Statistica Sinica Preprint No: SS-2025-0014	
Title	Detecting Structural Breaks in High-dimensional Functional Time Series Factor Models
Manuscript ID	SS-2025-0014
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0014
Complete List of Authors	Caixia Xu, Huacheng Su, Xu Liu and Jinhong You
Corresponding Authors	Huacheng Su
E-mails	suhuacheng79@163.com
Notice: Accepted author version.	

Detecting Structural Breaks in High-Dimensional Functional Time Series Factor Models

Caixia Xu[1], Huacheng Su[2,*] , Xu Liu[1], Jinhong You[1]

[1]*Shanghai University of Finance and Economics*

and [2]*Ningbo University of Technology*

Abstract:

This paper investigates the estimation of a functional factor model characterized by factor loadings that may change over time with the number of changes being unknown. We propose a novel procedure to detect potential breaks and identify their locations. In the first step, we compute factor loadings for each time point and analyze their differences between consecutive time points. In the second step, we employ the wild binary segmentation method (WBS) to estimate both the number and positions of change points in the sequence of these differences. In the third step, we utilize the estimated change point positions to re-estimate the functional factor model. This results in functional data where

*Corresponding author: Huacheng Su, School of Statistics and Data Science, Ningbo University of Technology, Ningbo 315211, Zhejiang, China, E-mail: suhuacheng79@163.com. ORCID: 0009-0005-2095-7592

Caixia Xu: ORCID: 0009-0005-5063-0507

Xu Liu: ORCID: 0000-0003-3829-1715

change points are known, leading to reduced fitting errors. It is crucial to emphasize that throughout the process of estimating the loading and number of factors, we have effectively leveraged the unique characteristics of complex functional data and mitigated the impact of unknown change points. We demonstrate that the proposed method can correctly identify the number of changes and accurately estimate their locations with probability approaching one. Simulation studies and empirical applications illustrate the excellent finite-sample performance of our proposed approach.

Key words and phrases: Break point; Functional factor model; Wild binary segmentation

1. Introduction

With advancements in socio-economic and technological fields, high-dimensional functional time series (FTS), which consist of sequences of curves exhibiting serial dependence over time, are becoming increasingly popular. Examples include intraday return density trajectories for various stocks, annual temperature curves recorded at different stations, and daily energy consumption curves for numerous households, among others. A powerful tool for analyzing functional time series is factor models. In this article, our focus lies on the identification of change points in functional time series factor models and the corresponding theoretical research.

Given the unique nature of functional time series, it is crucial to consider the characteristics of both time series and functional data when analyzing such data. In time series factor models, a few latent common factors drive the collective behavior of a high-dimensional vector of time series variables. This characteristic has positioned these models as a promising approach for analyzing high-dimensional time series data. Nevertheless, traditional factor models often assume that factor loadings, the coefficients linking variables to common factors, remain stable over extended periods. This assumption of time-invariant loadings may be restrictive, as it overlooks potential changes in these relationships. In reality, substantial research has explored the concept of change points in factor models. For instance, Stock and Watson (2009) investigated the reliability of forecasting in the presence of structural breaks in the factor loadings. Breitung and Eickmeier (2011) introduced three statistical measures for examining structural breaks in factor loadings, drawing inspiration from Andrews (1993). Chen et al. (2014) proposed a two-stage method for identifying significant breaks in factor loadings. Yamamoto and Tanaka (2015) introduced an enhanced adaptation of Breitung and Eickmeier (2011) tested to enhance its robustness. Su and Wang (2017) explored estimation and testing within time-varying factor models, developing a method capable of detecting multiple breaks in

factor loadings. Ma and Su (2018) presented a three-step method for structural break detection that can automatically verify the presence of breaks and estimate their precise locations.

In the context of high-dimensional FTS, functional factor models are employed to uncover the latent dynamic structure. These models use loadings to link functional factors with the observed variables. Retaining the functional and time series structure is crucial to model accurately the dynamics of the entire curve. This consideration enhances the significance of functional factor models in the context of functional time series (e.g., Kokoszka et al. (2015), Gao et al. (2021), Guo et al. (2021), Hallin et al. (2023), Chang et al. (2024) , Hu and Yao (2022), and Hu and Yao (2024)). Similar to traditional factor models, the loadings in functional factor models are also likely to vary over extended time spans. Taking the empirical study in Section 4 as an example where we analyze daily pollutant concentrations curves in the atmosphere across multiple locations, we summarize these variables by latent functional factors such as seasonal patterns and anthropogenic influences. The loadings that link variables to factors fluctuate because the influence of different factors can vary over time. As depicted in Figure 4 of Supplement material, when measuring the distance d_t between the estimated loadings at time t and $t-1$, most distances are non-zero,

where larger value of d_t indicates more significant change in loadings. This observation suggests that the assumption of constant loadings, as used in traditional factor models, needs to be re-evaluated.

The detection of change points in factor model loadings and in the mean function of functional time series are two distinct issues. On one hand, change points in the mean function reflect variations in the variables themselves, whereas change points in the loadings indicate alterations in the relationship between latent factors and variables. Taking the pollutant concentration data as an example, the emphasis on changes in loadings is to understand the mechanisms behind variations in pollutant concentrations and their relationship with latent factors, rather than solely focusing on the concentration changes themselves. As illustrated in Figure 1, the appearance of change points modifies the daily trends in pollutant concentrations, which indicates a significant correlation with fluctuations in the impact of factors on variables. On the other hand, it is enough to analyze single curve for identifying change points in the mean function, while functional factor models involve numerous curves. As such, the detection of change points in these models requires more complex methodologies. A substantial body of research has addressed change points in the mean function of functional time series, including studies by Berkes et al. (2009), Zhang et al. (2011),

Aston and Kirch (2012), Horváth et al. (2014), Aue et al. (2018) and Li et al. (2023)). There is also related work by Madrid Padilla et al. (2022), which proposes a kernel-based functional seeded binary segmentation method for detecting changes in mean functions indexed by a continuous covariate. In their framework, the functional object of interest is a deterministic mean function over the covariate domain, with dimensionality determined by the discretization of that domain.

However, to the best of our knowledge, there is currently a lack of research specifically addressing break detection and estimation in high-dimensional functional time series factor models. Due to the complex structure and redundant information inherent in functional time series, traditional change point identification methods used in factor models cannot be directly applied to functional factor models. Therefore, developing new methodologies tailored to the big challenges of functional factor models is essential for advancing this field.

In this study, we propose a novel approach for detecting structural breaks in functional time series factor models. Rather than being constrained by the inherent complexity and extensive volume of functional data, our method fully leverages the advantages of functional factor model to accurately identify both the number and positions of change points. Our

approach consists of three main steps. Firstly, we calculate factor loadings for each time point and examine the variations between consecutive time points to assess changes. Secondly, we employ the wild binary segmentation method (WBS) (Fryzlewicz (2014)) to identify the locations and number of structural breaks. Lastly, based on the estimated number and positions of the breaks, we re-estimate the factor loadings and functional factors to achieve a more accurate representation of the functional data with structural breaks. Unlike conventional factor models, where the data of each time point is depicted as a vector, functional data encapsulates rich and comprehensive information. This allows the construction of a conventional factor model at each time point and effectively turning the abundance of information in functional time series from a curse into a blessing. The efficacy of the proposed method is assessed through both simulated and real-world data examples.

In conclusion, this paper offers the following contributions:

1. We propose a novel method for identifying change points in factor models of functional time series. This approach capitalizes on the distinctive features of functional data, addressing the challenges inherent in such complex datasets;

2. We provide a rigorous theoretical framework for the proposed change

point identification method, including establishing the consistency of the estimation process and ensuring accuracy in detecting both the number and positions of change points;

3. Building on the change point identification framework, we devise a robust estimation method for functional factor models with corresponding asymptotic properties.

The subsequent sections of this paper are structured as follows. In Section 2, we present the detailed methodology of our procedure for detecting and estimating structural break points. Section 3 delves into the asymptotic theory underlying our approach. In Section 4, we present an empirical study using real-world data to further validate the effectiveness of our approach. In Section 5, we conclude with a summary of our findings and contributions. The proofs and technical details for the theoretical results are provided in the online Supplementary Material, along with simulation studies to demonstrate the efficacy of our proposed methods, and additional results of the real application.

2. The estimation of functional factor model and detecting procedure

In this section, we consider a large-dimensional functional factor model with an unknown number of breaks, and then propose a procedure for estimation. We first introduce some notations which will be used throughout the paper. For an $m \times n$ real matrix $\mathbf{X}$, we denote its transpose as $\mathbf{X}^T$, its rank as $\mathrm{r}(\mathbf{X})$, its Frobenius norm as $\|\mathbf{X}\| \equiv [\mathrm{tr}(\mathbf{X}\mathbf{X}^T)]^{1/2}$. For any positive numbers a_n and b_n, let $a_n \asymp b_n$ denote $\lim_{n\to\infty} a_n/b_n = c$ for a positive constant c, and let $a_n >> b_n$ denote $a_n^{-1} b_n = o(1)$. The operator $\xrightarrow{p}$ denotes convergence in probability. We use $(N,T) \to \infty$ to denote that N and T pass to infinity jointly. We denote the kth largest eigenvalue of a positive semidefinite matrix A as $\psi_k(A)$. For any real number a, $[a]$ denotes its integer part. Let $\eta_{a,b} = \sqrt{\min\{a,b\}}$ for constants a, b.

2.1 The functional factor model

Consider the time-varying functional factor model

$$X_{it}(s) = \boldsymbol{\lambda}_{it}^T \boldsymbol{f}_t(s) + u_{it}(s), i = 1, \cdots, N, t = 1, \cdots, T, s \in [0,1],$$

where $X_{it}(s)$ is the functional data which is the ith individual at time t, both $\boldsymbol{\lambda}_{it} = (\lambda_{it1}, \cdots, \lambda_{itr})^T$ and $\boldsymbol{f}_t(s) = (f_{t1}(s), \cdots, f_{tr}(s))^T$ are $r \times 1$ functional

vectors. $\boldsymbol{\lambda}_{it}$ is time-dependent factor loadings and $\boldsymbol{f}_t(s)$ is an unobserved common factors which is a function of s, $u_{it}(s)$ is the idiosyncratic error term independent of $\boldsymbol{f}_t(\cdot)$ with $E(u_{it}(s)) = 0$ and $\text{var}(u_{it}(s)) = \sigma^2$, and both N and T pass to infinity. For simplicity of technical proofs, we assume that r does not depend on N and T , but it is unknown, hence we need to estimate r from the data. Writing the above model in the vector form, we have

$$\mathbf{X}_t(s) = \boldsymbol{\lambda}_t \boldsymbol{f}_t(s) + \boldsymbol{u}_t(s), t = 1, \cdots, T, s \in [0, 1], \tag{2.1}$$

where $\mathbf{X}_t(s) = (X_{1t}(s), \cdots, X_{Nt}(s))^T$, $\boldsymbol{\lambda}_t = (\boldsymbol{\lambda}_{1t}, \cdots, \boldsymbol{\lambda}_{Nt})^T$ and $\mathbf{u}_t = (u_{1t}, \cdots, u_{Nt})^T$. To reflect the situation of irregular and possibly subject-specific time points, we assume that $X_t(\cdot)$ is measured at $\boldsymbol{s} = (s_1, \cdots, s_n)$.

We assume that there are m break points in the process $\{\boldsymbol{\lambda}_t\}$ and it satisfies $m \ll T$. When $m \geq 1$, denote the m break points by $\{t'_1, \cdots, t'_m\}$ that satisfy $1 \equiv t'_0 < t'_1 < \cdots < t'_m < t'_{m+1} \equiv T$, by which the whole time span is divided into $m + 1$ regimes, denoted by $I_k = [t'_k, t'_{k+1})$ for $k = 0, 1, \cdots, m - 1$ and $I_m = [t'_{m-1}, t'_m]$. We assume that $\boldsymbol{\lambda}_t = \boldsymbol{\alpha}_k^0$ for all $t \in I_k$ and $k = 0, \cdots, m$. When $m = 0$, we have $I_0 = I_m = [t_0, t_1) = [1, T]$ and $\boldsymbol{\lambda}_t = \boldsymbol{\alpha}_0$ for all $t \in [1, T]$ which means no break happens in this scenario. Let $\boldsymbol{\alpha}_k^0 = (\boldsymbol{\alpha}_{1k}^0, \cdots, \boldsymbol{\alpha}_{Nk}^0)^T$ for $k = 0, \cdots, m$ and $\boldsymbol{\alpha}_{ik}^0 \in \mathbb{R}^{r\times 1}$ for $i = 1, \cdots, N$. In practice, the number of factors r, the factors $\boldsymbol{f}_t(s)$, and the loading $\boldsymbol{\alpha}_k^0$ are unknown as well as the number of breaks m and the locations of the

breaks $t'_1, \cdots, t'_m$. We have to estimate the loadings at each time points and determine the number of factors through GR information criterion first. Then we use wild binary segmentation to estimate the number of break points and their locations.

2.2 Estimation of loadings at each point

In estimation of loadings, we assume that the number of factors r is known. Building upon this assumption, we proceed to estimate the loading at each time point t. The procedure is described as follows.

For each t, we estimate $\boldsymbol{\lambda}_t$ by minimizing

$$\sum_{l=1}^{n}(\boldsymbol{X}_t(s_l) - \boldsymbol{\lambda}_t \boldsymbol{f}_t(s_l))^T(\boldsymbol{X}_t(s_l) - \boldsymbol{\lambda}_t \boldsymbol{f}_t(s_l)). \tag{2.2}$$

Consider the identification conditions $N^{-1}\boldsymbol{\lambda}_t^T\boldsymbol{\lambda}_t = \mathbb{I}_r$ for each t which is imposed by Bai and Ng (2002) and is widely used in factor model, we have $\hat{\boldsymbol{f}}_t(s_l) = (N^{-1}\boldsymbol{\lambda}_t^T\boldsymbol{\lambda}_t)^{-1}(N^{-1}\boldsymbol{\lambda}_t^T\boldsymbol{X}_t(s)) = N^{-1}\boldsymbol{\lambda}_t^T\boldsymbol{X}_t(s)$. By substituting $\boldsymbol{f}_t(s_l)$ with $\hat{\boldsymbol{f}}_t(s_l)$, the above objective function (2.2) becomes

$$\begin{aligned}
&\sum_{l=1}^{n}\boldsymbol{X}_t^T(s_l)\boldsymbol{X}_t(s_l) - \sum_{l=1}^{n}\frac{1}{N}\boldsymbol{X}_t^T(s_l)\boldsymbol{\lambda}_t\boldsymbol{\lambda}_t^T\boldsymbol{X}_t(s_l) \\
&= \sum_{l=1}^{n}\boldsymbol{X}_t^T(s_l)\boldsymbol{X}_t(s_l) - \frac{1}{N}\mathrm{tr}(\boldsymbol{\lambda}_t^T\sum_{l=1}^{n}\boldsymbol{X}_t(s_l)\boldsymbol{X}_t^T(s_l)\boldsymbol{\lambda}_t).
\end{aligned}$$

Thus, $\hat{\boldsymbol{\lambda}}_t$ can be estimated by maximizing $\frac{1}{N}\mathrm{tr}(\boldsymbol{\lambda}_t^T\sum_{l=1}^{n}\boldsymbol{X}_t(s_l)\boldsymbol{X}_t^T(s_l)\boldsymbol{\lambda}_t)$ subject to $N^{-1}\boldsymbol{\lambda}_t^T\boldsymbol{\lambda}_t = \mathbb{I}_r$. When $\mathrm{rank}(\sum_{l=1}^{n}\boldsymbol{X}_t(s_l)\boldsymbol{X}_t^T(s_l)) \geq r$ for each

$t = 1, \cdots, T$, we have $\hat{\boldsymbol{\lambda}}_t = N^{1/2} E_{eigen}(\frac{1}{n}\sum_{l=1}^{n} \boldsymbol{X}_t(s_l)\boldsymbol{X}_t^T(s_l); r)$, where $E_{eigen}(A, q)$ is a matrix composed of the orthogonal eigenvectors corresponding to the q largest eigenvalues of matrix A. By summing the matrices corresponding to the time points $s_1, \cdots, s_n$, the estimation of the loading does not depend on $s_1, \cdots, s_n$.

Before we proceed to explain how we estimate the smoothed functional factors, we temporarily provide the factor estimates for each s_l by $\hat{\boldsymbol{f}}_t(s_l) = N^{-1}\hat{\boldsymbol{\lambda}}_t^T \boldsymbol{X}_t(s_l)$, which results in $\hat{\boldsymbol{X}}_t(s_l) = \hat{\boldsymbol{\lambda}}_t \hat{\boldsymbol{f}}_t(s_l)$, $l = 1, \cdots, n$.

In this stage, we establish objective function (2.2) to obtain estimators of loadings $\boldsymbol{\lambda}_t$ for each time point t. When the number of change points $m = 0$, indicating without variability, the loading matrices at each time point should be equal. Rewriting the objective function (2.2) as

$$\sum_{t=1}^{T}\sum_{l=1}^{n}(\boldsymbol{X}_t(s_l) - \boldsymbol{\lambda}_t \boldsymbol{f}_t(s_l))^T(\boldsymbol{X}_t(s_l) - \boldsymbol{\lambda}_t \boldsymbol{f}_t(s_l)),$$

we can derive a new estimator for the loading matrix by the same way, which can be expressed as $\hat{\boldsymbol{\alpha}} = N^{1/2} E_{eigen}(\frac{1}{T}\sum_{t=1}^{T}\frac{1}{n}\sum_{l=1}^{n} \boldsymbol{X}_t(s_l)\boldsymbol{X}_t^T(s_l); r)$. For functional time series data, even when a specific time point t is given, that is, when $T = 1$, we still have matrix $n^{-1}\sum_{l=1}^{n} \boldsymbol{X}_t(s_l)\boldsymbol{X}^T(s_l)$ which is non-singular when n is sufficiently large. Conversely, in the context of a typical unknown change point factor model, when a specific time point t is given, we only have one data information available, i.e., $\boldsymbol{X}_t\boldsymbol{X}_t^T$ which is singular.

Consequently, it is not possible to directly obtain the first r eigenvectors of $\boldsymbol{X}_t\boldsymbol{X}_t^T$. This is why we often use the piecewise approach to estimate the loading matrices within specific intervals.

2.3 Determination of the number of factors

In our preceding discussion, we have taken for granted that the count of factors, denoted as r, is pre-established. However, in a real-world scenario, r needs to be ascertained from the dataset itself. Two criterion functions used to estimate the number of factors r for each $t = 1, \cdots, T$ is typically given by Ahn and Horenstein (2013) which called ER and GR. The expressions for ER and GR are defined as follows:

$$ER_t(r) = \frac{\tilde{\psi}_r}{\tilde{\psi}_{r+1}}; GR_t(r) = \frac{ln(1+\tilde{\psi}_r/V(r))}{ln(1+\tilde{\psi}_{r+1}/V(r+1))} \tag{2.3}$$

where $\tilde{\psi}_r$ is the rth largest eigenvalues of $(Nn)^{-1}\sum_{l=1}^{n}\boldsymbol{X}_t(s_l)\boldsymbol{X}_t^T(s_l)$, $V(r) = \sum_{k=r}^{N}\tilde{\psi}_k$ and $\hat{r}_t = \arg\max_{1\le r\le r_{\max}} ER_t(r)$ or $\hat{r}_t = \arg\max_{1\le r\le r_{\max}} GR_t(r)$, $\hat{r}$ is the mode of $\boldsymbol{r} = (\hat{r}_1, \cdots, \hat{r}_T)$, denote by $\hat{r} = \mathcal{M}(\boldsymbol{r})$.

Remark 1. We emphasize that in the process of selecting the number of factors, we fully utilize the characteristics of functional data by considering their values at multiple time points. For each t, we estimate a corresponding number of factors, denoted as $\hat{r}_t$, and then take the mode of all $\hat{r}_t$ values.

This approach effectively mitigates the potential impact of unknown change points on the estimation of the number of factors. At least, it is reasonable to assume that the majority of data points do not exhibit time heterogeneity. Subsequent simulation studies further demonstrate this point.

2.4 Estimating the number and locations of break points

Due to the unknown change-point positions, we estimate the loading matrices for each time point t. As discussed in Section 2.1, when there are no structural breaks between two time points, indicating that both $\boldsymbol{\lambda}_t$ and $\boldsymbol{\lambda}_{t+1}$ belong to the same interval I_k for $k = 0, \cdots, m$, we observe that $\boldsymbol{\lambda}_t = \boldsymbol{\lambda}_{t+1} = \boldsymbol{\alpha}_k^0$. Consequently, the distance between these two loading matrices should be equal to zero. To calculate the difference between adjacent loadings, we define the distance as $d_t = \|P_{\hat{\boldsymbol{\lambda}}_{t+1}} - P_{\hat{\boldsymbol{\lambda}}_t}\|/\sqrt{2r}$, which results in a time series $\boldsymbol{d} = (d_1, \cdots, d_{T-1})$ with length $T-1$, where P_A is the orthogonal projection matrix onto A. Based on Theorem 1 in Section 3, the time series $\boldsymbol{d}$ should approach infinitesimal values when there are no change points. We consider the model $d_t = \mu_t + \epsilon_t, t = 1, \cdots, T-1$, where μ_t is a deterministic, one-dimensional, piecewise constant signal with change-points whose number m and locations $t'_1, ..., t'_m$ are unknown. Further technical assumptions on μ_t and ϵ_t will be specified later.

The fundamental component of the binary segmentation(BS) algorithm is the CUSUM statistic as presented below:

$$\tilde{D}^b_{a,c} = \sqrt{\frac{c-b}{h(b-a+1)}} \sum_{t=a}^{b} d_t - \sqrt{\frac{b-a+1}{h(c-b)}} \sum_{t=b+1}^{c} d_t,$$

where $(a, c) \subset \{1, \cdots, T\}$ with $a < c-1$, and any time point $b = a, \cdots, c-1$, $h = c-a+1$. The initial step of the BS algorithm involves calculating $\tilde{D}^b_{1,T}$, followed by selecting $b_{1,1} = \arg\max_{b:1\le b<T} |\tilde{D}^b_{1,T}|$ as the first candidate for a change point. This candidate's significance is evaluated against a specific criterion. If deemed significant, the interval $[1, T]$ is divided into two sub-intervals on either side of $b_{1,1}$, initiating a recursive process of computing $\tilde{D}^b_{1,b_{1,1}}$ and $\tilde{D}^b_{b_{1,1}+1,T}$, potentially leading to further subdivisions. In scenarios with a single change point, the output $b_{1,1}$ from the BS algorithm aligns with the Maximum Likelihood Estimation (MLE) when ϵ_t follows an independent and identically distributed Gaussian distribution, thus demonstrating strong performance.

However, as highlighted by Fryzlewicz (2014), since combined effect can be canceled out, the binary segmentation (BS) method may not be effective when there are multiple change points. To address this limitation, the WBS method improves upon BS by conducting multiple CUSUM tests over randomly selected sub-intervals $\{(\alpha_p, \beta_p)\}_{p=1}^{P}$, whose start- and end-point are independently and uniformly (with replacement) drawn from the

set $\{1, \cdots, T\}$. This approach ensures that each change point is, with high probability, the only change point within a selected interval. Please refer to Algorithm 1 (Fryzlewicz, 2014) for a formal description of the WBS method.

Algorithm 1 Wild Binary Segmentation WBS(a,c,$\{(\alpha_p, \beta_p)\}_{p=1}^{P}, \Delta_T$)

INPUT: Independent samples $\{d_t\}_{t=1}^{T-1}$, collection of intervals $\{(\alpha_p, \beta_p)\}_{p=1}^{P}$, threshold parameter Δ_T.
for $p = 1, \cdots, P$ **do**
  $(a_p, c_p) := [a, c] \cap [\alpha_p, \beta_p]$
  if $c_p - a_p < 1$ **then**
    STOP
  else
    $b_p := \arg\max_{b \in a_p, \cdots, c_p - 1} |\tilde{D}^b_{a_p, c_p}|$
    $e_p := |\tilde{D}^{b_p}_{a_p, c_p}|$
  end if
end for
$p^* := \arg\max_{p=1, \cdots, P} e_p$
if $e_{p*} > \Delta_T$ **then**
  add b_{p^*} to the set of estimated change-points
  Conduct WBS (a, b_{p^*}, Δ_T) and WBS $(b_{p^*} + 1, c, \Delta_T)$
end if
OUTPUT: The set of estimated change points.

Let $\hat{m}$ represent the estimated number of change points obtained through the WBS procedure, and let $\hat{t}'_1, \cdots, \hat{t}'_{\hat{m}}$ denote their locations, sorted in increasing order. Indeed, both estimated number and locations of change

points are inherently influenced by the choice of the selected threshold Δ_T. Denote $\hat{m}(\Delta_T) = \hat{m}$ and $C(\Delta_T) = \{\hat{t}'_1, \cdots, \hat{t}'_{\hat{m}}\}$. Consider any decreasing sequence $\{\Delta_T^k\}_{k=0}^K$ of thresholds such that $|C(\Delta_T^k)| = k$ for a certain fixed constant K, and assume $T \geq K$. We define $C_k = C(\Delta_T^k)$. In this section, we propose a method to select a model from the collection C_k for $k = 0$ to K by minimizing what we refer to as the "strengthened Schwarz information criterion (sSIC)". The sSIC is defined as follows:

$$\text{sSIC}(k) = \frac{T}{2}\log(\hat{\sigma}_k^2) + k\log^{\alpha}(T).$$

For any candidate model C_k, denote by k the estimation of the number of change points m, and by $\hat{t}^k = (\hat{t}'_0, \cdots, \hat{t}'_{k+1})$ the estimation of positions of change points, where $\hat{t}'_0 = 1$ and $\hat{t}'_{k+1} = T$. We can re-estimate the loading matrices for each of the $k+1$ intervals using the following equation:

$$\hat{\boldsymbol{\alpha}}_i^k = N^{1/2} E_{eigen}[\sum_{t=\hat{t}'_{i-1}}^{\hat{t}'_i} \sum_{l=1}^{n} \boldsymbol{X}_t(s_l)\boldsymbol{X}_t^T(s_l); \hat{r}] \tag{2.4}$$

for $i = 0, \cdots, k$. Then for fixed $t \in [t'_i, t'_{i+1}]$, $\hat{\sigma}_k^2$ is defined as

$$\hat{\sigma}_k^2 = (Tn)^{-1} \sum_{t=1}^{T} \sum_{l=1}^{n} \|\hat{\boldsymbol{X}}_t^k(s_l) - \boldsymbol{X}_t(s_l)\|_F^2, \tag{2.5}$$

where $\hat{\boldsymbol{X}}_t^k(s)$ is estimated by $\hat{\boldsymbol{X}}_t^k(s) = \hat{\boldsymbol{\alpha}}_i^k \hat{\boldsymbol{f}}_t^k(s)$ with $\hat{\boldsymbol{f}}_t^k(s) = \hat{\boldsymbol{\alpha}}_i^{kT} \boldsymbol{X}_t^k(s)/N$.

In sections 3, we will demonstrate the consistency of the number of change points selected by the WBS method. Additionally, we will show the

accuracy of change point location estimation through simulation studies in the online Supplementary Material.

2.5 Re-estimating the functional factor model

Since the change-point positions $\hat{t}_1, \cdots, \hat{t}_m$ have already been estimated, we can re-estimate loading $\hat{\boldsymbol{\alpha}}_k$ for each interval I_k by equation (2.5).

In order to obtain an estimate of the smoothed functional data $\boldsymbol{X}_t(s)$, we first apply Kauhunen-Loève expansion (Ash and Gardner (2014)) to the factor process $\boldsymbol{f}_t(s) = (f_{t1}(s), \cdots, f_{tr}(s))^T$ considering the dependence of functional data over time, that is $f_{tq}(s) = \sum_{k=1}^{\infty} \xi_{tqk}\phi_{qk}(s), q = 1, \cdots, r$, where $\phi_{qk}(s)$ is the kth orthonormal eigenfunction of the covariance function $C_q(s_1, s_2) = \text{cov}(f_{tq}(s_1), f_{tq}(s_2))$ for factor process q, which satisfies $\int_0^1 \phi_{qk}(s)\phi_{qk'}(s)ds = 1$ if $k = k'$ and 0 otherwise. ξ_{tqk} is the functional principal component score for the stochastic process $f_{tq}(s)$ with $E(\xi_{tqk}) = 0$, $\text{var}(\xi_{tqk}) = \rho_{qk}$ and $\text{cov}(\xi_{tqk}, \xi_{jqk'}) = 0$ if $k \neq k'$. ρ_{qk} is the eigenvalue corresponding to the eigenfunction $\phi_{qk}(\cdot)$, where $\rho_{q1} \geq \rho_{q2} \geq \cdots > 0$ and $\sum_{k=1}^{\infty} \rho_{qk} < \infty$ for any $q = 1, \cdots, r$, which implies that $\sup_{s\in[0,1]} E(\sum_{k=1}^{K} \xi_{tqk}$ $\phi_{qk}(s) - \sum_{k=1}^{\infty} \xi_{tqk}\phi_{qk}(s))^2 \to 0$ as $K \to \infty$, then we can denote $f_{tq}(s)$ as

$$f_{tq}(s) \approx \sum_{k=1}^{K} \xi_{tqk}\phi_{qk}(s), q = 1, \cdots, r. \tag{2.6}$$

Equation (2.6) has been extensively researched in the scenario where K is

constant. Here we allow $K \to \infty$ for improving flexibility. Besides, we use the same K for $q = 1, \cdots, r$ for simple notation.

Denote $\boldsymbol{\xi}_{tq} = (\xi_{tq1}, \cdots, \xi_{tqK})^T \in \mathbb{R}^K$, $\boldsymbol{\xi}_t = (\boldsymbol{\xi}_{t1}^T, \cdots, \boldsymbol{\xi}_{tr}^T)^T \in \mathbb{R}^{Kr}$ and $\boldsymbol{\Phi}(s) = \text{diag}(\boldsymbol{\Phi}_1(s), \cdots, \boldsymbol{\Phi}_r(s))$, where $\boldsymbol{\Phi}(s)$ is a $Kr \times r$ block diagonal matrix with block q being $\boldsymbol{\Phi}_q(s) = (\phi_{q1}(s), \cdots, \phi_{qK}(s))^T$. Then we have

$$\boldsymbol{X}_t(s) \approx \boldsymbol{\lambda}_t \boldsymbol{\Phi}^T(s) \boldsymbol{\xi}_t + \boldsymbol{u}_t(s) \tag{2.7}$$

and $\boldsymbol{\lambda}_t = \boldsymbol{\alpha}_k$ for $t \in I_k$, $k = 0, \cdots, m$.

Consider the identification conditions, (I1) $N^{-1} \boldsymbol{\lambda}_t^T \boldsymbol{\lambda}_t = \mathbb{I}_r$ for each t, and the maximum value of each column in matrix $\boldsymbol{\lambda}_t$ is positive; (I2) $\boldsymbol{\xi}\boldsymbol{\xi}^T$ is a $Kr \times Kr$ diagonal with decreasing diagonal entries, where $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_T)^T$;(I3) $\int \Phi(s)\Phi(s)^T ds = \mathbb{I}_{Kr}$ and $\phi_{qk}(0) > 0$. Conditions (I1) and (I2), as imposed by Bai and Ng (2002), are widely used in the literature of factor analysis, and condition (I3) is commonly used in FPCA for restricting the eigenfunctions.

We are ready to estimate $\Phi(t)$ and $\boldsymbol{\xi}$. To maintain generality, we make the assumption that the support of s is $[0, 1]$ through appropriate scaling. Define a vector of B-spline basis function $\boldsymbol{M}(\cdot) = (M_1(\cdot), \cdots, M_{\varphi_{T_k}}(\cdot))^T$ on $[0, 1]$, where $\varphi_{T_k} = O(T^v)$ for $k = 0, \cdots, m$, $0 < v < 1/2$. Then we can establish the relationship $\phi_{qk}(s) \approx \boldsymbol{\Theta}_{qk}^T \boldsymbol{M}(s)$. Let $\boldsymbol{\Theta}_q = (\boldsymbol{\Theta}_{q1}, \cdots, \boldsymbol{\Theta}_{qK})^T \in \mathbb{R}^{K \times \varphi_{T_k}}$, which results in $\boldsymbol{\phi}_q(s) \approx \boldsymbol{\Theta}_q \boldsymbol{M}(s)$. With the B-spline approxima-

tion, the identification condition (I3) for $\mathbf{\Phi}(\cdot)$ can be expressed as

$$\varphi_{T_k}^{-1}\mathbf{\Theta}_q\mathbf{\Theta}_q^T = \mathbb{I}_K, q = 1, \cdots, r; \quad \varphi_{T_k}\int \boldsymbol{M}(s)\boldsymbol{M}^T(s)ds = \mathbb{I}_{\varphi_{Tk}}. \tag{2.8}$$

Based on equation (2.6), we can express the function $f_tq(s)$ as follows:

$$f_{tq}(s) = \boldsymbol{M}^T(s)\mathbf{\Theta}_q^T\boldsymbol{\xi}_{tq}. \tag{2.9}$$

On the other hand, multiplying $N^{-1}\boldsymbol{\lambda}_t^T$ on both sides of equation (2.1), we can derive $N^{-1}\boldsymbol{\lambda}_t^T\boldsymbol{X}_t(s) \approx N^{-1}\boldsymbol{\lambda}_t^T\boldsymbol{\lambda}_t f_t(s)$. Consider $N^{-1}\boldsymbol{\lambda}_t^T\boldsymbol{\lambda}_t = \mathbb{I}_r$, we have $\boldsymbol{f}_t(s) = N^{-1}\boldsymbol{\lambda}_t^T\boldsymbol{X}_t(s)$ and

$$f_{tq}(s) = N^{-1}\sum_{i=1}^{N}\lambda_{ipq}X_{it}(s). \tag{2.10}$$

By combining equations (2.8) and (2.9), we can obtain the following equation:

$$N^{-1}\sum_{i=1}^{N}\lambda_{ipq}X_{it}(s_l) = \boldsymbol{M}^T(s_l)\mathbf{\Theta}_q^T\boldsymbol{\xi}_{tq}. \tag{2.11}$$

Then, by multiplying $\boldsymbol{M}(s_l)$ on both sides of equation (2.11) and summing over the observation times, we obtain:

$$(\sum_{l=1}^{n}\boldsymbol{M}(s_l)\boldsymbol{M}^T(s_l))^{-1}[N^{-1}\sum_{l=1}^{n}\boldsymbol{M}(s_l)\sum_{i=1}^{N}\lambda_{ipq}X_{it}(s_l)] = \mathbf{\Theta}_q^T\boldsymbol{\xi}_{tq}.$$

This results in a factor model with factor $\boldsymbol{\xi}_{tq}$, loading $\mathbf{\Theta}_q$ and response

$$\boldsymbol{w}_{tq} = (\sum_{l=1}^{n}\boldsymbol{M}(s_l)\boldsymbol{M}^T(s_l))^{-1}[N^{-1}\sum_{l=1}^{n}\boldsymbol{M}(s_l)\sum_{i=1}^{N}\lambda_{ipq}X_{it}(s_l)],$$

where $\boldsymbol{\lambda}_t = (\boldsymbol{\lambda}_{1t}, \cdots, \boldsymbol{\lambda}_{Nt})^T$ and $\boldsymbol{\lambda}_{it} = (\lambda_{it1}, \cdots, \lambda_{itr})^T$.

Define $\boldsymbol{W}_q = (\boldsymbol{\omega}_t)^T \in \mathbb{R}^{T \times \varphi_{T_k}}$, $\boldsymbol{\xi}_{[q]} = (\boldsymbol{\xi}_{tq})^T \in \mathbb{R}^{T \times K}$. Then we estimate $(\boldsymbol{\Theta}_q, \boldsymbol{\xi}_{[q]})$, denoted by $(\hat{\boldsymbol{\Theta}}_q^T, \hat{\boldsymbol{\xi}}_{[q]})$ the estimator, using the method proposed by Bai and Ng (2013), which are $\hat{\boldsymbol{\Theta}}_q^T = \varphi_{T_k}^{1/2} E_{eigen}[\boldsymbol{W}_q^T \boldsymbol{W}_q; K]$ and $\hat{\boldsymbol{\xi}}_{[q]} = \varphi_{T_k}^{-1} \boldsymbol{W}_q \times \hat{\boldsymbol{\Theta}}_q^T$ for $q = 1, \cdots, r$. Finally, we estimate $\phi_{qk}(s)$ by $\hat{\phi}_{qk}(s) = \hat{\boldsymbol{\Theta}}_{qk}^T \boldsymbol{M}(s)$. In this context, since we estimate score by factor analysis, K can be selected, denoted by $\hat{K}$, by calculating the proportion of variability explained according to the principal components,

$$\min_{q \in \{1, \cdots, r\}} \sum_{k=1}^{\hat{K}} \psi_k(\boldsymbol{W}_q^T \boldsymbol{W}_q) / \sum_{k=1}^{\varphi_{T_k}} \psi_k(\boldsymbol{W}_q^T \boldsymbol{W}_q) > K_{\text{PCA}}\%,$$

where $\psi_k(A)$ is the kth largest eigenvalue of A. Herein, we determined the value of $K_{\text{PCA}}\%$ through simulation, ultimately arriving at a final figure of 97.5%. Further details on this process can be referred to the simulation section in the supplementary material.

3. Asymptotic properties

In this section, we study the asymptotic properties of our estimators. In order to establish the theoretical properties, the following assumptions are required:

Assumption 1. r is finite. For each given $t = 1, \cdots, T$, there exist non-

negative constants C_{min} and C_{max} that satisfy $C_{min} \leq \lim_{n\to\infty} \psi(\sum_{l=1}^{n} \boldsymbol{X}_t(s_l)\boldsymbol{X}_t(s_l)/(Nn)) \leq C_{max}$, where $\psi(A)$ denote the eigenvalues of A.

Assumption 2. (i) $E\|\boldsymbol{f}_t^0(s)\|^4 \leq C$ for each fixed s and $\sum_{s=s_l}^{s_{l'}} \boldsymbol{f}_t^{0T}(s)\boldsymbol{f}_t^0(s)/(l'-l) = \boldsymbol{\Sigma}_F + O_p(n^{-1/2})$ for some positive definite matrix $\boldsymbol{\Sigma}_F$ and for any $t = 1, \cdots, T$.

(ii) $\boldsymbol{\lambda}_{it}$'s are nonrandom such that $\max_{1\leq i\leq N, 1\leq t\leq T} \|\boldsymbol{\lambda}_{it}\| \leq C$ and $\|\boldsymbol{\alpha}_k^{0T}\boldsymbol{\alpha}_k^0/N - \boldsymbol{\Sigma}_k\| \to 0$ for $k = 0, \cdots, m$, where $\boldsymbol{\Sigma}_k$ is an $r \times r$ positive definite matrix.

Assumption 3. (i) Define $\boldsymbol{u}_{it}(\boldsymbol{s}) = (u_{it}(s_1), \cdots, u_{it}(s_n))^T \in \mathbb{R}^{n\times 1}$, $\boldsymbol{u}_t(\boldsymbol{s}) = (\boldsymbol{u}_{1t}(\boldsymbol{s}), \cdots, \boldsymbol{u}_{Nt}(\boldsymbol{s}))^T \in \mathbb{R}^{N\times n}$ and $\boldsymbol{u}_t(\boldsymbol{s}) = \boldsymbol{E}_N^{1/2}\boldsymbol{Q}(\boldsymbol{s})\boldsymbol{F}_n^{1/2}$, where $\boldsymbol{E}_N^{1/2}$ and $\boldsymbol{F}_n^{1/2}$ are the symmetric square roots of $N \times N$ and $n \times n$ positive semi-definite matrices of $\boldsymbol{E}_N$ and $\boldsymbol{F}_n$, respectively. $\boldsymbol{Q}(s) = (q_i(s_j))_{N\times n}$ and $q_i(s_j)$ are independent and identically distributed (i.i.d.) random variables with uniformly bounded moments up to the fourth order. $\psi_1(\boldsymbol{E}_N) < c_1$ and $\psi_1(\boldsymbol{F}_n) < c_1$ uniformly in N and n, respectively. $\psi_n(\boldsymbol{F}_n) > c_2$ for all n.

(ii) $0 < y \equiv \lim_{(N,n)\to\infty} \min(N,n)/\max(N,n) \leq 1$. Let $y^* = \min(y, 1)$. Then, there exists s a real number $c \in (1-y^*, 1]$ such that $\psi_{[c*n]}(\boldsymbol{F}_n) > c_1$ for all n.

Assumption 1 ensures the feasibility of estimating the loading at each time point t, considering the sufficiently large dimension of n. Assumption 2 parallels Assumptions A and B in Bai (2003), guaranteeing that each factor contributes nontrivially to the covariance of $\boldsymbol{X}_t(s)$. Assumption 3 is the same as the assumptions in Ahn and Horenstein (2013) and Onatski (2010). Assumption 3 (i) imposes restrictions on the covariance structure of the error terms. Assumption 3 (ii) holds with $c = 1$ provided that $X_{it}(s)$ are not perfectly multicollinear and none of them have zero idiosyncratic variances.

Assumption 4. (i) $E(u_{it}(s)) = 0$ and $\max_{1\le i\le N, 1\le t\le T} E|u_{it}^4(s)| \le C$ for each fixed s.

(ii) Let $\gamma_u^t(s,s') = N^{-1}E[\boldsymbol{u}_t^T(s)\boldsymbol{u}_t(s')]$, $\gamma_{u,f}^t(s,s') = N^{-1}E[\boldsymbol{f}_t^0(s)\boldsymbol{u}_t^T(s)\boldsymbol{u}_t(s')]$, $\gamma_{u,ff}^t(s,s') = N^{-1}E[\boldsymbol{f}_t^0(s)\boldsymbol{u}_t^T(s)\boldsymbol{u}_t(s')\boldsymbol{f}_t^{0T}(s')]$. For all $t = 1,\cdots,T$, $\max_s \sum_{s'=s_1}^{s_n} \|\gamma(s,s')\| \le C$ and $\max_{s'} \sum_{s=s_1}^{s_n} \|\gamma(s,s')\| \le C$ for $\gamma = \gamma_u^t, \gamma_{uf}^t, \gamma_{uff}^t$ for all $t = 1,\cdots,T$.

(iii) Let $\omega_{ij}^t(s) = E[u_{it}(s)u_{jt}(s)]$, $\omega_{ij}^t(s,s') = E[u_{it}(s)u_{jt}(s')]$. $\max_{1\le j\le N}\sum_{i=1}^N |\omega_{ij}^t(s)| \le C$ for each fixed s and $(Nn)^{-1}\sum_{i,j}\sum_{s,s'}|\omega_{ij}^t(s,s')| \le C$.

(iv) For all $s_1 \le s, s' \le s_n$, let $\nu_{i,j}^t(l,l') = (l'-l)^{-1}\sum_{s=s_l}^{s_{l'}}[u_{it}(s)u_{jt}(s) - E(u_{it}(s)u_{jt}(s))]$. $\max_{1\le i,j\le N} E|(l'-l)^{\frac{1}{2}}\nu_{ij}^t(l,l')|^4 \le C$ for all $s_l \le s_{l'}$ such

that $l' - l \to \infty$.

(v) Let $\zeta_{ss'}^t = N^{-1}[\boldsymbol{u}_t^T(s)\boldsymbol{u}_t(s') - E(\boldsymbol{u}_t^T(s)\boldsymbol{u}_t(s'))]$, $\zeta_{fss'}^t = N^{-1}[\boldsymbol{f}_t^0(s)\boldsymbol{u}_t^T(s)\boldsymbol{u}_t(s') - E(\boldsymbol{f}_t^0(s)\boldsymbol{u}_t^T(s)\boldsymbol{u}_t(s'))]$, $\zeta_{ffss'}^t = N^{-1}[\boldsymbol{f}_t^0(s)\boldsymbol{u}_t^T(s)\boldsymbol{u}_t(s')\boldsymbol{f}_t^{0T}(s') - E(\boldsymbol{f}_t^0(s)\boldsymbol{u}_t^T(s)\boldsymbol{u}_t(s')\boldsymbol{f}_t^{0T}(s'))]$. $\max_{s,s'} E\|\sqrt{N}\zeta^t\| \leq C$ and $E\|N^{-1/2}\boldsymbol{\alpha}_k^{0T}\boldsymbol{u}_t(s)\|^4 \leq C$ for each fixed s for $\zeta^t = \zeta_{ss'}^t, \zeta_{fss'}^t, \zeta_{ffss'}^t$, $t \in I_k$.

(vi) The eigenvalues of the $r \times r$ matrices $\boldsymbol{\Sigma}_k^{1/2}\boldsymbol{\Sigma}_F\boldsymbol{\Sigma}_k^{1/2}$ are distinct for $k = 1, ..., m+1$.

Assumption 4 (i) imposes moment conditions on $u_{it}(s)$, while Assumption 4 (ii)-(v) restricts the cross-sectional and serial dependence among $\{\boldsymbol{u}_t(s), \boldsymbol{f}_t(s)\}$. These conditions are consistent with those imposed in the literature Bai (2003) and Bai and Ng (2006). Assumption 4 (vi) is necessary to ensure the convergence of certain eigenvector estimates.

Theorem 1. *Suppose that Assumptions 1-4 hold with* $1 \leq r \leq r_{\max}$, *we have* $\lim_{(N,n)\to\infty} P(\hat{r} = r) = 1$.

Theorem 1 demonstrates that by maximizing $GR_t(r)$, we can consistently estimate the number of factors at time point t, denoted as $\hat{r}_t$. As mentioned in Remark 1, the ability to estimate the number of factors at each time point is indeed facilitated by the intricate structure of functional

data. Leveraging this complexity allows us to capture the underlying dynamics and dependencies within the data. Given our assumption that the number of factors remains constant, it is reasonable to select the estimated number of factors obtained from the majority of time points. Then we further quantify the distance between adjacent change points when there are no change points.

Theorem 2. *Suppose that Assumptions 1-4 hold. Then for all* $t = 2, \cdots, T$, *if* $t, t-1 \in I_k$, , *it holds that* $\frac{1}{N}\|\hat{\boldsymbol{\lambda}}_t - \hat{\boldsymbol{\lambda}}_{t-1}\|_F^2 = O_p(\eta_{Nn}^{-2})$, *as* $(N, n) \to \infty$.

Theorem 3. *Suppose that Assumptions 1-4 hold. As* $(N, n) \to \infty$, *we have* $\|\hat{\boldsymbol{f}}_t(s) - \boldsymbol{H}_t^{-1}\boldsymbol{f}_t(s)\|_F^2 = O_p(\eta_{Nn}^{-2})$ *for* $t = 1, \cdots, T$, $s = s_1, \cdots, s_n$.

In our setting, factor estimation is conducted separately at each time point based on discretized functional observations. Consequently, the error bounds $O_p(\eta_{Nn}^{-2})$ in Theorem 2 and 3 are consistent with the standard convergence rates $O_p(N^{-1} + T^{-1})$ for factor and loading estimation in approximate factor models (Bai and Ng, 2002; Bai, 2003).

Then we proceed to establish the consistency of change point estimation achieved through the binary segmentation method (WBS). To begin, we rely on the following assumptions:

Assumption 5. (i) The minimum spacing between two adjacent change-

points satisfies $\min_{i=2,\cdots,m}|t'_i - t'_{i-1}| \geq \delta_T$, and the magnitudes $\mu'_i = |\mu_{t'_i} - \mu_{t'_{i-1}}|$ of the jumps satisfy $\min_{i=1,\cdots,m}\mu'_i \geq \tilde{\mu}_{T,N}$, where δ_T and $\tilde{\mu}_{T,N}$ are linked by the requirement $\delta_T^{1/2}\tilde{\mu}_{T,N} \geq C\log^{1/2} T$ for a large enough C and $m \leq C$.

(ii) The number of change points m is bounded by a given non-negative constant $M_{\max}$ and $\Delta_1 > \log^{\alpha} T > \Delta_2$, where $T^{-1}N^{1/2}\Delta_1 = O_p(\eta_{N\delta_T}^{-1} + \tau_1/\delta)$ and $T^{-1}N^{1/2}\Delta_2 = O_p(\eta_{N\delta_T}^{-1})$. Here, we define $\tau_1(t')$ as the length from an undetected change point to the latest detected change point, which means $\tau_1(t') := \min\{|t' - \hat{t}'_k|, |t' - \hat{t}'_{k+1}|\}$ when $t' \in I_{k+1}$. Given that any $\tau_1(t')$ is bounded, we will use τ_1 to replace $\tau_1(t')$ consistently in the subsequent description of the asymptotic properties.

Remark 2. By assuming $\min_{i=2,\cdots,m}|t'_i - t'_{i-1}| \geq \delta_T$, we ensure control over the minimum distance between change points. This enables us to effectively limit the detected change points during the process of empirical testing, thereby mitigating the interference caused by noise. Formally, we express m as $m(T)$ and t'_i as $t'_i(T)$ for $i = 1, ..., m+1$. However, for the sake of brevity and consistency with other studies on change-point detection, we adopt the shorthand notation m and t'_i throughout the remainder of the paper, instead of the longer notation $m(T)$ and $t'_i(T)$. The term $\delta_T^{1/2}\tilde{\mu}_{T,N}$, as stated in Assumption 5, is well-established in the statistical signal de-

tection literature. Similar assumptions are made in the work conducted by Fryzlewicz (2014).

It is important to note that, unlike in Fryzlewicz (2014), we have defined an upper bound for the number of change points in Assumption 5 (ii), which is more realistic in practical applications. Additionally, we have further defined the relationship between N, δ_T, τ_1 and T to ensure the convergence of the sSIC criterion.

Theorem 4. *Suppose that Assumptions 5 hold. Let $m \leq M$, where M is a certain constant independent of T. Let $\alpha > 1$ be a constant such that $\log^{\alpha} T = o(\delta_T \tilde{\mu}_T^2)$. Let the candidate models $\{C(k)\}_{k=1}^{K}$ be produced by the WBS algorithm, and let $\hat{m} = \arg\min_{k=1,\cdots,K} sSIC(k)$. There exist two constants C_1, C_2 such that if $C_1 \log^{\alpha/2} T \leq \Delta_T \leq C_2 \delta_T^{1/2} \tilde{\mu}_{T,N}$, then*

$$P(\mathcal{A}_T) \geq 1 - CT^{-1} - T\delta_T^{-1}(1 - \delta_T^2 T^{-2}/9)^M,$$

where $\mathcal{A}_T = \{\hat{m} = m;\ \max_{i=1,\cdots,m} |\hat{t}_i' - t_i'| \leq C\tilde{\mu}_{T,N}^{-2} \log T\}$.

As indicated by Theorem 4, the convergence rate of the estimated change-point locations in the WBS method relies solely on the minimum jump height $\tilde{\mu}_{T,N}$. Consider the scenario where $\tilde{\mu}_{N,T} > \mu > 0$. In classical settings with a single change point and global signal strength, the optimal localization rate is $O_p(1)$ when the minimal spacing δ_T is of the same order

as T, see Korostelev (1988). In our setting where δ_T can be much smaller than $\log T$, Theorem 4 implies that $P(\exists_i|\hat{t}'_i - t_i| \geq C \ln T) \to 0$, achieving minimax optimal up to a logarithmic factor, which is consistent with the rates established in the WBS literature for multiple change-point problems, (see, e.g., (Fryzlewicz, 2014; Korkas and Fryzlewicz, 2017). When $\tilde{\mu}_{N,T}$ decreases with N at the rate N^{-1}, the localization accuracy necessarily deteriorates, as larger segment lengths are required to ensure detectability. In this case, the dependence on N reflects the intrinsic difficulty caused by diminishing effective signal strength in high-dimensional settings. To obtain more refined bounds, additional assumptions are required, similar to those imposed in Corollary 3 of Wang and Samworth (2018).

To quantify the degree of temporal dependence, we adopt the α-mixing coefficient which is a standard tool commonly used in the time series literature. Recall that a stochastic process $\{Z_t\}_{t\in\mathbb{Z}}$ is said to be α-mixing (strong mixing) if $\alpha(K) = \max_{t\in\mathbb{Z}} \alpha(\sigma(Z_s, s \leq t), \sigma(Z_s, s \geq t + K)) \to 0$ as $K \to \infty$, where we write $\alpha(\mathcal{A}, \mathcal{B}) = \sup_{A\in\mathcal{A}, B\in\mathcal{B}} |P(A \bigcap B) - P(A)P(B)|$ for any two σ-fields $\mathcal{A}$ and $\mathcal{B}$. Assumption 6 concerns the distributions of the functional covariates and the noise sequence.

Assumption 6. The sequence $\{(\boldsymbol{X}_t(s), \boldsymbol{u}_t(s))\}_{t=1}^T$ is stationary for $t \in I_k$, $k = 0, \cdots, m$ and α-mixing with $\sum_{K=1}^{\infty} K^{1/3}\alpha^{1/3}(K) < \infty$.

Assumption 7. Define $\mathcal{H}_\kappa = \{g(\cdot) : |g^{(h)}(x) - g^{(h)}(y)| \leq C|x-y|^k \quad \text{for any } x, y\}$, where $\kappa = h + k$ for $h \in \mathbb{N}_+$ and $k \in (0, 1]$. We suppose the true functions $\{\phi_{qk}(\cdot) : q = 1, \cdots, r; k = 1, \cdots, K\} \in \mathcal{H}_\kappa$ and $\kappa \leq 2$.

Similar assumptions regarding the mixing condition talked in Assumption 6 are posited in the study conducted by Kumar et al. (2024). Specifically, the α-mixing condition necessitates that $\alpha(k) = o(1/k^4)$, thereby enabling the mixing coefficient to gradually diminish at a polynomial rate. Assumption 7 is a regular smoothing condition on the functions.

Theorem 5. *Suppose that Assumptions 1-4 and 6-7 hold. Then as* $(N, T_k) \to \infty$, *it holds that* $\frac{1}{N}\|\hat{\boldsymbol{\alpha}}_k - \boldsymbol{\alpha}_k^0\|_F^2 = O_p(\eta_{NT_k}^{-2})$ *for all* $k = 0, \cdots, m$, $T_k = t'_{k+1} - t'_k$.

Indeed, the convergence rate of $\hat{\boldsymbol{\alpha}}_k$ can be decomposed into two terms: the estimation error $\eta_{NT_k}^{-2}$ and the approximation error $N_0 = 1/nT_k$. However, the latter, arising from the numerical approximation for $\tilde{\Sigma}_{\boldsymbol{X}}$, is negligible since $n > 1$. The convergence rate presented in Theorem 5 bears similarities to those observed in the linear factor model Bai and Ng (2013) and generalized factor model Liu et al. (2023). Building upon the results of Theorem 5, we can further establish the asymptotic normality of the loadings.

Theorem 6. *Suppose that Assumptions 1-4 and 6-7 hold. Then for all* $t \in I_k$, *as* $(N, T_k) \to \infty$, $N^{-2}T_k = o(1)$, *it holds that* $\sqrt{T_k}(\hat{\boldsymbol{\alpha}}_{it} - \boldsymbol{\alpha}_{ik}^0) \xrightarrow{d} N(0, \boldsymbol{\Lambda}_{\boldsymbol{\xi}}^{-1}\boldsymbol{\Psi}_{it}\boldsymbol{\Lambda}_{\boldsymbol{\xi}}^{-1})$, *where* $\boldsymbol{\Lambda}_{\boldsymbol{\xi}} = diag(\sum_{k=1}^K \rho_{1k}, \cdots, \sum_{k=1}^K \rho_{rk})$ *and* $(\boldsymbol{\Psi}_{it})_{r\times r} = E[\int_0^1 u_{it}(s)\Phi^T(s)ds\boldsymbol{\xi}_t\boldsymbol{\xi}_t^T \int_0^1 \Phi(s)u_{it}(s)ds]$.

Indeed, a similar asymptotic normality for the loadings in traditional factor models has been established in Theorem 2 of Bai (2003). Additionally, Wen and Lin (2022)) has also derived similar conclusions regarding the asymptotic distribution of loadings in the context of functional factor models. These findings highlight the robustness and generalizability of the asymptotic results for loadings across different types of factor models.

Theorem 7. *Suppose that Assumptions 1-4 and 6-7 hold. Then for* $t \in I_k$, *it holds that* $\|\hat{\boldsymbol{\xi}}_t - \hat{\boldsymbol{\xi}}_t^0\|_2 = O_p(\varphi_{T_k}^{1/2}\eta_{NT_k}^{-1} + \varphi_{T_k}^{-\kappa})K^{1/2}$.

Theorem 7 provides insights into the samplewise convergence rate of $\hat{\boldsymbol{\xi}}_t$, which comprises two components: the estimation error $K^{1/2}\varphi_{T_k}^{1/2}\eta_{NT_k}$ and the approximation error $K^{1/2}\varphi_{T_k}^{-\kappa}$. When φ_{T_k} is of the order $O_p(1)$ and K is of the order $O_p(1)$, the specification of $\boldsymbol{\Phi}(\cdot)$ is based on finite parameters. As a result, the model (2.7) can be reduced to the traditional factor model, and the estimation error becomes $\eta_{NT_k}^{-1}$. This aligns with the findings of Theorem 1 in Bai and Ng (2002).

Theorem 8. *Suppose that Assumptions 1-4 and 6-7 hold. Then for* $t \in I_k$, *we have* $\sum_{k=1}^{K} \int (\hat{\phi}_{qk}(s) - \phi_{qk}(s))^2 ds = O_p(\varphi_{T_k} \eta_{NT_k}^{-2} + \varphi_{T_k}^{-2\kappa})$.

Theorem 8 establishes the convergence rate for eigenfunctions $\hat{\phi}(s)$. This convergence rate is also composed of two terms: the estimation error $\varphi_{T_k} \eta_{NT_k}^{-2}$ and the approximation error $\varphi_{T_k}^{-2\kappa}$.

4. Application

In this section, we illustrate the proposed model and statistical inference methods by analyzing the following real-data application. Our dataset is sourced from `https://air.cnemc.cn:18007` and encompasses data on pollution emissions from 171 monitoring stations in Beijing, Tianjin, and 31 other cities. The dataset covers the period from March 2018 to February 2020, with data available for every 24-hour interval. For our analyses, we specifically focus on complete four-season periods, where spring spans from March to May, summer from June to August, autumn from September to November, and winter from December to February of the following year.

Due to missing data values, the dataset for the period from March 2018 to February 2019 consists of 360 days and is denoted as dataset1, resulting in N=171, T=360, and n=24. Similarly, the dataset for the period from March 2019 to February 2020 consists of 365 days and is denoted as dataset2,

resulting in N=171, T=365, and n=24. These cities are concentrated in the northern region of China, where atmospheric pollution is relatively severe due to the influence of climate, industrial composition, and topography.

Table 1 displays the number of change points and their locations for major pollution indicators in 2018 and 2019. It can be observed that most indicators have 1-3 change points within a year, which aligns with their respective seasonal characteristics. Remarkably, we can observe that for all three indicators, namely AQI, PM2.5, and PM10, we have identified change points on January 16, 2019, and November 25, 2019. These specific change points may be attributed to anthropogenic factors, given their timing and potential correlation with human activities.

Table 1: Number and Locations of Change Points for Pollution Indicators

Type	Year	$\hat{m}$	$\hat{\boldsymbol{t}}$		
AQI	2018	2	2018-03-29	**2019-01-16**	/
	2019	1	**2019-11-25**	/	/
PM2.5	2018	2	2018-10-27	**2019-01-16**	/
	2019	2	**2019-11-25**	2019-12-31	/
PM10	2018	3	2018-03-29	2018-04-06	**2019-01-16**
	2019	2	2019-10-05	**2019-11-25**	/

The influence of climate factors will be depicted in the figure provided

in the Supplementary Material. Here, we focus on anthropogenic influences, which also play a crucial role in shaping pollutant levels and can be more significantly observed through our method. Apart from the long-term control of atmospheric pollution and promotion of air purification through the implementation of relevant laws and regulations in China, effective and substantial reductions in pollutant emissions can also be achieved in the short term through the implementation of specific policies such as traffic restrictions and emission controls. The reasons for implementing such short-term controls can vary and may include factors such as hosting large international conferences or trial runs of new emission reduction initiatives.

As depicted in Figure 1, taking one of the change points of PM2.5, specifically January 16, 2019, as an example, it is evident that the daily concentration of PM2.5 significantly decreases compared to the five days preceding the change point. This decrease in PM2.5 concentration persists at least for the five days following the change point. Additionally, there is a noticeable decline in pollutant concentration during the afternoon hours, indicating a change in the trend of the daily pollutant concentration curve. This suggests that short-term anthropogenic factors, beyond seasonal influences, have played a role. The identified change points effectively capture these variations, providing valuable insights into the impact of short-term

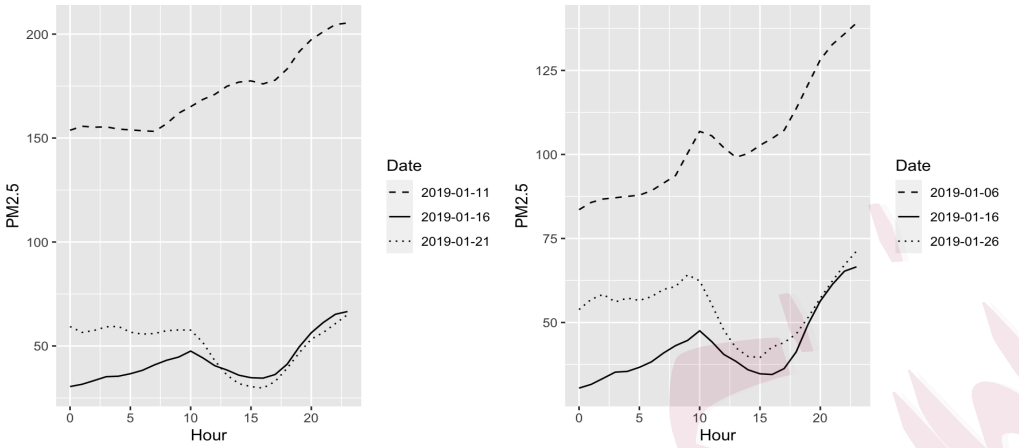

Figure 1: The left and right plots show the daily average PM2.5 concentration curves for the change point and the 5 days and 10 days before and after it, respectively. The change point day is represented by a solid line, with the days before the change point shown by a dashed line and the days after the change point shown by a dotted line.

intervention measures on pollutant levels.

However, it is important to note that short-term intervention measures are not sustainable in the long run, and pollutant levels may rebound after a certain period. Nevertheless, this analysis provides valuable information for improving long-term pollution control measures. It highlights the importance of reducing pollutant emissions in the long term and implementing effective measures for faster regulation of pollutant concentrations. This becomes a critical aspect in pollution prevention and control efforts.

5. Conclusion

This paper presents a novel approach for estimating the number and locations of structural breaks in functional time series factor model. By leveraging the factor model framework, we effectively capture the interplay between functional factors and temporal dimensions in functional data. Moreover, we exploit the intricate structure of functional data to propose a more practical method for identifying structural breaks in factor models. This approach offers a valuable tool for researchers in related fields, facilitating a better understanding of changing patterns in functional time series data. Furthermore, the approach proposed in this paper fully utilizes the intricate structure of functional data. It can be applied to other related models such as functional factor regression. This highlights the versatility of the methodology and its potential applications beyond the scope of this study.

Acknowledgments

The authors would like to thank the Editor, the Associate Editor and two anonymous referees for their constructive suggestions that substantially improved this paper. This work was supported by the 111-Center Project of China (B25066), the Innovative Research Team of SUFE, the NSFC (12271329, 72331005), the Shanghai Research Center for Data Science and

Decision Technology, the Fundamental Research Funds for the Central Universities (CXJJ-2024-455), Humanities and Social Sciences Fund of Ministry of Education 23YJA910003.

Supplementary Materials

The online Supplementary Material contains the numerical studies, additional results of the application, lemmas and technical proofs.

References

Ahn, S. C. and A. R. Horenstein (2013). Eigenvalue ratio test for the number of factors. *Econometrica 81*(3), 1203–1227.

Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica: Journal of the Econometric Society*, 821–856.

Ash, R. B. and M. F. Gardner (2014). *Topics in Stochastic Processes: Probability and Mathematical Statistics: A Series of Monographs and Textbooks*, Volume 27. Academic press.

Aston, J. A. and C. Kirch (2012). Detecting and estimating changes in dependent functional data. *Journal of Multivariate Analysis 109*, 204–220.

Aue, A., G. Rice, and O. Sönmez (2018). Detecting and dating structural breaks in functional data without dimension reduction. *Journal of the Royal Statistical Society Series B: Statistical Methodology 80*(3), 509–529.

REFERENCES

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*(1), 135–171.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221.

Bai, J. and S. Ng (2006). Determining the number of factors in approximate factor models, errata. *manuscript, Columbia University*.

Bai, J. and S. Ng (2013). Principal components estimation and identification of static factors. *Journal of Econometrics 176*(1), 18–29.

Berkes, I., R. Gabrys, L. Horváth, and P. Kokoszka (2009). Detecting changes in the mean of functional observations. *Journal of the Royal Statistical Society Series B: Statistical Methodology 71*(5), 927–946.

Breitung, J. and S. Eickmeier (2011). Testing for structural breaks in dynamic factor models. *Journal of Econometrics 163*(1), 71–84.

Chang, J., C. Chen, X. Qiao, and Q. Yao (2024). An autocovariance-based learning framework for high-dimensional functional time series. *Journal of Econometrics 239*(2), 105385.

Chen, L., J. J. Dolado, and J. Gonzalo (2014). Detecting big structural breaks in large factor models. *Journal of Econometrics 180*(1), 30–48.

Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. *The Annals of Statistics 42*(6), 2243–2281.

Gao, Y., H. L. Shang, and Y. Yang (2021). Factor-augmented smoothing model for functional data. *arXiv preprint arXiv:2102.02580*.

Guo, S., X. Qiao, and Q. Wang (2021). Factor modelling for high-dimensional functional time series. *arXiv preprint arXiv:2112.13651*.

Hallin, M., G. Nisol, and S. Tavakoli (2023). Factor models for high-dimensional functional time series i: Representation results. *Journal of Time Series Analysis 44*(5-6), 578–600.

Horváth, L., P. Kokoszka, and G. Rice (2014). Testing stationarity of functional time series. *Journal of Econometrics 179*(1), 66–82.

Hu, X. and F. Yao (2022). Sparse functional principal component analysis in high dimensions. *Statistica Sinica 32*(4), 1939–1960.

Hu, X. and F. Yao (2024). Dynamic principal component analysis in high dimensions. *Journal of the American Statistical Association 119*(545), 308–319.

Kokoszka, P., H. Miao, and X. Zhang (2015). Functional dynamic factor model for intraday price curves. *Journal of Financial Econometrics 13*(2), 456–477.

Korkas, K. K. and P. Fryzlewicz (2017). Multiple change-point detection for non-stationary time series using wild binary segmentation. *Statistica Sinica*, 287–311.

Korostelev, A. (1988). On minimax estimation of a discontinuous signal. *Theory of Probability & Its Applications 32*(4), 727–730.

Kumar, S., H. Xu, H. Cho, and D. Wang (2024). Estimation and inference for change points in

functional regression time series. *arXiv preprint arXiv:2405.05459*.

Li, D., R. Li, and H. L. Shang (2023). Detection and estimation of structural breaks in high-dimensional functional time series. *arXiv preprint arXiv:2304.07003*.

Liu, W., H. Lin, S. Zheng, and J. Liu (2023). Generalized factor model for ultra-high dimensional correlated variables with mixed types. *Journal of the American Statistical Association 118*(542), 1385–1401.

Ma, S. and L. Su (2018). Estimation of large dimensional factor models with an unknown number of breaks. *Journal of Econometrics 207*(1), 1–29.

Madrid Padilla, C. M., D. Wang, Z. Zhao, and Y. Yu (2022). Change-point detection for sparse and dense functional data in general dimensions. *Advances in Neural Information Processing Systems 35*, 37121–37133.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics 92*(4), 1004–1016.

Stock, J. H. and M. Watson (2009). Forecasting in dynamic factor models subject to structural instability. *The Methodology and Practice of Econometrics. A Festschrift in Honour of David F. Hendry 173*, 205.

Su, L. and X. Wang (2017). On time-varying factor models: Estimation and testing. *Journal of Econometrics 198*(1), 84–101.

Wang, T. and R. J. Samworth (2018). High dimensional change point estimation via sparse

projection. *Journal of the Royal Statistical Society Series B: Statistical Methodology 80*(1), 57–83.

Wen, S. and H. Lin (2022). Factor-guided functional pca for high-dimensional functional data. *arXiv preprint arXiv:2211.12012*.

Yamamoto, Y. and S. Tanaka (2015). Testing for factor loading structural change under common breaks. *Journal of Econometrics 189*(1), 187–206.

Zhang, X., X. Shao, K. Hayhoe, and D. J. Wuebbles (2011). Testing the structural stability of temporally dependent functional observations and application to climate projections. *Electronic Journal of Statistics 5*, 1765–1796.

School of Statistics and Data Science, Shanghai University of Finance and Economics, China

E-mail: xucaixia@163.sfue.edu.cn

School of Statistics and Data Science, Ningbo University of Technology, China

E-mail: suhuacheng79@163.com

School of Statistics and Data Science, Shanghai University of Finance and Economics, China

E-mail: liu.xu@sufe.edu.cn

School of Statistics and Data Science, Shanghai University of Finance and Economics, China

E-mail: johnyou07@163.com