSE and MAE), however, show that the fully functional procedure tends to produce smaller variations. This is in particular visible for $k^* = 50$. Both methods struggle to identify breaks away from the center of the data sequence at the low and medium sensitivity levels $c = 0.5$ and $c = 0.8$. Results are similar for $\sigma_{\text{slow}}$ under a prominent break. Here, the situation becomes more advantageous for the proposed method as it adjusts better to $k^* = 25$ and $k^* = 75$ for $c = 0.8$. Overall, it can be seen that the fully functional method is at least competitive with its fPCA counterpart.

The results for the moderate break case are summarized in Table 3.2. Here differences in performance become more apparent. For example, it can be seen that the fPCA based method still has difficulty in dating the breaks away from the center at the high sensitivity setting $c = 1.5$, for which the proposed method does much better. In general, the fully functional procedure is associated with smaller variation as evidenced in the reported measures of dispersion. Overall, the proposed method performs better than the fPCA counterpart in the moderate break case.

Figure 3.1 highlights the shapes of histograms of the two competing estimators for the break date $k^* = 50$ at various sensitivity levels $c$ for the prominent break in the fast setting. It is nicely seen that there is less dispersion in the histograms of the fully functional procedure. Figure 3.2 shows box plots of both break point estimators for the break date $k^* = 25$ at a progression of $c$ values ranging from 0.5 to 1.5 in the case of a prominent break in the fast setting and of a moderate break in the slow setting. Here the proposed procedure is seen to suffer less from both bias and variance. While the latter situation is by design preferable for the fully functional method, it is worthwhile noting that the estimation of the eigenstructure adds significant noise into the fPCA procedure even in situations for which the break appears in the dominant directions.
Table 3.1: Summary statistics for the fPCA and fully functional break point estimators in the prominent break case for the two eigenvalue settings (slow/fast decay), different break locations \( k^* \) and break magnitudes indexed by \( c \).

<table>
<thead>
<tr>
<th>( k^* )</th>
<th>( c )</th>
<th>method</th>
<th>mean</th>
<th>median</th>
<th>MSE</th>
<th>MAE</th>
<th>mean</th>
<th>median</th>
<th>MSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.5</td>
<td>fPCA</td>
<td>47.03</td>
<td>45.00</td>
<td>29.55</td>
<td>23.33</td>
<td>48.98</td>
<td>48.00</td>
<td>30.38</td>
<td>25.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>proposed</td>
<td>45.16</td>
<td>44.00</td>
<td>26.08</td>
<td>20.66</td>
<td>45.73</td>
<td>44.00</td>
<td>27.12</td>
<td>21.65</td>
</tr>
<tr>
<td>0.8</td>
<td></td>
<td>fPCA</td>
<td>41.03</td>
<td>35.00</td>
<td>24.68</td>
<td>17.09</td>
<td>48.25</td>
<td>47.00</td>
<td>29.86</td>
<td>24.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td>proposed</td>
<td>38.44</td>
<td>34.00</td>
<td>19.81</td>
<td>13.82</td>
<td>36.71</td>
<td>32.00</td>
<td>17.96</td>
<td>12.38</td>
</tr>
<tr>
<td>1.5</td>
<td></td>
<td>fPCA</td>
<td>29.36</td>
<td>25.00</td>
<td>9.61</td>
<td>4.49</td>
<td>25.05</td>
<td>25.00</td>
<td>0.25</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>proposed</td>
<td>29.28</td>
<td>25.00</td>
<td>8.71</td>
<td>4.32</td>
<td>25.05</td>
<td>25.00</td>
<td>0.23</td>
<td>0.05</td>
</tr>
<tr>
<td>50</td>
<td>0.5</td>
<td>fPCA</td>
<td>50.21</td>
<td>50.00</td>
<td>19.17</td>
<td>15.28</td>
<td>50.39</td>
<td>51.00</td>
<td>17.99</td>
<td>14.92</td>
</tr>
<tr>
<td></td>
<td></td>
<td>proposed</td>
<td>49.75</td>
<td>50.00</td>
<td>3.56</td>
<td>2.06</td>
<td>51.05</td>
<td>51.00</td>
<td>13.70</td>
<td>10.49</td>
</tr>
<tr>
<td>0.8</td>
<td></td>
<td>fPCA</td>
<td>49.88</td>
<td>50.00</td>
<td>10.91</td>
<td>6.31</td>
<td>50.23</td>
<td>50.00</td>
<td>17.04</td>
<td>13.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>proposed</td>
<td>49.98</td>
<td>50.00</td>
<td>1.41</td>
<td>0.74</td>
<td>50.30</td>
<td>50.00</td>
<td>5.66</td>
<td>3.31</td>
</tr>
<tr>
<td>1.5</td>
<td></td>
<td>fPCA</td>
<td>49.99</td>
<td>50.00</td>
<td>0.24</td>
<td>0.08</td>
<td>50.00</td>
<td>50.00</td>
<td>0.03</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>proposed</td>
<td>50.00</td>
<td>50.00</td>
<td>0.22</td>
<td>0.07</td>
<td>50.00</td>
<td>50.00</td>
<td>0.03</td>
<td>0.00</td>
</tr>
<tr>
<td>75</td>
<td>0.5</td>
<td>fPCA</td>
<td>54.41</td>
<td>56.00</td>
<td>28.52</td>
<td>22.34</td>
<td>51.42</td>
<td>52.00</td>
<td>29.93</td>
<td>24.92</td>
</tr>
<tr>
<td></td>
<td></td>
<td>proposed</td>
<td>56.15</td>
<td>58.00</td>
<td>24.80</td>
<td>19.42</td>
<td>54.99</td>
<td>56.00</td>
<td>26.18</td>
<td>20.98</td>
</tr>
<tr>
<td>0.8</td>
<td></td>
<td>fPCA</td>
<td>59.34</td>
<td>65.00</td>
<td>23.42</td>
<td>16.75</td>
<td>52.12</td>
<td>53.00</td>
<td>29.81</td>
<td>24.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>proposed</td>
<td>61.47</td>
<td>66.00</td>
<td>19.75</td>
<td>14.03</td>
<td>63.47</td>
<td>69.00</td>
<td>18.06</td>
<td>12.10</td>
</tr>
<tr>
<td>1.5</td>
<td></td>
<td>fPCA</td>
<td>70.84</td>
<td>75.00</td>
<td>9.32</td>
<td>4.32</td>
<td>74.95</td>
<td>75.00</td>
<td>0.26</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>proposed</td>
<td>70.84</td>
<td>75.00</td>
<td>8.57</td>
<td>4.21</td>
<td>74.95</td>
<td>75.00</td>
<td>0.26</td>
<td>0.05</td>
</tr>
</tbody>
</table>

3.3 Heavy tails

The heavy tail case is only considered here for independent curves (choosing \( \kappa_1 = \kappa_2 = 0 \)) in the moderate break case with fast decay of eigenvalues of the innovations. The case with slower decay of the eigenvalues produces results more in favor of the proposed method. Instead of the normal distributions specified in Section 3.1, \( \zeta_1, \ldots, \zeta_n \) were chosen \( t \)-distributed with 4 degrees of freedom and \( \varepsilon_1, \ldots, \varepsilon_{100} \) were defined accordingly. Modifications of the simulation settings presented in this section could potentially be useful for applications to intra-day financial data. Due to the reduced number of finite moments in this setting, the fPCA based procedure is not theoretically justified.

Results in Table 3.3 are given for \( k^* = 25, 50, 75 \) and two choices of break sensitivity \( c = 0.1, 0.5 \). The summary statistics show the proposed method to be superior if the break is small (\( c = 0.1 \)) and if \( k^* \) is removed from the middle of the sample (\( k^* = 25, 75 \)). Measures of centrality for \( k^* = 50 \) are similar for both procedures, but differences become evident in the reported values for both MSE and MAE, with the proposed method performing significantly better. For the larger break (\( c = 0.5 \)) both methods behave almost identical. The proposed method looks in general more favorable in the heavy-tail case than in the time series case of the
Table 3.2: Same as in Table 3.1 but for the moderate break case.

<table>
<thead>
<tr>
<th>$k^*$</th>
<th>$c$</th>
<th>method</th>
<th>mean</th>
<th>median</th>
<th>MSE</th>
<th>MAE</th>
<th>mean</th>
<th>median</th>
<th>MSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.5</td>
<td>fPCA</td>
<td>49.25</td>
<td>49.00</td>
<td>31.38</td>
<td>25.53</td>
<td>50.73</td>
<td>51.00</td>
<td>32.76</td>
<td>27.23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>proposed</td>
<td>44.57</td>
<td>43.00</td>
<td>25.00</td>
<td>19.91</td>
<td>43.15</td>
<td>41.00</td>
<td>24.52</td>
<td>18.79</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>fPCA</td>
<td>45.98</td>
<td>44.00</td>
<td>28.69</td>
<td>22.40</td>
<td>46.48</td>
<td>45.00</td>
<td>29.52</td>
<td>23.37</td>
</tr>
<tr>
<td></td>
<td></td>
<td>proposed</td>
<td>37.21</td>
<td>32.00</td>
<td>18.51</td>
<td>12.46</td>
<td>33.63</td>
<td>28.00</td>
<td>14.89</td>
<td>8.80</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>fPCA</td>
<td>29.96</td>
<td>25.00</td>
<td>11.78</td>
<td>5.24</td>
<td>37.33</td>
<td>29.00</td>
<td>21.26</td>
<td>13.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>proposed</td>
<td>27.64</td>
<td>25.00</td>
<td>6.67</td>
<td>2.64</td>
<td>26.14</td>
<td>25.00</td>
<td>4.16</td>
<td>1.14</td>
</tr>
<tr>
<td>50</td>
<td>0.5</td>
<td>fPCA</td>
<td>50.06</td>
<td>50.00</td>
<td>21.05</td>
<td>17.51</td>
<td>50.91</td>
<td>51.00</td>
<td>19.60</td>
<td>16.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>proposed</td>
<td>50.02</td>
<td>50.00</td>
<td>2.59</td>
<td>1.37</td>
<td>50.52</td>
<td>50.00</td>
<td>9.82</td>
<td>6.45</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>fPCA</td>
<td>48.97</td>
<td>50.00</td>
<td>18.05</td>
<td>13.88</td>
<td>50.80</td>
<td>50.00</td>
<td>18.28</td>
<td>14.65</td>
</tr>
<tr>
<td></td>
<td></td>
<td>proposed</td>
<td>50.00</td>
<td>50.00</td>
<td>0.58</td>
<td>0.24</td>
<td>50.29</td>
<td>50.00</td>
<td>4.61</td>
<td>1.89</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>fPCA</td>
<td>50.00</td>
<td>50.00</td>
<td>0.12</td>
<td>0.02</td>
<td>50.17</td>
<td>50.00</td>
<td>7.00</td>
<td>2.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td>proposed</td>
<td>50.00</td>
<td>50.00</td>
<td>0.08</td>
<td>0.01</td>
<td>50.01</td>
<td>50.00</td>
<td>0.17</td>
<td>0.01</td>
</tr>
<tr>
<td>75</td>
<td>0.5</td>
<td>fPCA</td>
<td>52.25</td>
<td>52.00</td>
<td>30.18</td>
<td>24.77</td>
<td>50.74</td>
<td>52.00</td>
<td>31.88</td>
<td>26.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>proposed</td>
<td>56.51</td>
<td>58.00</td>
<td>24.22</td>
<td>19.17</td>
<td>58.21</td>
<td>62.00</td>
<td>23.29</td>
<td>17.37</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>fPCA</td>
<td>55.70</td>
<td>59.00</td>
<td>26.98</td>
<td>20.93</td>
<td>53.07</td>
<td>54.50</td>
<td>29.30</td>
<td>23.72</td>
</tr>
<tr>
<td></td>
<td></td>
<td>proposed</td>
<td>62.88</td>
<td>67.00</td>
<td>17.99</td>
<td>12.41</td>
<td>65.75</td>
<td>72.00</td>
<td>15.89</td>
<td>9.44</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>fPCA</td>
<td>70.85</td>
<td>75.00</td>
<td>9.76</td>
<td>4.39</td>
<td>64.38</td>
<td>72.00</td>
<td>19.42</td>
<td>11.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>proposed</td>
<td>72.91</td>
<td>75.00</td>
<td>5.25</td>
<td>2.09</td>
<td>73.94</td>
<td>75.00</td>
<td>4.21</td>
<td>1.06</td>
</tr>
</tbody>
</table>

Table 3.3: Summary statistics for the fPCA and fully functional break point estimators for different break locations $k^*$ and break magnitudes indexed by $c$.

<table>
<thead>
<tr>
<th>$k^*$</th>
<th>$c$</th>
<th>method</th>
<th>mean</th>
<th>median</th>
<th>MSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.1</td>
<td>fPCA</td>
<td>48.25</td>
<td>47.00</td>
<td>29.86</td>
<td>24.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td>proposed</td>
<td>36.71</td>
<td>32.00</td>
<td>17.96</td>
<td>12.38</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>fPCA</td>
<td>25.05</td>
<td>25.00</td>
<td>0.25</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>proposed</td>
<td>25.05</td>
<td>25.00</td>
<td>0.23</td>
<td>0.05</td>
</tr>
<tr>
<td>50</td>
<td>0.1</td>
<td>fPCA</td>
<td>50.23</td>
<td>50.00</td>
<td>17.04</td>
<td>13.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>proposed</td>
<td>50.30</td>
<td>50.00</td>
<td>5.66</td>
<td>3.31</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>fPCA</td>
<td>50.00</td>
<td>50.00</td>
<td>0.03</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>proposed</td>
<td>50.00</td>
<td>50.00</td>
<td>0.03</td>
<td>0.00</td>
</tr>
<tr>
<td>75</td>
<td>0.1</td>
<td>fPCA</td>
<td>52.12</td>
<td>53.00</td>
<td>29.81</td>
<td>24.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>proposed</td>
<td>63.47</td>
<td>69.00</td>
<td>18.06</td>
<td>12.10</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>fPCA</td>
<td>74.95</td>
<td>75.00</td>
<td>0.26</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>proposed</td>
<td>74.95</td>
<td>75.00</td>
<td>0.26</td>
<td>0.05</td>
</tr>
</tbody>
</table>

previous section due to the deteriorated performance in estimating eigenvalues and eigenfunctions.

Figure 3.3 shows histograms for $c = 0.5$ and three different degrees of freedom for the $t$-innovations. It can be seen that in all cases the fPCA based procedure fails to produce reasonable results. The performance is
worst for $df = 2$ and somewhat comparable for $df = 3$ and $df = 4$. The proposed method is seen to work for the latter two cases but it’s performance deteriorates for $df = 2$, a situation that is not theoretically justified.

## 4 Application

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 $i = 1, \ldots, 100$. The (scaled) cumulative intra-day returns were then computed as

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

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 4.1. 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 4.2 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 a 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

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

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 between 1 and 5 (corresponding to the second to fifth sample eigenfunctions of the original data) are summarized in Table 4.1. For any $d > 5$, the estimated change is 09/18/2001.
Table 4.1: 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.

<table>
<thead>
<tr>
<th>$d$</th>
<th>TVE</th>
<th>$k^*_1$</th>
<th>calendar date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.47</td>
<td>40</td>
<td>08/08/2001</td>
</tr>
<tr>
<td>2</td>
<td>0.68</td>
<td>58</td>
<td>09/04/2001</td>
</tr>
<tr>
<td>3</td>
<td>0.75</td>
<td>65</td>
<td>09/19/2001</td>
</tr>
<tr>
<td>4</td>
<td>0.81</td>
<td>61</td>
<td>09/07/2001</td>
</tr>
<tr>
<td>5</td>
<td>0.85</td>
<td>64</td>
<td>09/18/2001</td>
</tr>
</tbody>
</table>

Using $k^*_1 = 64$ 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 4.4 shows 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 $\ell$, for the latter plot noting that, by Parseval’s identity, $||\delta||^2 = \sum_\ell |\langle \delta, \hat{\varphi}_\ell \rangle|^2$. Therefore the $\pi_\ell$ measure the proportion of the squared norm of $\hat{\delta}$ explained by the $\ell$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.

5 Conclusions

In this paper, a fully functional methodology was introduced to date mean curve breaks for functional data. The assumptions made allow for time series specifications of the curves and are formulated using the optimal rates for approximations of the data with $\ell$-dependent sequences. The assumptions are notably weaker than those usually made in the fPCA context and include heavy-tailed functional observations, making the asymptotic theory developed here widely applicable. In a comprehensive simulation study it is shown that the fully functional method tends to perform better than its fPCA counterpart, with significant performance gains for breaks that do not align well with the directions specified by the largest (few) eigenvalue(s) of the data covariance operator. It is shown in an application to intra-day log-return data of Microsoft stock that such a situation is of practical relevance. The Microsoft example is by no means a special case, as other stocks such as IBM show similar behavior. (These results are not reported here, but are available upon request.) It is hoped
that the proposed methodology will find widespread use in the future. More generally, this work provides an in-depth study in a specific context of the overarching principle that whenever the signal of interest is not dominant or is “sparse”, in the sense that it is not entirely contained in the leading principal components, then alternatives to dimension reduction based methods should be considered and are likely more effective.

6 Proofs

The proofs of Theorems 2.1 and 2.2 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.

6.1 A useful decomposition

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

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

where

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

A standard calculation common in the analysis of structural breaks (see, for example, Csörgő and Horváth, 1997), using the definitions of $S_{n,k}^0$ and the model equation (2.1), yields that, for $1 \leq k < k^*$,

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

(6.1)

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

$$D_k^{(1)} = -(k - k^*) \frac{n - k^*}{n} \delta$$

and

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

are deterministic drift functions and

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

and

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

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

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

(6.2)
into respective drift and random functions

\[ D^{(3)}_k = -\left(k - k^\ast\right) \frac{k^\ast}{n} \delta \quad \text{and} \quad D^{(4)}_k = -(2n - k - k^\ast) \frac{k^\ast}{n} \delta, \]

\[ E^{(3)}_k = -\sum_{i=k^\ast+1}^{k} \varepsilon_i - \frac{k - k^\ast}{n} \sum_{i=1}^{n} \varepsilon_i \quad \text{and} \quad E^{(4)}_k = \sum_{i=1}^{k} \varepsilon_i + \sum_{i=1}^{k^\ast} \varepsilon_i - \frac{k + k^\ast}{n} \sum_{i=1}^{n} \varepsilon_i. \]

In the proofs below, the focus is on the asymptotics for \( R_{n,k} \) when \( 1 \leq k < k^\ast \) using (6.1), noting that similar arguments may be applied when \( k^\ast < k \leq n \) using (6.2). Throughout the proofs, for \( j \in \mathbb{N}_0 \), \( c_j \) denote positive absolute constants.

### 6.2 Proof of Theorem 2.1

**Lemma 6.1.** If the assumptions of Theorem 2.1 are satisfied, \( |\hat{k}_n^\ast - k^\ast| \) is bounded in probability.

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

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

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

\[
\max_{1 \leq k \leq k^\ast - N} \frac{1}{n} \left( D^{(1)}_k, D^{(2)}_k \right) = \max_{1 \leq k \leq k^\ast - N} (k - k^\ast) \frac{k + k^\ast}{n} \left( \frac{n - k^\ast}{n} \right)^2 \|\delta\|^2
\]

\[
= -2N\theta(1 - \theta)^2\|\delta\|^2 + o(1).
\]

Therefore,

\[
\lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} \max_{1 \leq k \leq k^\ast - N} \frac{1}{n} \left( D^{(1)}_k, D^{(2)}_k \right) = -\infty. \tag{6.3}
\]

In the following, it will be shown that the deterministic component whose asymptotics is established in (6.3) is the dominant term on the right hand side of (6.1). In this direction, it is first established that, for all \( x > 0 \),

\[
\lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} P\left( \max_{1 \leq k \leq k^\ast - N} \frac{\langle E^{(1)}_k, E^{(2)}_k \rangle}{\langle D^{(1)}_k, D^{(2)}_k \rangle} > x \right) = 0. \tag{6.4}
\]

Note that, for all \( n \geq 1 \) and \( 1 \leq k < k^\ast \),

\[
\left( D^{(1)}_k, D^{(2)}_k \right) = n(k^\ast - k) \frac{k + k^\ast}{n} \frac{n - k^\ast}{n} \|\delta\|^2 \geq c_0 n(k^\ast - k)
\]

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

\[
\frac{\langle E^{(1)}_k, E^{(2)}_k \rangle}{\langle D^{(1)}_k, D^{(2)}_k \rangle} \leq c_1 \frac{\langle E^{(1)}_k, E^{(2)}_k \rangle}{n(k^\ast - k)} \leq c_1 \frac{\|E^{(1)}_k\| \|E^{(2)}_k\|}{n(k^\ast - k)}.
\]
Taking the maximum on right- and left-hand side leads to
\[
\max_{1 \leq k \leq k^* - N} \frac{\langle E_k^{(1)}, E_k^{(2)} \rangle}{k^* - k} \leq c_2 \max_{1 \leq k \leq k^* - N} \left\| E_k^{(1)} \right\| \max_{1 \leq k \leq k^* - N} \frac{\left\| E_k^{(2)} \right\|}{n}
\]
The definition of $E_k^{(1)}$ and the triangle inequality give
\[
\max_{1 \leq k \leq k^* - N} \frac{\left\| E_k^{(1)} \right\|}{k^* - k} \leq \max_{1 \leq k \leq k^* - N} \frac{1}{k^* - k} \left\| \sum_{i=k+1}^{k^*} \varepsilon_i \right\| + \frac{1}{n} \left\| n \varepsilon_i \right\|
\]  \hspace{1cm} (6.5)

Now, for all $\alpha \in (1/2, 1),$
\[
\max_{1 \leq k \leq k^* - N} \frac{1}{k^* - k} \left\| \sum_{i=k+1}^{k^*} \varepsilon_i \right\| \overset{\text{def}}{=} \max_{N \leq k \leq k^*} \frac{1}{k} \left\| \sum_{i=1}^{k} \varepsilon_i \right\| \leq \frac{1}{N^{1-\alpha}} \sup_{k \geq 1} \frac{1}{k^\alpha} \left\| \sum_{i=1}^{k} \varepsilon_i \right\|
\]

By Lemma A.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 (6.5), it follows that
\[
\lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} P \left( \max_{1 \leq k \leq k^* - N} \frac{\left\| E_k^{(1)} \right\|}{k^* - k} > x \right) = 0.
\]

One may apply similar arguments to show that
\[
\lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} P \left( \max_{1 \leq k \leq k^* - N} \frac{\left\| E_k^{(2)} \right\|}{n} > x \right) = 0,
\]
which yields (6.4). Similar arguments also imply that
\[
\lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} P \left( \max_{1 \leq k \leq k^* - N} \frac{\langle E_k^{(1)}, D_k^{(2)} \rangle}{\langle D_k^{(1)}, D_k^{(2)} \rangle} > x \right) = 0,
\]  \hspace{1cm} (6.6)

and
\[
\lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} P \left( \max_{1 \leq k \leq k^* - N} \frac{\langle D_k^{(1)}, E_k^{(2)} \rangle}{\langle D_k^{(1)}, D_k^{(2)} \rangle} > x \right) = 0.
\]  \hspace{1cm} (6.7)

Combining (6.4), (6.6) and (6.7) with (6.1) yields that, for all $M < 0,$
\[
\lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} P \left( \max_{1 \leq k \leq k^* - N} R_{n,k} > M \right) = 0.
\]  \hspace{1cm} (6.8)

It may be proven analogously using $D_k^{(3)}, D_k^{(4)}, E_k^{(3)}, E_k^{(4)}$ and (6.2) that
\[
\lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} P \left( \max_{k^* + N \leq k \leq n} R_{n,k} > M \right) = 0.
\]  \hspace{1cm} (6.9)

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, (6.8) and (6.9) 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.
Lemma 6.2. If the assumptions of Theorem 2.1 are satisfied, then

\[(R_{n,k^*+k}: k \in [-N, N]) \overset{\mathbb{P}}{\to} (2\theta(1-\theta)P(k): k \in [-N, N])\]

for any \(N \geq 1\) as \(n \to \infty\).

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

\[
\max_{k^*-N \leq k \leq k^*} \left| \frac{1}{n} \left< D_k^{(1)}, D_k^{(2)} \right> - 2\theta(1-\theta)^2 \|\delta\|^2(k-k^*) \right| \leq 2\theta^2(1-\theta)\|\delta\|^2(k-k^*) = o(1),
\]

as \(n \to \infty\). One may obtain similarly that

\[
\max_{k^* \leq k \leq k^*+N} \left| \frac{1}{n} \left< D_k^{(3)}, D_k^{(4)} \right> - [ - 2\theta^2(1-\theta)\|\delta\|^2(k-k^*) ] \right| = o(1).
\]

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

\[
\frac{1}{n} \left< E_k^{(1)}, D_k^{(2)} \right> = \frac{n-k^*}{n} \frac{k^*}{n} \sum_{i=k+1}^{k^*} \langle \epsilon_i, \delta \rangle + \frac{n-k^*}{n} \frac{k^*}{n} \sum_{i=1}^{n} \langle \epsilon_i, \delta \rangle.
\]

The first term on the right-hand side of (6.12) can be estimated as

\[
\max_{k^*-N \leq k \leq k^*} \left| \frac{n-k^*}{n} \frac{k^*}{n} \sum_{i=k+1}^{k^*} \langle \epsilon_i, \delta \rangle - 2(1-\theta)\|\delta\|^2(k-k^*) \right| = o_P(1).
\]

For the second term on the right-hand side of (6.12), Fubini’s theorem shows that \(\mathbb{E}[\langle \epsilon_i, \delta \rangle] = 0\), and hence by the Ergodic Theorem, \((1/n) \sum_{i=1}^{n} \langle \epsilon_i, \delta \rangle \to 0\) as \(n \to \infty\) with probability one. Consequently,

\[
\max_{k^*-N \leq k \leq k^*} \left| \frac{n-k^*}{n} \frac{k^*}{n} (k-k^*)^{1/n} \sum_{i=1}^{n} \langle \epsilon_i, \delta \rangle \right| = o_P(1).
\]

Taken together, (6.12)–(6.14) imply that

\[
\max_{k^*-N \leq k \leq k^*} \left| \frac{1}{n} \left< E_k^{(1)}, D_k^{(2)} \right> - 2(1-\theta)\|\delta\|^2 \sum_{i=k+1}^{k^*} \langle \epsilon_i, \delta \rangle \right| = o_P(1).
\]
Similar arguments lead to
\[
\max_{k^* - N \leq k \leq k^* + N} \left| \frac{1}{n} \left\langle E_k^{(1)}, D_k^{(2)} \right\rangle - 2(1 - \theta) \sum_{i=k^* + 1}^{k} \langle \varepsilon_i, \delta \rangle \right| = o_P(1). \tag{6.16}
\]
It remains to be show that the other four terms on the concluding lines of (6.1) and (6.2) do not contribute asymptotically. To this end, by the Cauchy–Schwarz inequality,
\[
\max_{k^* - N \leq k \leq k^*} \left| \frac{1}{n} \left\langle E_k^{(1)}, E_k^{(2)} \right\rangle \right| \leq \max_{k^* - N \leq k \leq k^*} \frac{1}{n} \| E_k^{(1)} \| \| E_k^{(2)} \|. \tag{6.17}
\]

The triangle inequality yields that
\[
\max_{k^* - N \leq k \leq k^*} \| E_k^{(1)} \| \leq \max_{k^* - N \leq k \leq k^*} \left\| \sum_{i=k^* + 1}^{k} \varepsilon_i \right\| + \frac{N}{n} \left\| \sum_{i=1}^{n} \varepsilon_i \right\| = O_P(1) + o_P(1). \tag{6.19}
\]

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,
\[
\max_{k^* - N \leq k \leq k^*} \frac{1}{n} \| E_k^{(2)} \| \leq \max_{k^* - N \leq k \leq k^*} \left\| \sum_{i=1}^{k} \varepsilon_i \right\| + \frac{1}{n} \left\| \sum_{i=1}^{k^*} \varepsilon_i \right\| + \max_{k^* - N \leq k \leq k^*} \frac{k + k^*}{n^2} \left\| \sum_{i=1}^{k^*} \varepsilon_i \right\| = o_P(1), \tag{6.20}
\]

since the first term on the right-hand side is \(o_P(1)\) by Lemma A.1, and the second and third terms are \(o_P(1)\) by the Ergodic Theorem. Equations (6.17)–(6.20) imply that
\[
\max_{k^* - N \leq k \leq k^*} \left| \frac{1}{n} \left\langle E_k^{(1)}, E_k^{(2)} \right\rangle \right| = o_P(1).
\]

Additionally, in a similar fashion to (6.17),
\[
\max_{k^* - N \leq k \leq k^*} \left| \frac{1}{n} \left\langle D_k^{(1)}, E_k^{(2)} \right\rangle \right| \leq \max_{k^* - N \leq k \leq k^*} \| D_k^{(1)} \| \max_{k^* - N \leq k \leq k^*} \frac{1}{n} \| E_k^{(2)} \| \tag{6.21}
\]
\[
\leq \| \delta \|^2 N \max_{k^* - N \leq k \leq k^*} \frac{n - k^*}{n} \max_{k^* - N \leq k \leq k^*} \frac{1}{n} \| E_k^{(2)} \|
\]
\[
= o_P(1),
\]
according to (6.20). Equations (6.10), (6.15), (6.17), and (6.21) imply the lemma when \(-N \leq k - k^* \leq 0\), and (6.11), (6.16) plus similar arguments applied to the remaining two terms in (6.2) imply the result when \(0 \leq k - k^* \leq N\).

**Proof of Theorem 2.1.** Lemma 6.1 implies that it suffices to consider the weak convergence of \(\hat{k}_n - k^*\) on a bounded subset of the integers, and Lemma 6.2 along with the Continuous Mapping Theorem gives this the weak convergence to the limit on all bounded subsets and therefore proves the theorem.
6.3 Proof of Theorem 2.2

The proof of Theorem 2.2 is carried out analogously to Theorem 2.1, 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 6.3.** If the assumptions of Theorem 2.2 are satisfied, \( \|\delta_n\| | |\hat{k}_n^* - k^*| \) is bounded in probability.

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

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

as \( n \to \infty \). It follows that

\[
\max_{1 \leq k \leq k^*-N_\delta} \frac{1}{n} \left( D_k^{(1)} \right) \left( D_k^{(2)} \right) = \max_{1 \leq k \leq k^*-N_\delta} \|\delta_n\|^2 \left( \frac{k + k^*}{n} \right) \left( \frac{n - k^*}{n} \right)^2
\]

\[
= \|\delta_n\|^2 \left( \frac{2k^* - N_\delta}{n} \right) \left( \frac{n - k^*}{n} \right)^2
\]

\[
\to -2N\theta(1 - \theta)^2
\]

as \( n \to \infty \). Using similar arguments as those used to establish (6.4), (6.6), and (6.7), one can show that this term is the asymptotically dominant term in equation (6.1). Thus, for all \( M < 0 \),

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

Moreover,

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

by applying the same reasoning to the terms in (6.2). The preceding two equations imply the lemma. \( \square \)

**Lemma 6.4.** If the assumptions of Theorem 2.2 are satisfied, then

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

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

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

\[
\sup_{x \in [-N,0]} \left| \frac{1}{n} \left( D_k^{(1)} D_k^{(2)} \right) - (-2\theta(1 - \theta)^2 x) \right|
\]

\[
= \sup_{x \in [-N,0]} \left| k(x) \frac{2k^* + k(x)}{n} \left( \frac{n - k^*}{n} \right)^2 \|\delta_n\|^2 - (-2\theta(1 - \theta)^2 x) \right|
\]

\[
= o(1),
\]

18
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

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

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 A.2 yields that there exist two independent sequences of two-parameter Gaussian processes \((\Gamma^{(1)}_n(\cdot, \cdot) : n \in \mathbb{N})\) and \((\Gamma^{(2)}_n(\cdot, \cdot) : n \in \mathbb{N})\) satisfying

\[
\sup_{x \in [-N,0]} \int \left( \|\delta_n\| \sum_{i=k^*+k(x)+1}^{k^*+k(x)+1} \varepsilon_i(t) - \Gamma^{(1)}_n(-x, t) \right)^2 dt = o_P(1) \tag{6.22}
\]

and

\[
\sup_{x \in [0,N]} \int \left( \|\delta_n\| \sum_{i=k^*}^{k^*+k(x)+1} \varepsilon_i(t) - \Gamma^{(2)}_n(x, t) \right)^2 dt = o_P(1), \tag{6.23}
\]

such that \( \mathbb{E}[\Gamma^{(j)}_n(x, t) = 0, and \text{Cov}(\Gamma^{(j)}_n(x, t), \Gamma^{(j)}_n(x', t')) = \min(x, x')C_\varepsilon(t, t') \), 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

\[
\sup_{x \in [-N,0]} \left| \frac{1}{n} \left\langle E^{(1)}_{k^*+k(x)}, D^{(2)}_{k^*+k(x)} \right\rangle - 2\theta(1-\theta) \int \Gamma^{(1)}_n(-x, t) \frac{\delta_n(t)}{\|\delta_n\|} dt \right| = o_P(1). \tag{6.24}
\]

The definitions of \( E^{(1)}_k \) and \( D^{(2)}_k \) give that

\[
\sup_{x \in [-N,0]} \left| \frac{1}{n} \left\langle E^{(1)}_{k^*+k(x)}, D^{(2)}_{k^*+k(x)} \right\rangle - 2\theta(1-\theta) \int \Gamma^{(1)}_n(-x, t) \frac{\delta_n(t)}{\|\delta_n\|} dt \right| \tag{6.25}
\]

\[
= \sup_{x \in [-N,0]} \left| \frac{(2k^* + k(x))(n-k^*)}{n^2} \int \|\delta_n\| \sum_{i=k^*+k(x)+1}^{k^*+k(x)+1} \varepsilon_i(t) \frac{\delta_n(t)}{\|\delta_n\|} dt \right.
\]

\[
- 2\theta(1-\theta) \int \Gamma^{(1)}_n(-x, t) \frac{\delta_n(t)}{\|\delta_n\|} dt + \frac{k(x)(2k^* + k(x))(n-k^*)}{n^3} \int \sum_{i=1}^{n} \varepsilon_i(t)\delta_n(t) dt \right|.
\]

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

where

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

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

According (6.22) and since \( k(x)/n = O(1/\|\delta_n\|^2) \) uniformly in \( x \in [-N, 0] \), it follows that

\[
\sup_{x \in [-N,0]} H_2(x) = O(1) \frac{1}{\sqrt{n}} \int \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i(t) \delta_n(t) dt = o_P(1). \tag{6.26}
\]
Furthermore, \((2k^* + k(x))(n - k^*) / n^2 \rightarrow 2\theta(1 - \theta)\) as \(n \rightarrow \infty\). Hence,

\[
\sup_{x \in [-N,0]} H_1(x) = O(1) \sup_{x \in [-N,0]} \left| \int \left( \|\delta_n\| \sum_{i=k^*+k(x)+1}^{k^*} \varepsilon_i(t) - \Gamma_n^{(1)}(-x,t) \right) \|\delta_n\| dt \right|. \tag{6.27}
\]

It follows from the Cauchy–Schwarz inequality that

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

This combined with (6.26) and (6.22) implies that

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

which with (6.26) and (6.25) establishes (6.24). A parallel argument shows that

\[
\sup_{x \in [0,N]} \left| \frac{1}{n} \left\{ E_k^{(3)}(x) - D_k^{(4)}(x) \right\} - 2\theta(1 - \theta) \int \Gamma_n^{(2)}(x,t) \|\delta_n\| dt \right| = o_P(1). \tag{6.28}
\]

For \(x \in [0, N]\), let

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

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 \(\sigma^2\) 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 \(D[0,N]\). The same argument applies when \(\Psi_n\) is defined using \(\Gamma_n^{(2)}\). Therefore, (6.24) and (6.28) imply, upon showing that the remaining terms in (6.1) and (6.2) 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,

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

where

\[
T_1 = \max_{k^* - N_\delta \leq k \leq k^*} \frac{k - k^*}{n} \frac{n - k^*}{n} \int \sum_{i=1}^{k^*} \varepsilon_i(t) \|\delta_n\| dt,
\]

\[
T_2 = \max_{k^* - N_\delta \leq k \leq k^*} \frac{k - k^*}{n} \frac{n - k^*}{n} \int \sum_{i=1}^{k^*} \varepsilon_i(t) \|\delta_n\| dt,
\]

\[
T_3 = \max_{k^* - N_\delta \leq k \leq k^*} \frac{k + k^*}{n} \frac{k - k^*}{n} \frac{n - k^*}{n} \int \sum_{i=1}^{n} \varepsilon_i(t) \|\delta_n\| dt.
\]

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

\[
\left| \frac{k - k^*}{n} \right| \leq \frac{N}{n\|\delta_n\|^2}.
\]
Additionally, by the Cauchy–Schwarz and triangle inequalities,
\[ T_1 \leq O(1) \frac{N}{n\|\delta_n\|^2} \max_{k^* - N \leq k \leq k^*} \int \sum_{i=1}^{k} \varepsilon_i(t) \delta_n(t) dt \]
\[ = O(1) \frac{N}{\sqrt{n\|\delta_n\|^2}} \max_{k^* - N \leq k \leq k^*} \int \frac{1}{\sqrt{n}} \sum_{i=1}^{k} \varepsilon_i(t) \|\delta_n\| dt = o_P(1), \]

where in the last line Lemma A.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,
\[ \max_{k^* - N \leq k \leq k^*} \left| \frac{1}{n} \langle D_k^{(1)}, E_k^{(2)} \rangle \right| = o_P(1). \]

Additionally, by the Cauchy–Schwarz and triangle inequalities,
\[ \left| \frac{1}{n} \langle E_k^{(1)}, E_k^{(2)} \rangle \right| \leq \frac{1}{n} \int \left( \sum_{i=k+1}^{k^*} \varepsilon_i(t) + \frac{k - k^*}{n} \sum_{i=1}^{n} \varepsilon_i(t) \right) \left( \sum_{i=1}^{k} \varepsilon_i(t) + \sum_{i=1}^{n} \varepsilon_i(t) - \frac{k + k^*}{n} \sum_{i=1}^{n} \varepsilon_i(t) \right) dt \]
\[ \leq \frac{1}{n} \left\| \sum_{i=k+1}^{k^*} \varepsilon_i \right\| + \left\| \frac{k - k^*}{n} \sum_{i=1}^{n} \varepsilon_i \right\| \left( \left\| \sum_{i=1}^{k} \varepsilon_i \right\| + \left\| \sum_{i=1}^{n} \varepsilon_i \right\| + \left\| \frac{k + k^*}{n} \sum_{i=1}^{n} \varepsilon_i \right\| \right). \]

Therefore,
\[ \max_{k^* - N \leq k \leq k^*} \left| \frac{1}{n} \langle E_k^{(1)}, E_k^{(2)} \rangle \right| \leq \frac{1}{n} \left( \max_{k^* - N \leq k \leq k^*} \left\| \sum_{i=k+1}^{k^*} \varepsilon_i \right\| + \max_{k^* - N \leq k \leq k^*} \left\| \frac{k - k^*}{n} \sum_{i=1}^{n} \varepsilon_i \right\| \right) \]
\[ \times \left( \max_{k^* - N \leq k \leq k^*} \left\| \sum_{i=1}^{k} \varepsilon_i \right\| + \left\| \sum_{i=1}^{n} \varepsilon_i \right\| + \max_{k^* - N \leq k \leq k^*} \left\| \frac{k + k^*}{n} \sum_{i=1}^{n} \varepsilon_i \right\| \right). \]
\[ = \frac{1}{n} O_P \left( \sqrt{N \delta} + \frac{N \delta}{\sqrt{n}} \right) O_P(\sqrt{n}) = O_P \left( \sqrt{\frac{N \delta}{n}} + \frac{N \delta}{n} \right) \]
\[ = o_P(1), \]

where Lemma A.2 and the Continuous Mapping Theorem were again used to obtain upper bounds on the maximum norm terms. The remaining terms coming from (6.2) can be shown to be \( o_P(1) \) in a similar way.

**Proof of Theorem 2.2.** The proof of the theorem follows from the lemmas in this subsection as the proof of Theorem 2.1 from the lemmas in the previous subsection.

**6.4 Proof of Theorem 2.3**

**Proof of Theorem 2.3.** The proof is obtained from an application of the Continuous Mapping Theorem and Theorem 1.2 of Jirak (2013).
A Technical Lemmas

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

\[
\max_{1 \leq k \leq n} \frac{1}{\sqrt{k}} \left\| \sum_{i=1}^{k} \varepsilon_i \right\| = O_p(\log^{1/p}(n)) \quad (n \to \infty).
\]

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

\[
\max_{1 \leq k \leq n} \frac{1}{\sqrt{k}} \left\| \sum_{i=1}^{k} \varepsilon_i \right\| \leq \max_{1 \leq j \leq c \log(n)} \max_{\rho^{j-1} < k \leq \rho^j} \frac{1}{\sqrt{k}} \left\| \sum_{i=1}^{k} \varepsilon_i \right\|
\]

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

\[
P\left( \max_{i \in A} X_i > x \right) \leq P\left( \bigcup_{i \in A} \{ X_i > x \} \right) \leq \sum_{i \in A} P(X_i > x),
\]

and Chebyshev’s inequality that

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

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 (A.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 A.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

\[
\sup_{0 \leq x \leq N} \int \left( S_n^{(1)}(x, t) - \Gamma_n^{(1)}(x, t) \right)^2 dt = o_P(1), \quad (A.2)
\]
and

$$\sup_{0 \leq x \leq N} \int_{0}^{t} \left( S_{n}^{(2)}(x, t) - \Gamma_{n}^{(2)}(x, t) \right)^{2} \, dt = o_{P}(1), \quad (A.3)$$

where $E[\Gamma^{(j)}_{n}(x,t)] = 0$ and $\text{Cov}(\Gamma^{(j)}_{n}(x,t), \Gamma^{(j)}_{n}(x',t')) = C_{\varepsilon}(t,t')$, $j = 1, 2$, with $C_{\varepsilon}(t,t')$ defined in (2.7).

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

$$S_{n}^{(1,*)}(x,t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{nx} \varepsilon_{i}^{(i)}(t),$$

where $\varepsilon_{i}^{(i)}$ 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

$$E \left[ \sup_{0 \leq x \leq N} \left\| S_{n}^{(1)}(x, \cdot) - S_{n}^{(1,*)}(x, \cdot) \right\| \right] \leq \frac{1}{\sqrt{n}} E \left[ \sup_{0 \leq x \leq N} \sum_{i=1}^{nx} \| \varepsilon_{i} - \varepsilon_{i}^{(i)} \| \right] \leq \frac{1}{\sqrt{n}} \sum_{i=1}^{\infty} \left( E \left[ \| \varepsilon_{i} - \varepsilon_{i}^{(i)} \|^{2} \right] \right)^{1/2} = o(1).$$

Now Markov’s inequality implies that

$$\sup_{0 \leq x \leq N} \int_{0}^{t} \left( S_{n}^{(1)}(x, t) - S_{n}^{(1,*)}(x, t) \right)^{2} \, dt = o_{P}(1). \quad (A.4)$$

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

$$\sup_{0 \leq x \leq N} \int_{0}^{t} \left( S_{n}^{(1,*)}(x, t) - \Gamma_{n}^{(1)}(x, t) \right)^{2} \, dt = o_{P}(1).$$

This and (A.4) imply (A.2), and the sequences $\Gamma_{n}^{(1)}$ and $\Gamma_{n}^{(2)}$ have been constructed to be independent. \qed

**References**


Figure 2.1: Top: plots of the basis functions $b_1$ (red) and $b_2$ (blue). Middle: the functional observations corresponding to $c = 2$ (left) and $c = 4$ (right); observations prior to (post) $k^* = 50$ are in red (blue). Bottom: histograms of the break point estimators $\hat{k}_{100}$ (blue) and $\tilde{k}_{100}$ (red) for $c = 2$ (left) and $c = 4$ (right).
Figure 3.1: Histograms of the fully functional (blue) and the fPCA based (red) break point estimator for FAR(2) processes in the prominent break case with $k^* = 50$, $\sigma_{\text{fast}}$ and sensitivity levels $c = 0.2$ (top left), $c = 0.5$ (top right), $c = 0.8$ (bottom left) and $c = 1.5$ (bottom right).
Figure 3.2: Box plots of the fully functional (blue) and the fPCA based (red) break point estimator for FAR(2) processes and \( k^* = 25 \) in the prominent break case with \( \sigma_{\text{fast}} \) (left) and in the moderate break case with \( \sigma_{\text{slow}} \) (right). The \( x \)-axis specifies the sensitivity level \( c \), the \( y \)-axis the estimated break point fraction (relative to \( n \)). The true break point fraction \( \theta = 0.25 \) is indicated by a dashed horizontal line.

Figure 3.3: Histograms of the break date estimators for \( t \)-distributed innovations with 2 (left), 3 (middle) and 4 (right) degrees of freedom. The proposed method is in blue, the fPCA method in red.
Figure 4.1: 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 4.2: Left: the (smoothed) first eigenfunction obtained from fPCA. Right: mean curves prior to (red) and post (blue) the estimated break date 09/18/2001.
Figure 4.3: 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.

Figure 4.4: Estimated break function $\hat{\delta}$ (left) and proportion of variation in $\hat{\delta}$ explained by the $\ell$th sample eigenfunction (right) for the transformed cumulative Microsoft log-return curves.