

**Matteo Barigozzi, Haeran Cho and Piotr Fryzlewicz**  
**Simultaneous multiple change-point and  
factor analysis for high-dimensional time  
series**

**Article (Published version)  
(Refereed)**

**Original citation:**

Barigozzi, Matteo and Cho, Haeran and Fryzlewicz, Piotr (2018) *Simultaneous multiple change-point and factor analysis for high-dimensional time series*. [Journal of Econometrics](#). ISSN 0304-4076 (In Press)

DOI: [10.1016/j.jeconom.2018.05.003](https://doi.org/10.1016/j.jeconom.2018.05.003)

© 2018 the Authors  
CC BY 4.0

This version available at: <http://eprints.lse.ac.uk/88110/>  
Available in LSE Research Online: July 2018

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.



Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: [www.elsevier.com/locate/jeconom](http://www.elsevier.com/locate/jeconom)

# Simultaneous multiple change-point and factor analysis for high-dimensional time series

Matteo Barigozzi<sup>a</sup>, Haeran Cho<sup>b,\*</sup>, Piotr Fryzlewicz<sup>a</sup>

<sup>a</sup> Department of Statistics, London School of Economics, Houghton Street, London WC2A 2AE, UK

<sup>b</sup> School of Mathematics, University of Bristol, University Walk, Bristol, BS8 1TW, UK

## ARTICLE INFO

### Article history:

Received 16 February 2017

Received in revised form 2 February 2018

Accepted 28 May 2018

Available online xxxx

### Keywords:

Piecewise stationary factor model

Change-point detection

Principal component analysis

Wavelet transformation

Double CUSUM binary segmentation

## ABSTRACT

We propose the first comprehensive treatment of high-dimensional time series factor models with multiple change-points in their second-order structure. We operate under the most flexible definition of piecewise stationarity, and estimate the number and locations of change-points consistently as well as identifying whether they originate in the common or idiosyncratic components. Through the use of wavelets, we transform the problem of change-point detection in the second-order structure of a high-dimensional time series, into the (relatively easier) problem of change-point detection in the means of high-dimensional panel data. Also, our methodology circumvents the difficult issue of the accurate estimation of the true number of factors in the presence of multiple change-points by adopting a screening procedure. We further show that consistent factor analysis is achieved over each segment defined by the change-points estimated by the proposed methodology. In extensive simulation studies, we observe that factor analysis prior to change-point detection improves the detectability of change-points, and identify and describe an interesting ‘spillover’ effect in which substantial breaks in the idiosyncratic components get, naturally enough, identified as change-points in the common components, which prompts us to regard the corresponding change-points as also acting as a form of ‘factors’. Our methodology is implemented in the R package `factorcpt`, available from CRAN.

© 2018 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

High-dimensional time series data abound in modern data science, including finance (e.g., simultaneously measured returns on a large number of assets (Fan et al., 2011b; Barigozzi and Hallin, 2017)), economics (e.g., country-level macroeconomic data (Stock and Watson, 2009) or retail price index data (Groen et al., 2013)), neuroimaging (e.g., measurements of brain activity (Schröder and Ombao, 2016; Barnett and Onnela, 2016)) and biology (e.g., transcriptomics data (Omranian et al., 2015) or Hi-C data matrices (Brault et al., 2016)).

Factor modelling, in which the individual elements of a high-dimensional time series are modelled as sums of a common component (a linear combination of a small number of possibly unknown factors), plus each individual element’s own idiosyncratic noise, is a well-established technique for dimension reduction in time series. Time series factor models are classified, in relation to the effect of factors on the observed time series, into ‘static’ (only a contemporaneous effect, see e.g., (Stock and Watson, 2002; Bai and Ng, 2002; Bai, 2003)) or ‘dynamic’ (lagged factors may also have an effect, see e.g., (Forni and Lippi, 2001; Hallin and Lippi, 2013)) factor models.

\* Corresponding author.

E-mail addresses: [m.barigozzi@lse.ac.uk](mailto:m.barigozzi@lse.ac.uk) (M. Barigozzi), [haeran.cho@bristol.ac.uk](mailto:haeran.cho@bristol.ac.uk) (H. Cho), [p.fryzlewicz@lse.ac.uk](mailto:p.fryzlewicz@lse.ac.uk) (P. Fryzlewicz).

<https://doi.org/10.1016/j.jeconom.2018.05.003>

0304-4076/© 2018 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

It is increasingly recognised that in several important application areas, such as those mentioned at the beginning of this section, nonstationary time series data are commonly observed. Arguably the simplest realistic departure from stationarity, which also leads to sparse and interpretable time series modelling, is piecewise-stationarity, in which the time series is modelled as approximately stationary between neighbouring change-points, and changing its distribution (e.g., the mean or covariance structure) at each change-point.

The main aim of this work is to provide the first comprehensive framework for the estimation of time series factor models with multiple change-points in their second-order structure. The existing literature on time series factor modelling has only partially embraced nonstationarity. One way in which a change-point is typically handled in the literature is via the assumption that the structural break in the loadings is ‘moderate’ and it affects only a limited number of series, so that it does not adversely impact the quality of traditional stationary principal component analysis (PCA)-type estimation (Stock and Watson, 2002, 2009; Bates et al., 2013).

However, opinions diverge on the empirically relevant degrees of temporal instability in the factor loadings, and several authors observe that ‘large’ changes in the stochastic data-generating mechanism have the potential to severely distort the estimation of the factor structure. Investigations into the effect of a *single* break in the loadings or the number of factors on the factor structure, with accompanying change-point tests and estimators for its location, can be found in Breitung and Eickmeier (2011), Chen et al. (2014), Han and Inoue (2014), Corradi and Swanson (2014), Yamamoto and Tanaka (2015), Baltagi et al. (2017), Bai et al. (2017) and Massacci (2017). Lasso-type estimation is considered for change-point analysis under factor modelling in Cheng et al. (2016) and Ma and Su (2016): the former concerns single change-point detection in the loadings and the number of factors, while the latter considers multiple change-point detection in loadings only. Note that the  $\ell_1$ -penalty of the Lasso is not optimal for change-point detection, as investigated in Brodsky and Darkhovsky (1993) and Cho and Fryzlewicz (2011). In summary, apart from Ma and Su (2016) and Sun et al. (2016) (the latter considers factor models with multiple change-points but for a small number of time series only), the existing change-point methods proposed for factor models focus on detecting a single break of a particular type, namely a break in the loadings or the number of factors.

We now describe in detail the contribution and findings of this work, at the same time giving an overview of the organisation of the paper.

- (a) *We propose a comprehensive methodology for the consistent estimation of multiple change-points in the second-order structure of a high-dimensional time series governed by a factor model.* This is in contrast to the substantial time series factor model literature, which is overwhelmingly concerned with testing for a single change-point. In practice, the possibility of the presence of multiple change-points cannot be ruled out from a dataset consisting of observations over a long stretch of time, as illustrated in our applications to financial and macroeconomic time series data in Section 8. Our estimators are ‘interpretable’ in the sense that they enable the identification of whether each change-point originates from the common or idiosyncratic components. Through simulation studies (Section 7), it is demonstrated that in high-dimensional time series segmentation, factor analysis prior to change-point detection improves the detectability of change-points that appear only in either of the common or the idiosyncratic components.
- (b) *We operate under the most flexible definition of piecewise-stationarity, embracing all possible structural instabilities under factor modelling:* it allows factors and idiosyncratic components to have unrestricted piecewise-stationary second-order structures, including changes in their autocorrelation structures (Section 2).
- (c) *We derive a uniform convergence rate for the PCA-based estimator of the common components under the factor model with multiple change-points in Theorems 1 and 2 (Section 3.1).* A key to the derivation of the theoretical results is the introduction of the ‘capped’ PCA estimator of the common components, which controls for the possible contribution of spurious factors to individual common components even when the number of factors is over-specified.
- (d) *Through the use of wavelets, we transform the problem of change-point detection in the second-order structure of a high-dimensional time series, into the (relatively easier) problem of change-point detection in the means of high-dimensional panel data.* More specifically, in the first stage of the proposed methodology, we decompose the observed time series into common and idiosyncratic components, and we compute wavelet transforms for each component (separately), see e.g., Nason et al. (2000), Fryzlewicz and Nason (2006) and Van Bellegem and von Sachs (2008) for the use of locally stationary wavelet models for time series data. In this way, any change-point in the second-order structure of common or idiosyncratic components is detectable as a change-point in the means of the wavelet-transformed series (Sections 3.2 and 3.3).
- (e) *Each of the panels of transformed common and idiosyncratic components serves as an input to the second stage of our methodology, which requires an algorithm for multiple change-point detection in the means of high-dimensional panel data.* For this, a number of methodologies have been investigated in the literature, such as Horváth and Hušková (2012), Enikeeva and Harchaoui (2015), Jirak (2015), Cho and Fryzlewicz (2015) and Wang and Samworth (2018). In Section 4.1, we adopt the Double CUSUM Binary Segmentation procedure proposed in Cho (2016), which achieves consistency in estimating the total number and locations of the multiple change-points while permitting both within-series and cross-sectional correlations. In Section 4.2, we prove that this consistency result carries over to the consistency in multiple change-point detection in the common and idiosyncratic components, as the dimensions of the data,  $n$  and  $T$ , diverge (Theorem 3).
- (f) *Motivated by the theoretical finding noted in (c), our methodology is equipped with a step that screens the results of change-point analysis over a range of factor numbers employed for the estimation of common components*

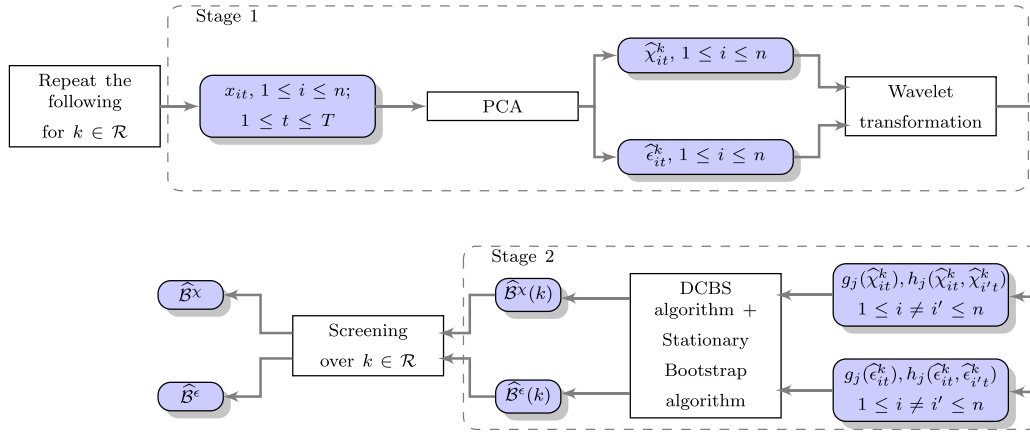


Fig. 1. Flowchart of the proposed methodology.

(Section 4.3). It enables us to circumvent the challenging problem of accurately estimating the true number of factors in contrast to much of the existing literature, and plays the key role for the proposed methodology to achieve consistent change-point estimation.

- (g) Once all the change-points are estimated, we show that the common components are consistently estimated via PCA on each estimated segment (Section 5). Such result is new to the literature on factor models with structural breaks.
- (h) We identify and describe an interesting ‘spillover’ effect in which change-points attributed to large breaks in the second-order structure of the idiosyncratic components get (seemingly falsely) identified as change-points in the common components (Section 7). We argue that this phenomenon is only too natural and expected, and can be ascribed to the prominent change-points playing the role of ‘common factors’, regardless of whether they originate in the common or idiosyncratic components. We refer to this new point of view as *regarding change-points as factors*.
- (i) We provide an R package named `factorcpt`, which implements our methodology. The package is available from CRAN (Cho et al., 2016).

## An overview of the proposed methodology

We provide an overview of the change-point detection methodology with an accompanying flowchart in Fig. 1. For each step, we provide a reference to the relevant section in the paper.

Input: time series  $\{x_{it}, i = 1, \dots, n, t = 1, \dots, T\}$  and a set of factor number candidates  $\mathcal{R} = \{\underline{r}, \underline{r} + 1, \dots, \bar{r}\}$ .

Iteration: repeat Stages 1–2 for all  $k \in \mathcal{R}$ .

Stage 1: Factor analysis and wavelet transformation. PCA:  $x_{it}$  is decomposed into the common component estimated with  $k$  factors ( $\hat{x}_{it}^k$ ) and the idiosyncratic component ( $\hat{e}_{it}^k$ ) via PCA (Section 3.1).

Wavelet transformation:  $\hat{x}_{it}^k$  and  $\hat{e}_{it}^k$  are separately transformed into panels with (almost) piecewise constant signals via wavelet-based transformations  $g_j(\cdot)$  and  $h_j(\cdot, \cdot)$  (Section 3.3).

Stage 2: High-dimensional panel data segmentation. The joint application of the Double CUSUM Binary Segmentation (Section 4.1) and stationary bootstrap (Section 6.1) algorithms on the panels obtained in Stage 1, returns the sets of change-points detected for the common ( $\hat{B}^x(k)$ ) and idiosyncratic ( $\hat{B}^e(k)$ ) components.

Screening over  $k \in \mathcal{R}$ : the sets  $\hat{B}^x(k)$  are screened to obtain the final estimates  $\hat{B}^x$  and  $\hat{B}^e$  (Section 4.3).

## Notation

For a given  $m \times n$  matrix  $\mathbf{B}$  with  $b_{ij} = [\mathbf{B}]_{i,j}$  denoting its  $(i, j)$  element, its spectral norm is defined as  $\|\mathbf{B}\| = \sqrt{\mu_1(\mathbf{B}\mathbf{B}^\top)}$ , where  $\mu_k(\mathbf{C})$  denotes the  $k$ th largest eigenvalue of  $\mathbf{C}$ , and its Frobenius norm as  $\|\mathbf{B}\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n b_{ij}^2}$ . For a given set  $\Pi$ , we denote its cardinality by  $|\Pi|$ . The indicator function is denoted by  $\mathbb{I}(\cdot)$ . Also, we use the notations  $a \vee b = \max(a, b)$  and  $a \wedge b = \min(a, b)$ . Besides,  $a \sim b$  indicates that  $a$  is of the order of  $b$ , and  $a \gg b$  that  $a^{-1}b \rightarrow 0$ . We denote an  $n \times n$ -matrix of zeros by  $\mathbf{O}_n$ .

## 2. Piecewise stationary factor model

In this section, we define a piecewise stationary factor model which provides a framework for developing our change-point detection methodology. Throughout, we assume that we observe an  $n$ -dimensional vector of time series  $\mathbf{x}_t = (x_{1t}, \dots, x_{nt})^\top$ , following a factor model and undergoing an unknown number of change-points in its second-order structure.

The location of the change-points is also unknown. Each of element  $x_{it}$  can be written as:

$$x_{it} = \chi_{it} + \epsilon_{it} = \lambda_i^\top \mathbf{f}_t + \epsilon_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T, \quad (1)$$

where  $\mathbf{x}_t = (\chi_{1t}, \dots, \chi_{nt})^\top$  and  $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \dots, \epsilon_{nt})^\top$  denote the common and the idiosyncratic components of  $\mathbf{x}_t$ , with  $E(\chi_{it}) = E(\epsilon_{it}) = 0$  for all  $i, t$ . We refer to  $r$  as the number of factors in model (1). Then the common components are driven by the  $r$  factors  $\mathbf{f}_t = (f_{1t}, \dots, f_{rt})^\top$  with  $E(\mathbf{f}_t) = \mathbf{0}$ , and  $\lambda_i = (\lambda_{i1}, \dots, \lambda_{ir})^\top$  denotes the  $r$ -dimensional vector of loadings. We denote the  $n \times r$  matrix of factor loadings by  $\mathbf{A} = [\lambda_1, \dots, \lambda_n]^\top$ . The change-points in the second-order structure of  $\mathbf{x}_t$  are classified into  $\mathcal{B}^\chi = \{\eta_1^\chi, \dots, \eta_{B_\chi}^\chi\}$ , those in the common components, and  $\mathcal{B}^\epsilon = \{\eta_1^\epsilon, \dots, \eta_{B_\epsilon}^\epsilon\}$ , those in the idiosyncratic components.

**Remark 1.** In the factor model (1), both the loadings and the factor number are time-invariant. However, it is well established in the literature on factor models with structural breaks, that any change in the number of factors or loadings can be represented by a factor model with stable loadings and time-varying factors. In doing so,  $r$ , the dimension of the factor space under (1), satisfies  $r \geq r_b$ , where  $r_b$  denotes the minimum number of factors of each segment  $[\eta_b^\chi + 1, \eta_{b+1}^\chi]$ ,  $b = 0, \dots, B_\chi$ , over which the common components remain stationary. We provide a comprehensive example illustrating this point in [Appendix A](#). Many tests and estimation methods for changes in the loadings rely on this equivalence between the two representations of time-varying factor models, see e.g., [Han and Inoue \(2014\)](#), [Chen et al. \(2014\)](#). Therefore, we work with the representation in (1), where all change-points in the common components are imposed on the second-order structure of  $\mathbf{f}_t$  as detailed in [Assumption 1](#). At the same time, we occasionally speak of changes in the loadings or the number of factors, referring to those changes in  $\mathbf{f}_t$  that can be ascribed back to such changes.

We denote by  $\beta(t) = \max\{0 \leq b \leq B_\chi : \eta_b^\chi + 1 \leq t\}$  the index of the change-point in  $\chi_{it}$  that is nearest to, and strictly to the left of  $t$ , and by  $\bar{\eta}^\chi(t) = \max\{\eta_b^\chi : \eta_b^\chi + 1 \leq t, b = 0, \dots, B_\chi\}$  the latest change-point in  $\chi_{it}$  strictly before  $t$ , with the notational convention  $\bar{\eta}_0^\chi = 0$  and  $\eta_{B_\chi+1}^\chi = T$ . Similarly,  $\eta^\epsilon(t)$  is defined with respect to  $\eta_b^\epsilon \in \mathcal{B}^\epsilon$ , and  $\gamma(t)$  denotes the index of the change-point in  $\epsilon_{it}$  that is nearest to  $t$  while being strictly to its left. Then, we impose the following conditions on  $\mathbf{f}_t$  and  $\boldsymbol{\epsilon}_t$ .

**Assumption 1.**

- (i) There exist weakly stationary processes  $\mathbf{f}_t^b = (f_{1t}^b, \dots, f_{rt}^b)^\top$  associated with the intervals  $[\eta_b^\chi + 1, \eta_{b+1}^\chi]$  for  $b = 0, \dots, B_\chi$  such that  $E(f_{jt}^b) = 0$  for all  $j, t$ , and

$$\max_{1 \leq j \leq r} E \left\{ f_{jt} - f_{jt}^{\beta(t)} \right\}^2 = O \left( \rho_f^{t - \bar{\eta}^\chi(t)} \right)$$

for some fixed  $\rho_f \in [0, 1)$ .

- (ii) Let  $\mathbf{x}_t^b = (\chi_{1t}^b, \dots, \chi_{nt}^b)^\top = \mathbf{A} \mathbf{f}_t^b$  and denote the (auto)covariance matrix of  $\mathbf{x}_t^b$  by  $\tilde{\mathbf{T}}_\chi^b(\tau) = E\{\mathbf{x}_{t+\tau}^b (\mathbf{x}_t^b)^\top\} = \mathbf{A} E\{\mathbf{f}_{t+\tau}^b (\mathbf{f}_t^b)^\top\} \mathbf{A}^\top$ . Then, there exists fixed  $\bar{\tau}_\chi < \infty$  such that for any  $b = 1, \dots, B_\chi$ , we have  $\tilde{\mathbf{T}}_\chi^b(\tau) - \tilde{\mathbf{T}}_\chi^{b-1}(\tau) \neq \mathbf{0}_n$  for some  $|\tau| \leq \bar{\tau}_\chi$ .

- (iii) There exist weakly stationary processes  $\boldsymbol{\epsilon}_t^b = (\epsilon_{1t}^b, \dots, \epsilon_{nt}^b)^\top$  associated with the intervals  $[\eta_b^\epsilon + 1, \eta_{b+1}^\epsilon]$  for  $b = 0, \dots, B_\epsilon$  such that  $E(\epsilon_{it}^b) = 0$  for all  $i, t$ , and

$$\max_{1 \leq i \leq n} E \left\{ \epsilon_{it} - \epsilon_{it}^{\gamma(t)} \right\}^2 = O \left( \rho_\epsilon^{t - \bar{\eta}^\epsilon(t)} \right)$$

for some fixed  $\rho_\epsilon \in [0, 1)$ .

- (iv) Denote the (auto)covariance matrix of  $\boldsymbol{\epsilon}_t^b$  by  $\tilde{\mathbf{T}}_\epsilon^b(\tau) = E\{\boldsymbol{\epsilon}_{t+\tau}^b (\boldsymbol{\epsilon}_t^b)^\top\}$ . Then, there exists fixed  $\bar{\tau}_\epsilon < \infty$  such that for any  $b = 1, \dots, B_\epsilon$ , we have  $\tilde{\mathbf{T}}_\epsilon^b(\tau) - \tilde{\mathbf{T}}_\epsilon^{b-1}(\tau) \neq \mathbf{0}_n$  for some  $|\tau| \leq \bar{\tau}_\epsilon$ .

**Assumption 1**(i)–(ii) indicates that for each  $b = 0, \dots, B_\chi$ , the common component  $\chi_{it}$  is ‘close’ to a stationary process  $\chi_{it}^b$  over the segment  $[\eta_b^\chi + 1, \eta_{b+1}^\chi]$  such that the effect of transition from one segment to another diminishes at a geometric rate as  $t$  moves away from  $\eta_b^\chi$ , and any  $\eta_b^\chi \in \mathcal{B}^\chi$  coincides with a change-point in the autocovariance or cross-covariance matrices of  $\mathbf{x}_t^{\beta(t)}$ . The same arguments apply to  $\epsilon_{it}$  under [Assumption 1](#) (iii)–(iv). The treatment of such approximately piecewise stationary  $\mathbf{f}_t$  and  $\boldsymbol{\epsilon}_t$  is similar to that in [Fryzlewicz and Subba Rao \(2014\)](#).

Note that the literature on factor models with structural breaks have primarily focused on the case of a single change-point in the loadings or factor number see e.g., [Breitung and Eickmeier \(2011\)](#), [Chen et al. \(2014\)](#) and [Han and Inoue \(2014\)](#). We emphasise that to the best of our knowledge, the model (1) equipped with [Assumption 1](#) is the first one to offer a comprehensive framework that allows for multiple change-points that are not confined to breaks in the loadings or the emergence of a new factor, but also includes breaks in the second-order structure of the factors and idiosyncratic components.

We now list and motivate the assumptions imposed on the piecewise stationary factor model see e.g., [Stock and Watson \(2002\)](#), [Bai and Ng \(2002\)](#), [Bai \(2003\)](#), [Forni et al. \(2009\)](#) and [Fan et al. \(2013\)](#) for similar conditions on stationary factor models.



**Assumption 2.**

- (i)  $T^{-1} \sum_{t=1}^T E(\mathbf{f}_t \mathbf{f}_t^\top) = \mathbf{I}_r$ .
- (ii) There exists some fixed  $c_f, \beta_f \in (0, \infty)$  such that, for any  $z > 0$  and  $j = 1, \dots, r$ ,
 
$$P(|f_{jt}| > z) \leq \exp\{1 - (z/c_f)^{\beta_f}\}.$$

**Assumption 3.**

- (i) There exists a positive definite  $r \times r$  matrix  $\mathbf{H}$  with distinct eigenvalues, such that  $n^{-1} \mathbf{A}^\top \mathbf{A} \rightarrow \mathbf{H}$  as  $n \rightarrow \infty$ .
- (ii) There exists  $\tilde{\lambda} \in (0, \infty)$  such that  $|\lambda_{ij}| < \tilde{\lambda}$  for all  $i, j$ .

**Assumption 4.**

- (i) There exists  $C_\epsilon \in (0, \infty)$  such that
 
$$\max_{1 \leq s \leq e \leq T} \left\{ \frac{1}{e-s+1} \sum_{i,i'=1}^n \sum_{t,t'=s}^e a_i a_{i'} E(\epsilon_{it} \epsilon_{i't'}) \right\} < C_\epsilon,$$
 for any sequence of coefficients  $\{a_i\}_{i=1}^n$  satisfying  $\sum_{i=1}^n a_i^2 = 1$ .
- (ii)  $\epsilon_{it}$  are normally distributed.

**Assumption 5.**

- (i)  $\{\mathbf{f}_t, 1 \leq t \leq T\}$  and  $\{\epsilon_t, 1 \leq t \leq T\}$  are independent.
- (ii) Denoting the  $\sigma$ -algebra generated by  $\{(\mathbf{f}_t, \epsilon_t), s \leq t \leq e\}$  by  $\mathcal{F}_s^e$ , let

$$\alpha(k) = \max_{1 \leq t \leq T} \sup_{\substack{A \in \mathcal{F}_{t-k}^t \\ B \in \mathcal{F}_{t+k}^\infty}} |P(A)P(B) - P(A \cap B)|.$$

Then, there exists some fixed  $c_\alpha, \beta \in (0, \infty)$  satisfying  $3\beta_f^{-1} + \beta^{-1} > 1$ , such that, for all  $k \in \mathbb{Z}^+$ , we have  $\alpha(k) \leq \exp(-c_\alpha k^\beta)$ .

We adopt the normalisation given in [Assumption 2\(i\)](#) for the purpose of identification; in general, factors and loadings are recoverable up to a linear invertible transformation only. Similar assumptions are found in the factor model literature, see e.g., equation (2.1) of [Fan et al. \(2013\)](#). In order to motivate [Assumptions 2\(i\), 3 and 4 \(i\)](#), we introduce the notations

$$\begin{aligned} \mathbf{I}_f^b &= \frac{1}{\eta_{b+1}^\chi - \eta_b^\chi} \sum_{t=\eta_b^\chi+1}^{\eta_{b+1}^\chi} E(\mathbf{f}_t \mathbf{f}_t^\top), \quad \mathbf{I}_\epsilon^b = \frac{1}{\eta_{b+1}^\epsilon - \eta_b^\epsilon} \sum_{t=\eta_b^\epsilon+1}^{\eta_{b+1}^\epsilon} E(\epsilon_t \epsilon_t^\top), \quad \mathbf{I}_\chi^b = \mathbf{A} \mathbf{I}_f^b \mathbf{A}^\top, \\ \mathbf{I}_\chi &= \frac{1}{T} \sum_{b=0}^{B_\chi} (\eta_{b+1}^\chi - \eta_b^\chi) \mathbf{I}_\chi^b, \quad \mathbf{I}_\epsilon = \frac{1}{T} \sum_{b=0}^{B_\epsilon} (\eta_{b+1}^\epsilon - \eta_b^\epsilon) \mathbf{I}_\epsilon^b \quad \text{and} \quad \mathbf{I}_x = \mathbf{I}_\chi + \mathbf{I}_\epsilon. \end{aligned}$$

We denote the eigenvalues (in non-increasing order) of  $\mathbf{I}_\chi^b, \mathbf{I}_\epsilon^b, \mathbf{I}_\chi, \mathbf{I}_\epsilon$  and  $\mathbf{I}_x$  by  $\mu_{\chi,j}^b, \mu_{\epsilon,j}^b, \mu_{\chi,j}, \mu_{\epsilon,j}$  and  $\mu_{x,j}$ , respectively. Then, [Assumptions 2\(i\) and 3](#) imply that  $\mu_{\chi,j}^b, j = 1, \dots, r_b$  are diverging as  $n \rightarrow \infty$  and, in particular, are of order  $n$  for all  $b = 0, \dots, B_\chi$ . [Assumption 3](#) implicitly rules out change-points which are due to weak changes in the loadings, in the sense that the magnitudes of the changes in the loadings are small, or only a small fraction of  $\chi_{it}$  undergoes the change. A similar requirement can be found in e.g., Assumption 1 of [Chen et al. \(2014\)](#) and Assumption 10 of [Han and Inoue \(2014\)](#). We note that [Bai et al. \(2017\)](#) studies the consistency of a least squares estimator for a single change-point attributed to a possibly weak break in the loadings in the above sense. In Section 7, we provide numerical results on the effect of the size of the break on our proposed methodology.

[Assumption 4\(i\)](#) guarantees that, when  $\mathbf{a} = (a_1, \dots, a_n)^\top$  is a normalised eigenvector of  $\mathbf{I}_\epsilon^b$ , then the largest eigenvalue of  $\mathbf{I}_\epsilon^b$  is bounded for all  $b = 0, \dots, B_\epsilon$  and  $n$ , that is  $\mu_{\epsilon,1}^b < C_\epsilon$ . This is the same assumption as those made in [Chamberlain and Rothschild \(1983\)](#) and [Forni et al. \(2009\)](#) and comparable to Assumption C.4 of [Bai \(2003\)](#) and Assumption 2.1 of [Fan et al. \(2011a\)](#) in the stationary case. Note that [Assumption 4\(i\)](#) is sufficient in guaranteeing the commonness of  $\chi_{it}$  and the idiosyncrasy of  $\epsilon_{it}$  according to Definitions 2.1 and 2.2 of [Hallin and Lippi \(2013\)](#). [Assumption 4 \(ii\)](#) may be relaxed to allow for  $\epsilon_{it}$  of exponential-type tails, provided that the tail behaviour carries over to the cross-sectional sums of  $\epsilon_{it}$ . Note that the normality of the idiosyncratic component does not necessarily imply the normality of the data since the factors are allowed to be non-normal.

[Assumption 5](#) is commonly found in the factor model literature (see e.g., Assumptions 3.1–3.2 of [Fan et al. 2011a](#)). In particular, the exponential-type tail conditions in [Assumptions 2 \(ii\) and 4 \(ii\)](#), along with the mixing condition in [Assumption 5 \(ii\)](#), allow us to control the deviation of sample covariance estimates from their population counterparts, via Bernstein-type inequality (see e.g., Theorem 1.4 of [Bosq \(1998\)](#) and Theorem 1 of [Merlevède et al. \(2011\)](#)).

We also assume the following on the minimum distance between two adjacent change-points.

**Assumption 6.** There exists a fixed  $c_\eta \in (0, 1)$  such that

$$\min \left\{ \min_{0 \leq b \leq B_\chi} |\eta_{b+1}^\chi - \eta_b^\chi|, \min_{0 \leq b \leq B_\epsilon} |\eta_{b+1}^\epsilon - \eta_b^\epsilon| \right\} \geq c_\eta T.$$

Under [Assumptions 2–6](#), the eigenvalues of  $\mathbf{T}_\chi$  and  $\mathbf{T}_\epsilon$  satisfy:

(C1) there exist some fixed  $\underline{c}_j, \bar{c}_j$  such that for  $j = 1, \dots, r$ ,

$$0 < \underline{c}_j < \liminf_{n \rightarrow \infty} \frac{\mu_{\chi,j}}{n} \leq \limsup_{n \rightarrow \infty} \frac{\mu_{\chi,j}}{n} < \bar{c}_j < \infty,$$

and  $\bar{c}_{j+1} < \underline{c}_j$  for  $j = 1, \dots, r-1$ ;

(C2)  $\mu_{\epsilon,1} < C_\epsilon$ , for any  $n$ .

Moreover, by Weyl's inequality, we have the following asymptotic behaviour of the eigenvalues of  $\mathbf{T}_\chi$ :

(C3) the  $r$  largest eigenvalues,  $\mu_{\chi,1}, \dots, \mu_{\chi,r}$ , are diverging linearly in  $n$  as  $n \rightarrow \infty$ ;

(C4) the  $(r+1)$ th largest eigenvalue,  $\mu_{\chi,r+1}$ , stays bounded for any  $n$ .

From (C1)–(C4) above, it is clear that for identification of the common and idiosyncratic components, factor models need to be studied in the asymptotic limit where  $n \rightarrow \infty$  and  $T \rightarrow \infty$ . This requirement, in practice, recommends the use of large cross-sections to apply PCA for factor analysis. In particular, we require

**Assumption 7.**  $n \rightarrow \infty$  as  $T \rightarrow \infty$ , with  $n = O(T^\kappa)$  for some  $\kappa \in (0, \infty)$ .

Under [Assumption 7](#), we are able to establish the consistency in estimation of the common components as well as that of the change-points in high-dimensional settings where  $n > T$ , even when the factor number  $r$  is unknown.

**Remark 2.** The multiple change-point detection algorithm adopted in Stage 2 of our methodology still achieves consistency when [Assumption 6](#) is relaxed to allow  $\min\{|\eta_{b+1}^\chi - \eta_b^\chi|, |\eta_{b+1}^\epsilon - \eta_b^\epsilon|\} \geq c_\eta T^\nu$  for some fixed  $c > 0$  and  $\nu \in (6/7, 1]$ , with  $B \equiv B_T$  increasing in  $T$  such that  $B_T = O(\log^2 T)$ . However, under these relaxed conditions, it is no longer guaranteed that  $\mathbf{T}_\chi$  and, consequently,  $\mathbf{T}_\epsilon$  have  $r$  diverging eigenvalues. Therefore, in this paper we limit the scope of our theoretical results to the more restricted setting of [Assumption 6](#).

### 3. Factor analysis and wavelet transformation

#### 3.1. Estimation of the common and idiosyncratic components

Decomposing  $x_{it}$  into common and idiosyncratic components is an essential step for the separate treatment of the change-points in  $\chi_{it}$  and  $\epsilon_{it}$ , such that we can identify the origins of detected change-points. Therefore, in this section, we establish the asymptotic bounds on the estimation error of the common and idiosyncratic components, when using PCA under the piecewise stationary factor model in [\(1\)](#).

Let  $\hat{\mathbf{w}}_{\chi,j}$  denote the  $n$ -dimensional normalised eigenvector corresponding to the  $j$ th largest eigenvalue of the sample covariance matrix,  $\hat{\mathbf{T}}_\chi$ , with its entries  $\hat{w}_{\chi,ij}$ ,  $i = 1, \dots, n$ . When the number of factors  $r$  is known, the PCA estimator of  $\chi_{it}$  is defined as  $\hat{\chi}_{it} = \sum_{j=1}^r \hat{w}_{\chi,ij} \hat{\mathbf{w}}_{\chi,j}^\top \mathbf{x}_t$ , for which the following theorem holds.

**Theorem 1.** Under [Assumptions 1–7](#), the PCA estimator of  $\chi_{it}$  with  $r$  known satisfies

$$\max_{1 \leq i \leq n} \max_{1 \leq t \leq T} |\hat{\chi}_{it} - \chi_{it}| = O_p \left\{ \left( \sqrt{\frac{\log n}{T}} \vee \frac{1}{\sqrt{n}} \right) \log^\theta T \right\},$$

for some fixed  $\theta \geq 1 + (\beta_f^{-1} \vee 1/2)$ .

Proofs of [Theorem 1](#) and all other theoretical results are provided in [Appendix B](#).

Despite the presence of multiple change-points, allowed both in the variance and autocorrelations of  $\mathbf{f}_t$ , the rate of convergence for  $\hat{\chi}_{it}$  is almost as fast as the one derived for the stationary case, e.g., [Theorem 3](#) of [Bai \(2003\)](#), and [Theorem 1](#) of [Pelger \(2015\)](#) and [Theorem 5](#) of [Aït-Sahalia and Xiu \(2017\)](#) in the context of factor modelling high-frequency data. We highlight that [Theorem 1](#) derives a uniform bound on the estimation error over  $i = 1, \dots, n$  and  $t = 1, \dots, T$ ; similar results are found in [Theorem 4](#) of [Fan et al. \(2013\)](#).

However, the true number of factors  $r$  is typically unknown and, since the seminal paper by [Bai and Ng \(2002\)](#), estimation of the number of factors has been one of the most researched problems in the factor model literature see also [Alessi et al. \(2010\)](#), [Onatski \(2010\)](#) and [Ahn and Horenstein \(2013\)](#). Although, ([Han and Inoue, 2014](#)) (in their Proposition 1) and ([Chen](#)

et al., 2014)(in their Proposition 2) showed that the information criteria of Bai and Ng (2002) achieve asymptotic consistency in estimating  $r$  in the presence of a single break in the loadings, it has been observed that such an approach tends to exhibit poor finite sample performance when the idiosyncratic components are both serially and cross-sectionally correlated, or when  $\text{Var}(\epsilon_{it})$  is large compared to  $\text{Var}(\chi_{it})$ . Also, it was noted that the factor number estimates heavily depend on the relative magnitude of  $n$  and  $T$  and the choice of penalty terms as demonstrated in the numerical studies of Bai and Ng (2002) and Ahn and Horenstein (2013). While the majority of change-point methods for factor models rely on the consistent estimation of the number of factors (Breitung and Eickmeier, 2011; Han and Inoue, 2014; Chen et al., 2014; Corradi and Swanson, 2014), our empirical study on simulated data, available in the supplementary document, indicates that this dependence on a single estimate may lead to failure in change-point detection.

To remedy this, we propose to consider a range of factor number candidates in our change-point analysis, in particular, allowing for the over-specification of the number of factors. In high dimensions, the estimated eigenvectors  $\hat{\mathbf{w}}_{x,j}$  for  $j > r$  are in general not consistent, as implied by Theorem 2 of Yu et al. (2015). Thus, estimates of common components based on more than  $r$  principal components may be subject to a non-negligible overestimation error. In order to control the effect of over-specifying the number of factors, we propose a modified PCA estimator of  $\chi_{it}$  defined as  $\hat{\chi}_{it}^k = \sum_{j=1}^k \tilde{w}_{x,ij} \tilde{\mathbf{w}}_{x,j}^\top \mathbf{x}_t$ , where each element of  $\tilde{\mathbf{w}}_{x,j}$  is ‘capped’ according to the rule

$$\tilde{w}_{x,ij} = \hat{w}_{x,ij} \mathbb{I}\left(|\hat{w}_{x,ij}| \leq \frac{c_w}{\sqrt{n}}\right) + \text{sign}(\hat{w}_{x,ij}) \cdot \frac{c_w}{\sqrt{n}} \mathbb{I}\left(|\hat{w}_{x,ij}| > \frac{c_w}{\sqrt{n}}\right),$$

for some fixed  $c_w > 0$ . The estimated loadings are then given by  $\hat{\lambda}_{ij} = \sqrt{n} \tilde{w}_{x,ij}$  and the factors by  $\hat{f}_{jt} = n^{-1/2} \tilde{\mathbf{w}}_{x,j}^\top \mathbf{x}_t$ . A practical way of choosing  $c_w$  is discussed in Section 6.2.

Note that, thanks to the result in (C3) and Lemmas 2 and 3 in Appendix B, we have  $\max_{1 \leq j \leq r} \max_{1 \leq i \leq n} |\hat{w}_{x,ij}| = O_p(1/\sqrt{n})$ .

In other words, asymptotically, the capping does not alter the contribution from the  $r$  leading eigenvectors of  $\hat{\mathbf{T}}_x$  to  $\hat{\chi}_{it}^k$ , even when the capping is applied without the knowledge of  $r$ . On the other hand, by means of this capping, we control the contribution from the ‘spurious’ eigenvectors when  $k > r$ , which allows us to establish the following bound on the partial sums of the estimation errors even when the factor number is over-specified.

**Theorem 2.** Suppose that Assumptions 1–7 hold, and let  $\theta$  be defined as in Theorem 1. Then, the capped estimator  $\hat{\chi}_{it}^k$  satisfies the following:

(i) when  $k = r$ ,

$$\max_{1 \leq i \leq n} \max_{1 \leq s < e \leq T} \frac{1}{\sqrt{e-s+1}} \left| \sum_{t=s}^e (\hat{\chi}_{it}^r - \chi_{it}) \right| = O_p \left\{ \left( \sqrt{\frac{\log n}{T}} \vee \frac{1}{\sqrt{n}} \right) \log^\theta T \right\}; \quad (2)$$

(ii) when  $k > r$ ,

$$\max_{1 \leq i \leq n} \max_{1 \leq s < e \leq T} \frac{1}{\sqrt{e-s+1}} \left| \sum_{t=s}^e (\hat{\chi}_{it}^k - \chi_{it}) \right| = O_p(\log^\theta T); \quad (3)$$

with  $\theta$  given in Theorem 1.

The bound in (2) concerns the case in which we correctly specify the number of factors and is in agreement with Theorem 1. Turning to when the factor number is over-specified, Forni et al. (2000) (in their Corollary 2) and (Onatski, 2015) (in his Proposition 1) reported similar results for the stationary case. However, the uniform consistency of  $\hat{\chi}_{it}^k$ ,  $k \geq r$ , in the presence of multiple change-points, has not been spelled out before in the factor model literature to the best of our knowledge; we achieve this via the proposed capping. On the other hand, it is possible to show that with  $k < r$ , the estimation error in  $\hat{\chi}_{it}^k$  is non-negligible. Lastly, thanks to Lemma 4 in Appendix B, it is straightforward to show that analogous bounds hold for the idiosyncratic component  $\hat{\epsilon}_{it}^k = x_{it} - \hat{\chi}_{it}^k$ .

Although the over-specification of  $k > r$  brings the bound in (2) to increase by  $\sqrt{n} \wedge \sqrt{T}$ , we can still guarantee that all change-points in  $\chi_{it}$  ( $\epsilon_{it}$ ) are detectable from  $\hat{\chi}_{it}^k$  ( $\hat{\epsilon}_{it}^k$ ) provided that  $k \geq r$ , as shown in Proposition 1 and Theorem 3. In what follows, we continue describing our methodology by supposing that  $k \geq r$  is given, and we refer to Section 4.3 for the complete description of the proposed ‘screening’ procedure that considers a range of values for  $k$ .

### 3.2. Wavelet periodograms and cross-periodograms

As the first stage of the proposed methodology, we construct a wavelet-based transformation (WT) of the estimated common and idiosyncratic components  $\hat{\chi}_{it}^k$  and  $\hat{\epsilon}_{it}^k$  from Section 3.1, which serves as an input to the algorithm for high-dimensional change-point analysis in Stage 2. In order to motivate the WT, which will be formally introduced in Section 3.3, we limit our discussion in this section to the (unobservable) common component  $\chi_{it}$ ; the same arguments hold verbatim for the idiosyncratic one. In practice, the WT is performed on the estimated common component,  $\hat{\chi}_{it}^k$ , and the effect of considering estimated quantities rather than the true ones is studied in Section 3.3.



Nason et al. (2000) have proposed the use of wavelets as building blocks in nonstationary time series analogous to Fourier exponentials in the classical Cramér representation for stationary processes. The simplest example of a wavelet system, Haar wavelets, are defined as

$$\psi_{j,l} = 2^{j/2} \mathbb{I}(0 \leq l \leq 2^{-j-1} - 1) - 2^{j/2} \mathbb{I}(2^{-j-1} \leq l \leq 2^{-j} - 1),$$

with  $j \in \{-1, -2, \dots\}$  denoting the wavelet scale, and  $l \in \mathbb{Z}$  denoting the location. Small negative values of the scale parameter  $j$  denote fine scales where the wavelet vectors are more localised and oscillatory, while large negative values denote coarser scales with longer, less oscillatory wavelet vectors.

Recall the notation  $\beta(t) = \max\{0 \leq b \leq B_\chi : \eta_b^\chi + 1 \leq t\}$ . Wavelet coefficients of  $\chi_{it}^{\beta(t)}$  (introduced in Assumption 1) are defined as  $d_{j,it} = \sum_{l=0}^{\mathcal{L}_j-1} \chi_{i,t-l}^{\beta(t)} \psi_{j,l}$ , with respect to  $\psi_j = (\psi_{j,0}, \dots, \psi_{j,\mathcal{L}_j-1})^\top$ , a vector of discrete wavelets at scale  $j$ . Note that the support of  $\psi_j$  is of length  $\mathcal{L}_j = M2^{-j}$  for a fixed  $M > 0$  (which depends on the choice of wavelet family), so that we have access to wavelet coefficients from at most  $\lfloor \log_2 T \rfloor$  scales for a time series of length  $T$ . In other words, wavelet coefficients are obtained by filtering  $\chi_{it}^{\beta(t)}$  with respect to wavelet vectors of finite lengths.

Wavelet periodogram and cross-periodogram sequences of  $\chi_{it}^{\beta(t)}$  are defined as  $I_{j,it} = |d_{j,it}|^2$  and  $I_{j,i't} = d_{i,it} d_{j,i't}$ . It has been shown that the expectations of these sequences have a one-to-one correspondence with the second-order structure of the input time series, see Cho and Fryzlewicz (2012) for the case of univariate time series and (Cho and Fryzlewicz, 2015) for the high-dimensional case. To illustrate, suppose that  $t - \mathcal{L}_j + 1 \geq \underline{\eta}^\chi(t) + 1$ . Then,

$$\begin{aligned} \mathbb{E}|d_{j,it}|^2 &= \sum_{l,l'=0}^{\mathcal{L}_j-1} \mathbb{E}(\chi_{i,t-l}^{\beta(t-l)} \chi_{i,t-l'}^{\beta(t-l')}) \psi_{j,l} \psi_{j,l'} = \sum_{l=0}^{\mathcal{L}_j-1} \sum_{|\tau| < \mathcal{L}_j} \mathbb{E}(\chi_{i,t-l}^{\beta(t-l)} \chi_{i,t-l-\tau}^{\beta(t-l-\tau)}) \psi_{j,l} \psi_{j,l+\tau} \\ &= \sum_{|\tau| < \mathcal{L}_j} [\tilde{\mathbf{T}}_\chi^{\beta(t)}(\tau)]_{l,i} \psi_j(\tau), \end{aligned} \quad (4)$$

where  $\psi_j(\tau) = \sum_{l \in \mathbb{Z}} \psi_{j,l} \psi_{j,l+\tau}$  (with  $\psi_{j,l} = 0$  unless  $0 \leq l \leq \mathcal{L}_j - 1$ ) denotes the autocorrelation wavelets (Nason et al., 2000) and  $\tilde{\mathbf{T}}_\chi^b(\tau) = \mathbb{E}[\chi_{t+\tau}^b (\chi_t^b)^\top]$ . That is, provided that  $t$  is sufficiently distanced (by  $\mathcal{L}_j$ ) from any change-points to its left,  $\mathbb{E}|d_{j,it}|^2$  is a wavelet transformation of the (auto)covariances  $\mathbb{E}(\chi_{it}^{\beta(t)} \chi_{i,t+\tau}^{\beta(t)})$ ,  $|\tau| \leq \mathcal{L}_j - 1$ . Similar arguments hold between wavelet cross-periodograms and cross-covariances of  $\chi_{it}^{\beta(t)}$ . Following Cho and Fryzlewicz (2015), we conclude that under Assumption 1 (ii), any jump in the autocovariance and cross-covariance structures of  $\chi_t^{\beta(t)}$  is translated to a jump in the means of its wavelet periodogram and cross-periodogram sequences at some wavelet scale  $j$ , in the following sense:  $\{\mathbb{E}|d_{j,it}|^2, 1 \leq i \leq n, \mathbb{E}(d_{j,it} d_{j,i't}), 1 \leq i < i' \leq n\}$  are ‘almost’ piecewise constant with their change-points coinciding with those in  $\mathcal{B}^\chi$ , apart from intervals of length  $\mathcal{L}_j$  around the change-points.

It is reasonable to limit our consideration to wavelets at the first  $J_T^*$  finest scales  $j = -1, \dots, -J_T^*$  (with  $J_T^* \leq \lfloor \log_2 T \rfloor$ ), in order to control the possible bias in change-point estimation that arises from the transition intervals of length  $\mathcal{L}_j$ . On the other hand, due to the compactness of the support of  $\psi_j$ , a change in  $\tilde{\mathbf{T}}_\chi^{\beta(t)}(\tau)$  that appears only at some large lags ( $|\tau| \geq \mathcal{L}_{-J_T^*} = M2^{J_T^*}$ ), is not detectable as a change-point in the wavelet periodogram and cross-periodogram sequences at the few finest scales ( $j \geq -J_T^*$ ), see (4). To strike a balance between the above quantities related to the choice of  $J_T^*$ , we set  $J_T^* = \lfloor C \log_2 \log^\nu T \rfloor$  for some  $\nu \in (0, 1]$  and some fixed  $C > 0$ . We refer to Section 6.2 for the choice of  $C$ . Then,

- (a) any change in  $\mathbf{T}_\chi^{\beta(t)}(\tau)$  that occurs at (at least) one out of an increasing number of lags ( $|\tau| \leq M \log^\nu T$ ), is registered as a change-point in the expectations of the wavelet (cross-)periodograms of  $\chi_{it}^{\beta(t)}$ ;
- (b) the possible bias in the registered locations of the change-points is controlled to be at most  $O(\log^\nu T)$ .

### 3.3. Wavelet-based transformation for change-point analysis

In this section, we propose the WT of estimated common and idiosyncratic components, and show that the change-points in the complex (autocovariance and cross-covariance) structure of (unobservable)  $\chi_{it}$  and  $\epsilon_{it}$ , are made detectable as the change-points in the relatively simple structure (means) of the panel of the wavelet transformed  $\hat{\chi}_{it}^k$  and  $\hat{\epsilon}_{it}^k$ . This panel then serves as an input to Stage 2 of our methodology, as described in Section 4. As in Section 3.2, we limit the discussion of the WT and its properties when applied to the change-point analysis of the common components; the same arguments are applicable to that of the idiosyncratic components.

Let  $\hat{d}_{j,it} = \sum_{l=0}^{\mathcal{L}_j-1} \hat{\chi}_{i,t-l}^k \psi_{j,l}$  denote the wavelet coefficients of  $\hat{\chi}_{it}^k$ . Then, for each  $j = -1, -2, \dots, -J_T^*$ , we propose the following transformation which takes  $\hat{\chi}_{it}^k$  and produces a panel of  $n(n+1)/2$ -dimensional sequences with elements:

$$g_j(\hat{\chi}_{it}^k) \equiv g_j(\hat{\chi}_{it}^k, \dots, \hat{\chi}_{i,t-\mathcal{L}_j+1}^k) = |\hat{d}_{j,it}| = \left| \sum_{l=0}^{\mathcal{L}_j-1} \hat{\chi}_{i,t-l}^k \psi_{j,l} \right|, \quad 1 \leq i \leq n,$$

$$h_j(\widehat{\chi}_{it}^k, \widehat{\chi}_{i't}^k) \equiv h_j(\widehat{\chi}_{it}^k, \dots, \widehat{\chi}_{i,t-\mathcal{L}_j+1}^k, \widehat{\chi}_{i't}^k, \dots, \widehat{\chi}_{i',t-\mathcal{L}_j+1}^k) = |\widehat{d}_{j,it} + s_{iit'} \widehat{d}_{j,i't}|$$

$$= \left| \sum_{l=0}^{\mathcal{L}_j-1} \widehat{\chi}_{i,t-l}^k \psi_{j,l} + s_{iit'} \sum_{l'=0}^{\mathcal{L}_j-1} \widehat{\chi}_{i',t-l'}^k \psi_{j,l'} \right|, \quad 1 \leq i < i' \leq n,$$

where  $s_{iit'} \in \{-1, 1\}$ . For example, with Haar wavelets at scale  $j = -1$ ,

$$g_{-1}(\widehat{\chi}_{it}^k) = \frac{1}{\sqrt{2}} |\widehat{\chi}_{it}^k - \widehat{\chi}_{i,t-1}^k| \quad \text{and} \quad h_{-1}(\widehat{\chi}_{it}^k, \widehat{\chi}_{i't}^k) = \frac{1}{\sqrt{2}} |\widehat{\chi}_{it}^k - \widehat{\chi}_{i,t-1}^k + s_{iit'}(\widehat{\chi}_{i't}^k - \widehat{\chi}_{i',t-1}^k)|.$$

**Remark 3.** Notice that  $g_j(\widehat{\chi}_{it}^k)$  is simply the squared root of the wavelet periodogram of  $\widehat{\chi}_{it}^k$  at scale  $j$ . Arguments supporting the choice of  $h_j$  in place of wavelet cross-periodogram sequences can be found in Section 3.1.2 of [Cho and Fryzlewicz \(2015\)](#), where it is guaranteed that any change detectable from  $\widehat{d}_{j,it} \cdot \widehat{d}_{j,i't}$  can also be detected from  $h_j(\widehat{\chi}_{it}^k, \widehat{\chi}_{i't}^k)$  with either of  $s_{iit'} \in \{1, -1\}$ . As per the recommendation made therein, we select  $s_{iit'} = -\text{sign}\{\widehat{\text{cor}}(\widehat{\chi}_{it}^k, \widehat{\chi}_{i't}^k)\}$ , where  $\widehat{\text{cor}}(\widehat{\chi}_{it}^k, \widehat{\chi}_{i't}^k)$  denotes the sample correlation between  $\widehat{\chi}_{it}^k$  and  $\widehat{\chi}_{i't}^k$  over  $t = 1, \dots, T$ . This is in an empirical effort to select  $s_{iit'}$  that better brings out any change in  $E[h_j(\chi_{it}, \chi_{i't})]$  for the given data.

As with wavelet periodograms and cross-periodograms discussed in Section 3.2, the transformed series  $g_j(\widehat{\chi}_{it}^k)$  and  $h_j(\widehat{\chi}_{it}^k, \widehat{\chi}_{i't}^k)$  contain the change-points in the second-order structure of  $\chi_{it}^{\beta(t)}$  as change-points in their 'signals'. This is formalised in the following proposition.

**Proposition 1.** Suppose that all the conditions in [Theorem 2](#) hold. For some fixed  $k \geq r$  and  $J_T^* = \lfloor \text{Clog}_2 \log^v T \rfloor$ , consider the  $N = J_T^* n(n+1)/2$ -dimensional panel

$$\{g_j(\widehat{\chi}_{it}^k), 1 \leq i \leq n, h_j(\widehat{\chi}_{it}^k, \widehat{\chi}_{i't}^k), 1 \leq i < i' \leq n; -J_T^* \leq j \leq -1; 1 \leq t \leq T\} \quad (5)$$

and denote as  $y_{\ell t}$  a generic element of (5). Then, we have the following decomposition:

$$y_{\ell t} = z_{\ell t} + \varepsilon_{\ell t}, \quad \ell = 1, \dots, N, \quad t = 1, \dots, T. \quad (6)$$

(i)  $z_{\ell t}$  are piecewise constant as the corresponding elements of

$$\left\{ E\{g_j(\chi_{it}^{\beta(t)})\}, 1 \leq i \leq n, E\{h_j(\chi_{it}^{\beta(t)}, \chi_{i't}^{\beta(t)})\}, 1 \leq i < i' \leq n; -J_T^* \leq j \leq -1; 1 \leq t \leq T \right\}. \quad (7)$$

That is, all change-points in  $z_{\ell t}$  belong to  $\mathcal{B}^X = \{\eta_1^X, \dots, \eta_{B_X}^X\}$  and for each  $b \in \{1, \dots, B_X\}$ , there exists at least a single index  $\ell \in \{1, \dots, N\}$  for which  $|z_{\ell, \eta_b^X+1} - z_{\ell, \eta_b^X}| \neq 0$ .

(ii)  $\max_{1 \leq \ell \leq N} \max_{1 \leq s < e \leq T} (e - s + 1)^{-1/2} |\sum_{t=s}^e \varepsilon_{\ell t}| = O_p(\log^{\theta+v} T)$ .

Therefore, the panel data in (5) contains all change-points in the second-order structure of  $\chi_{it}^{\beta(t)}$  as the change-points in its piecewise constant signals represented by  $z_{\ell t}$ .

Let us denote by  $\widetilde{y}_{\ell t}$  an element of the panel obtained by transforming  $\chi_{it}$  in place of  $\widehat{\chi}_{it}^k$  in (5). Then, the proof of [Proposition 1](#) is based on the following decomposition

$$\varepsilon_{\ell t} = y_{\ell t} - z_{\ell t} = \{E(\widetilde{y}_{\ell t}) - z_{\ell t}\} + \{\widetilde{y}_{\ell t} - E(\widetilde{y}_{\ell t})\} + (y_{\ell t} - \widetilde{y}_{\ell t}) = I + II + III.$$

Then,  $I$  accounts for the discrepancy between  $\chi_{it}$  and  $\chi_{it}^{\beta(t)}$  and is controlled by [Assumption 1\(i\)](#), and  $II$  follows from the weak dependence structure and the tail behaviour of  $\mathbf{f}_t$  assumed in [Assumptions 2 \(ii\)](#) and [5](#). Term  $III$  arises from the estimation error in  $\widehat{\chi}_{it}^k$  which can be bounded as shown in [Theorem 2](#), which further motivates the WT of  $\widehat{\chi}_{it}^k$  via  $g_j$  and  $h_j$  rather than using its wavelet (cross-)periodograms.

To conclude, note that the WT of  $\widehat{\chi}_{it}^k$  are collected into the  $N$ -dimensional panel

$$\{g_j(\widehat{\epsilon}_{it}^k), 1 \leq i \leq n, h_j(\widehat{\epsilon}_{it}^k, \widehat{\epsilon}_{i't}^k), 1 \leq i < i' \leq n; -J_T^* \leq j \leq -1; 1 \leq t \leq T\} \quad (8)$$

for change-point analysis, where the elements of (8) are also decomposed into piecewise constant signals

$$\left\{ E\{g_j(\epsilon_{it}^{\gamma(t)})\}, 1 \leq i \leq n, E\{h_j(\epsilon_{it}^{\gamma(t)}, \epsilon_{i't}^{\gamma(t)})\}, 1 \leq i < i' \leq n; -J_T^* \leq j \leq -1; 1 \leq t \leq T \right\}, \quad (9)$$

and error terms of bounded partial sums; see [Appendix C](#) where we present the result analogous to [Proposition 1](#) for the idiosyncratic components.

## 4. High-dimensional panel data segmentation

### 4.1. Double CUSUM binary segmentation

Cumulative sum (CUSUM) statistics have been widely adopted for change-point detection in both univariate and multivariate data. In order to detect change-points in the  $N$ -dimensional additive panel data in (5), we compute  $N$  univariate

CUSUM series and aggregate the high-dimensional CUSUM series via the Double CUSUM statistic proposed by [Cho \(2016\)](#), which achieves this through point-wise, data-driven partitioning of the panel data. When dealing with multiple change-point detection, the Double CUSUM statistic is used jointly with binary segmentation in an algorithm, which we refer to as the Double CUSUM Binary Segmentation (DCBS) algorithm. The DCBS algorithm guarantees the consistency in multiple change-point detection in high-dimensional settings, as shown in [Theorem 3](#), while allowing for the cross-sectional size of change to decrease with increasing sample size (see [Assumption 9](#)). Further, it admits both serial- and cross-correlations in the data, which is a case highly relevant for the time series factor model considered in this paper.

Consider the  $N$ -dimensional input panel  $\{y_{\ell t}, \ell = 1, \dots, N; t = 1, \dots, T\}$ , computed from  $\hat{\chi}_{it}^k$  or  $\hat{\epsilon}_{it}^k$  for the change-point analysis in either the common or the idiosyncratic components via WT; see [\(5\)](#) and [\(8\)](#) for their definitions. We define the CUSUM series of  $y_{\ell t}$  over a generic segment  $[s, e]$  for some  $1 \leq s < e \leq T$ , as

$$\mathcal{Y}_{s,b,e}^{\ell} = \frac{1}{\sigma_{\ell}} \sqrt{\frac{(b-s+1)(e-b)}{e-s+1}} \left( \frac{1}{b-s+1} \sum_{t=s}^b y_{\ell t} - \frac{1}{e-b} \sum_{t=b+1}^e y_{\ell t} \right),$$

for  $b = s, \dots, e-1$ , where  $\sigma_{\ell}$  denotes a scaling constant for treating all rows of the panel data  $y_{\ell t}$ ,  $1 \leq \ell \leq N$  on equal footing; see [Remark 4](#) for its choice. We note that if  $\varepsilon_{\ell t}$  in [\(6\)](#) were i.i.d. Gaussian random variables, the maximum likelihood estimator of the change-point location for  $\{y_{\ell t}, s \leq t \leq e\}$  would coincide with  $\arg \max_{b \in [s,e]} |\mathcal{Y}_{s,b,e}^{\ell}|$ .

Proposed in [Cho \(2016\)](#), the Double CUSUM (DC) operator aggregates the  $N$  series of CUSUM statistics from  $y_{\ell t}$  and returns a two-dimensional array of DC statistics:

$$\mathcal{D}_{s,b,e}(m) = \sqrt{\frac{m(2N-m)}{2N}} \left( \frac{1}{m} \sum_{\ell=1}^m |\mathcal{Y}_{s,b,e}^{(\ell)}| - \frac{1}{2N-m} \sum_{\ell=m+1}^N |\mathcal{Y}_{s,b,e}^{(\ell)}| \right),$$

for  $b = s, \dots, e-1$  and  $m = 1, \dots, N$ , where  $|\mathcal{Y}_{s,b,e}^{(\ell)}|$  denotes the CUSUM statistics at  $b$  ordered according to their moduli, i.e.,  $|\mathcal{Y}_{s,b,e}^{(1)}| \geq \dots \geq |\mathcal{Y}_{s,b,e}^{(N)}|$ . Notice that  $\mathcal{D}_{s,b,e}(m)$  takes the contrast between the  $m$  largest CUSUM values  $|\mathcal{Y}_{s,b,e}^{(\ell)}|$ ,  $1 \leq \ell \leq m$  and the rest at each  $b$ , and thus partitions the coordinates into the  $m$  that are the most likely to contain a change-point and those which are not in a point-wise manner. Then, the test statistic is derived by maximising the two-dimensional array of DC statistics over both time and cross-sectional indices, as

$$\mathcal{T}_{s,e} = \max_{b \in [s,e]} \max_{1 \leq m \leq N} \mathcal{D}_{s,b,e}(m), \quad (10)$$

which is compared against a threshold  $\pi_{N,T}$  for determining the presence of a change-point over the interval  $[s, e]$ . If  $\mathcal{T}_{s,e} > \pi_{N,T}$ , the location of the change-point is identified as

$$\hat{\eta} = \arg \max_{b \in [s,e]} \max_{1 \leq m \leq N} \mathcal{D}_{s,b,e}(m).$$

**Remark 4** (Choice of  $\sigma_{\ell}$ ). [Cho \(2016\)](#) assumes second-order stationarity on the error term  $\varepsilon_{\ell t}$  in [\(6\)](#), which enables the use of its long-run variance estimator as the scaling term  $\sigma_{\ell}$ . However, it is not trivial to define a similar quantity in the problem considered here, particularly due to the possible nonstationarities in  $\varepsilon_{\ell t}$ . Following [Cho and Fryzlewicz \(2015\)](#), in order for the CUSUM series computed on  $y_{\ell t}$  not to depend on the level of  $E(y_{\ell t}^2)$ , we also adopt the choice of  $\sigma_{\ell} = \sqrt{T^{-1} \sum_{t=1}^T y_{\ell t}^2}$ . Note that the asymptotic consistency of the DCBS algorithm in [Theorem 3](#) does not depend on the choice of  $\sigma_{\ell}$ , provided that it is bounded away from zero and from the above for all  $\ell = 1, \dots, N$  with probability tending to one (see [Assumption \(A6\)](#) of [Cho \(2016\)](#)). By adopting arguments similar to those in the proof of Lemma 6 in [Cho and Fryzlewicz \(2012\)](#), it can be shown that our choice of  $\sigma_{\ell}$  satisfies these properties.

We now formulate the DCBS algorithm which is equipped with the threshold  $\pi_{N,T}$ . The index  $u$  is used to denote the level (indicating the progression of the segmentation procedure) and  $v$  to denote the location of the node at each level.

The Double CUSUM Binary Segmentation (DCBS) algorithm

Step 0: Set  $(u, v) = (1, 1)$ ,  $s_{u,v} = 1$ ,  $e_{u,v} = T$  and  $\hat{\mathcal{B}} = \emptyset$ .

Step 1: At the current level  $u$ , repeat the following for all  $v$ .

Step 1.1: Letting  $s = s_{u,v}$  and  $e = e_{u,v}$ , obtain the series of CUSUMs  $\mathcal{Y}_{s,b,e}^{\ell}$  for  $b \in [s, e]$  and  $\ell = 1, \dots, N$ , on which  $\mathcal{D}_{s,b,e}(m)$  is computed over all  $b$  and  $m$ .

Step 1.2: Obtain the test statistic  $\mathcal{T}_{s,e} = \max_{b \in [s,e]} \max_{1 \leq m \leq N} \mathcal{D}_{s,b,e}(m)$ .

Step 1.3: If  $\mathcal{T}_{s,e} \leq \pi_{N,T}$ , quit searching for change-points on the interval  $[s, e]$ . On the other hand, if  $\mathcal{T}_{s,e} > \pi_{N,T}$ , locate  $\hat{\eta} = \arg \max_{b \in [s,e]} \max_{1 \leq m \leq N} \mathcal{D}_{s,b,e}(m)$ , add it to the set of estimated change-points  $\hat{\mathcal{B}}$ , and proceed to Step 1.4.

Step 1.4: Divide the interval  $[s_{u,v}, e_{u,v}]$  into two sub-intervals  $[s_{u+1,2v-1}, e_{u+1,2v-1}]$  and  $[s_{u+1,2v}, e_{u+1,2v}]$ , where  $s_{u+1,2v-1} = s_{u,v}$ ,  $e_{u+1,2v-1} = \hat{\eta}$ ,  $s_{u+1,2v} = \hat{\eta} + 1$  and  $e_{u+1,2v} = e_{u,v}$ .

Step 2: Once  $[s_{u,v}, e_{u,v}]$  for all  $v$  are examined at level  $u$ , set  $u \leftarrow u + 1$  and go to Step 1.

Step 1.3 provides a stopping rule to the DCBS algorithm, by which the search for further change-points is terminated once  $\mathcal{T}_{s,e} \leq \pi_{N,T}$  on every segment defined by two adjacent estimated change-points in  $\hat{\mathcal{B}}$ . Depending on the choice of the input panel data  $y_{\ell t}$  as in [\(5\)](#) or [\(8\)](#), the DCBS algorithm returns  $\hat{\mathcal{B}}^{\chi}(k)$  or  $\hat{\mathcal{B}}^{\epsilon}(k)$ , the sets of change-points detected from  $\hat{\chi}_{it}^k$  and  $\hat{\epsilon}_{it}^k$ , respectively.

#### 4.2. Consistency in multiple change-point detection

In this section, we show the consistency of change-points estimated for the common and idiosyncratic components by the DCBS algorithm, in terms of their total number and locations. Consider the panel  $\{y_{\ell t}, \ell = 1, \dots, N; t = 1, \dots, T\}$  that represents the WTs of  $\hat{\chi}_{it}^k$  and  $\hat{\epsilon}_{it}^k$  as in (5) and (8), respectively. In either case, let  $\eta_b, b = 1, \dots, B$  denote the change-points in the piecewise constant signals  $z_{\ell t}$  underlying  $y_{\ell t}$  (see (7) and (9) for the precise definitions of  $z_{\ell t}$ ), so that we have either  $\eta_b = \eta_b^\chi$  with  $B = B_\chi$ , or  $\eta_b = \eta_b^\epsilon$  with  $B = B_\epsilon$ . We impose the following assumptions on the signals  $z_{\ell t}$  and the change-point structure therein.

**Assumption 8.** There exists a fixed constant  $\bar{g} > 0$  such that  $\max_{1 \leq \ell \leq N} \max_{1 \leq t \leq T} |z_{\ell t}| \leq \bar{g}$ .

**Assumption 9.** At each  $\eta_b$ , define  $\Pi_b = \{1 \leq \ell \leq N : \delta_{\ell, b} = |z_{\ell, \eta_b+1} - z_{\ell, \eta_b}| > 0\}$ , the index set of those  $z_{\ell t}$  that undergo a break at  $t = \eta_b$ , and denote its cardinality by  $m_b = |\Pi_b|$ . Further, let  $\tilde{\delta}_b = m_b^{-1} \sum_{\ell \in \Pi_b} |\delta_{\ell, b}|$ , the average size of jumps in  $z_{\ell t}$  at  $t = \eta_b$ . Then  $\Delta_{N, T} = \min_{1 \leq b \leq B} \sqrt{m_b} \tilde{\delta}_b$  satisfies  $(\sqrt{N} \log^{\theta+u} T)^{-1} \Delta_{N, T} T^{1/4} \rightarrow \infty$  as  $T \rightarrow \infty$ .

Assumption 8 requires the expectations of the WTs of  $\chi_{it}^b$  and  $\epsilon_{it}^b$  defined in Assumption 1 to be bounded, which in fact holds trivially as  $\sum_{l=0}^{C_j-1} \psi_{j,l}^2 = 1$  for all  $j$ .

Assumption 9 imposes a condition on the minimal cross-sectional size of the changes in  $z_{\ell t}$ , represented by  $\Delta_{N, T}$ . Under Assumption 3, the change-points which are due to breaks in the loadings or (dis)appearance of new factors (see Remark 1), are implicitly required to be ‘dense’, in the sense that the number of coordinates in  $\mathbf{x}_t$  affected by the changes is of order  $n$ . Consequently, such changes appear in a large number (of order  $n^2$ ) of elements of  $\tilde{\mathbf{T}}_\chi^{\beta(t)}(\tau)$  and, therefore, that of the WT of the second-order structure,  $z_{\ell t}$ . Similarly, noting that  $[\tilde{\mathbf{T}}_\chi^{\beta(t)}]_{i,i'} = \sum_{j,j'=1}^r \lambda_{ij} \lambda_{i'j'} E(f_{jt}^{b,j} f_{j't}^{b,j'})$ , a break in the autocovariance functions of  $\mathbf{f}_t^{\beta(t)}$  results in a dense change-point that affects a large number of  $z_{\ell t}$ . In other words, for change-point analysis in the common components, Assumption 9 is reduced to requiring  $(\sqrt{J_T^*} \log^{\theta+u} T)^{-1} T^{1/4} \min_{1 \leq b \leq B_\chi} \tilde{\delta}_b \rightarrow \infty$ , as  $T \rightarrow \infty$ , allowing for the average jump size in individual coordinates  $z_{\ell t}$  to tend to zero at a rate slower than  $T^{-1/4}$ , and is no longer dependent on  $n$ . On the other hand, our model in (1) does not impose any assumption on the ‘denseness’ of  $\eta_b^\epsilon, b = 1, \dots, B_\epsilon$ , and sparse change-points (with  $m_b \ll n^2$ ) in the idiosyncratic components are detectable provided that their sparsity is compensated by the size of jumps,  $\tilde{\delta}_b$ .

Let  $\hat{\mathcal{B}} = \{\hat{\eta}_b, b = 1, \dots, \hat{B}, 1 < \hat{\eta}_1 < \dots < \hat{\eta}_{\hat{B}} < T\}$  denote the change-points detected from  $\{y_{\ell t}, \ell = 1, \dots, N; t = 1, \dots, T\}$  by the DCBS algorithm, i.e.,  $\hat{\mathcal{B}} = \hat{\mathcal{B}}^\chi(k)$  or  $\hat{\mathcal{B}} = \hat{\mathcal{B}}^\epsilon(k)$ , depending on whether  $y_{\ell t}$  is the WT of  $\hat{\chi}_{it}^k$  or  $\hat{\epsilon}_{it}^k$  for some fixed  $k$ . Accordingly, we have either  $\eta_b = \eta_b^\chi$  with  $B = B_\chi$ , or  $\eta_b = \eta_b^\epsilon$  with  $B = B_\epsilon$ . Then the following theorem establishes that the DCBS algorithm performs consistent change-point analysis for both the common and the idiosyncratic components.

**Theorem 3.** Suppose that Assumptions 1–9 hold, and let  $\theta$  be defined as in Theorem 1. Also, let the threshold  $\pi_{N, T}$  satisfy  $C' N \Delta_{N, T}^{-1} \log^{2\theta+2u} T < \pi_{N, T} < C'' \Delta_{N, T} T^{1/2}$  for some fixed  $C', C'' > 0$ . Then, there exists  $c_1 > 0$  such that

$$P(\hat{\mathcal{B}} = \mathcal{B}; |\hat{\eta}_b - \eta_b| < c_1 \omega_{N, T} \text{ for } b = 1, \dots, \hat{B}) \rightarrow 1$$

as  $T \rightarrow \infty$ , where  $\omega_{N, T} = N \Delta_{N, T}^{-2} \log^{2\theta+2u} T$ .

Theorem 3 shows both the total number and the locations of the change-points in  $\mathcal{B}^\chi$  and  $\mathcal{B}^\epsilon$  are consistently estimated; in the rescaled time  $t/T \in [0, 1]$ , the bound on the bias in estimated change-point locations satisfies  $\omega_{N, T}/T \rightarrow 0$  as  $T \rightarrow \infty$  under Assumption 9. The optimality in change-point detection may be defined as when the true and estimated change-points are within the distance of  $O_p(1)$  (Korostelev, 1987). With our approach, near-optimality in change-point estimation is achieved up to a logarithmic factor when the change-points are cross-sectionally dense ( $m_b \sim N$ ) with average size  $\tilde{\delta}_b$  bounded away from zero, so that  $\Delta_{N, T} \sim \sqrt{N}$ .

#### 4.3. Screening over a range of factor number candidates

In this section, we detail a screening procedure motivated by Theorem 2, which enables us to bypass the challenging task of estimating the number of factors in the presence of (possibly) multiple change-points in the factor structure.

The performance of most methods proposed for change-point analysis under factor modelling, such as those listed in Introduction, relies heavily on accurate estimation of the factor number. Thanks to Theorem 2, however, mis-specifying the factor number in our methodology does not influence the theoretical consistency as reported in Theorem 3, provided that we choose  $k$  sufficiently large to satisfy  $k \geq r$ . Based on this observation, we propose to screen  $\hat{\mathcal{B}}^\chi(k)$ , the set of change-points detected from  $\hat{\chi}_{it}^k$ , for a range of factor number candidates denoted as  $\mathcal{R} = \{r, r+1, \dots, \bar{r}\}$ . Specifically, we select the  $\hat{\mathcal{B}}^\chi(k)$  with the largest cardinality over  $k$  and, if there is a tie, we select  $\hat{\mathcal{B}}^\chi(k)$  with the largest  $k$ . Denoting  $k^* = \max\{k \in \mathcal{R} : |\hat{\mathcal{B}}^\chi(k)| = \max_{k' \in \mathcal{R}} |\hat{\mathcal{B}}^\chi(k')|\}$ , change-points in the idiosyncratic components are detected from  $\hat{\epsilon}_{it}^* \equiv \hat{\epsilon}_{it}^{k^*} = x_{it} - \hat{\chi}_{it}^{k^*}$ , where  $\hat{\chi}_{it}^* \equiv \hat{\chi}_{it}^{k^*}$ .

As noted below [Theorem 2](#), using  $k < r$  factors leads to  $\hat{\chi}_{it}^k$  with non-negligible estimation error which may not contain all the change-points in  $\mathcal{B}^X$  as change-points in its second-order structure. Moreover, with such  $k$ , some  $\eta_b^X$ 's may appear as change-points in  $\hat{\epsilon}_{it}^k$ , which we refer to as the spillage of change-points in the common components over to the idiosyncratic components. This justifies the proposed screening procedure and the choice of  $k^*$ .

For  $\underline{r}$ , we use the number of factors estimated by minimising the information criterion of [Bai and Ng \(2002\)](#) using as penalty  $p(n, T) = (n \wedge T)^{-1} \log(n \wedge T)$ ; in principle, any other procedure for factor number estimation may be adopted to select  $\underline{r}$ . The range  $\mathcal{R}$  also involves the choice of the maximum number of factors,  $\bar{r}$ . In the factor modelling literature, this choice is a commonly faced problem, e.g., when estimating the number of factors using an information criterion-type estimator. In the literature on stationary factor models, the maximum number of factors is usually fixed at a small number ( $\bar{r} = 8$  is used in [Bai and Ng \(2002\)](#)) for practical purposes. However, the presence of change-points tends to increase the number of factors, as shown with an example in [Appendix A](#). Therefore, we set  $\bar{r} = 20 \vee \lfloor \sqrt{n \wedge T} \rfloor$ , where the second term dominates the first when  $n$  and  $T$  are large.

## 5. Factor analysis after change-point detection

Once the change-points are detected, we can estimate the factor space over each segment  $I_b^X = [\hat{\eta}_b^X + 1, \hat{\eta}_{b+1}^X]$ ,  $b = 0, \dots, \hat{B}_X$ , defined by two consecutive change-points estimated from the common components. Denoting the sample covariance matrix over  $t \in I_b^X$  by  $\hat{\Gamma}_b^X = (\hat{\eta}_{b+1}^X - \hat{\eta}_b^X)^{-1} \sum_{t \in I_b^X} \mathbf{x}_t \mathbf{x}_t^\top$ , let  $\hat{\mathbf{w}}_{x,j}^b$  denote the eigenvector of  $\hat{\Gamma}_b^X$  corresponding to its  $j$ th largest eigenvalue. Then, for a fixed number of factors  $k$ , the segment-specific estimator of  $\chi_{it}$  is obtained via PCA

as  $\hat{\chi}_{it}^{(b,k)} = \sum_{j=1}^k \hat{w}_{x,ij}^b (\hat{\mathbf{w}}_{x,j}^b)^\top \mathbf{x}_t$ .

The number of factors  $r_b$  in the  $b$ th segment can be estimated by means of the information criterion proposed in [Bai and Ng \(2002\)](#):

$$\hat{r}_b = \arg \min_{1 \leq k \leq \bar{r}} \left[ \log \left\{ \frac{1}{n(\hat{\eta}_{b+1}^X - \hat{\eta}_b^X)} \sum_{i=1}^n \sum_{t=\hat{\eta}_b^X+1}^{\hat{\eta}_{b+1}^X} (x_{it} - \hat{\chi}_{it}^{(b,k)})^2 \right\} + kp(n, T) \right], \quad (11)$$

where  $\bar{r}$  denotes the maximum allowable factor number, and the penalty function  $p(n, T)$  satisfies  $p(n, T) \rightarrow 0$  as well as  $(n^2 \wedge \sqrt{T}) \log^{-2/\beta_f} T \cdot p(n, T) \rightarrow \infty$ . Motivated by the formulation of penalties in [Bai and Ng \(2002\)](#), we may use  $p(n, T) = (n + \sqrt{T}) / (n\sqrt{T}) \log^a(n \wedge \sqrt{T})$  for some  $a > 2/\beta_f$ .

Let  $\mathbf{A}_b$  be the  $n \times r_b$  matrix of loadings for the  $b$ th segment. In order to discuss the theoretical properties of segment-wise factor analysis, we require the following assumption that extends [Assumption 3\(i\)](#) to each segment  $I_b^X$ .

**Assumption 10.** There exists a positive definite  $r_b \times r_b$  matrix  $\mathbf{H}_b$  such that  $n^{-1} \mathbf{A}_b^\top \mathbf{A}_b \rightarrow \mathbf{H}_b$  as  $n \rightarrow \infty$ .

Then, we obtain the asymptotic results in [Propositions 2](#) and [3](#) for the segment-wise estimators of the factor number and the common components.

**Proposition 2.** Suppose that all the conditions in [Theorem 3](#) and [Assumption 10](#) hold. For all  $b = 0, \dots, \hat{B}_X$ ,  $\hat{r}_b$  returned by (11) satisfies  $P(\hat{r}_b = r_b) \rightarrow 1$  as  $n, T \rightarrow \infty$ .

**Proposition 3.** Suppose that all the conditions in [Theorem 3](#) and [Assumption 10](#) hold. Then, for all  $b = 0, \dots, \hat{B}_X$ ,

$$\max_{1 \leq i \leq n} \max_{t \in I_b^X} |\hat{\chi}_{it}^{(b, \hat{r}_b)} - \chi_{it}| = O_p \left\{ \left( \sqrt{\frac{\log n}{T}} \vee \frac{1}{\sqrt{n}} \right) \log^\theta T \right\}.$$

From [Propositions 2](#) and [3](#), we have the guarantee that the factor space is consistently estimated by PCA over each estimated segment. We may further refine the post change-point analysis by first determining whether  $\hat{\eta}_b^X$  can be associated with (a) a break in the loadings or factor number, or (b) that in the autocorrelation structure of the factors only. This can be accomplished by comparing  $\hat{r}_{b-1}$  and  $\hat{r}_b$  against the factor number estimated from the pooled segment  $I_{b-1}^X \cup I_b^X$ : a break in the loadings or factor number necessarily brings in a change in the number of factors from the pooled segment. However, if (b) is the case, the segments before and after  $\hat{\eta}_b^X$  as well as the pooled one return the identical number of factors, and we can perform the joint factor analysis of the two segments.

## 6. Computational aspects

### 6.1. Bootstrap procedure for threshold selection

The range of theoretical rates supplied for  $\pi_{N,T}$  in [Theorem 3](#), involves typically unattainable knowledge of the minimum cross-sectional size of the changes. Hence, we propose a bootstrap procedure for the selection of  $\pi_{N,T}^X$  and  $\pi_{N,T}^\epsilon$ , the thresholds for change-point analysis of  $\hat{\chi}_{it}^k$  and  $\hat{\epsilon}_{it}^k$ , respectively. We omit  $N$  and  $T$  from their subscripts for notational convenience when



there is no confusion. Although a formal proof on the validity of the proposed bootstrap algorithm is beyond the scope of the current paper, simulation studies reported in Section 7 demonstrate its good performance when applied jointly with our proposed methodology. We refer to [Trapani \(2013\)](#), [Corradi and Swanson \(2014\)](#), [Gonçalves and Perron \(2014, 2016\)](#) for alternative bootstrap methods under factor models and [Jentsch and Politis \(2015\)](#) for the linear process bootstrap for multivariate time series in general.

We propose a new bootstrap procedure which is specifically motivated by the separate treatment of common and idiosyncratic components in our change-point detection methodology. Namely, the resampling method produces bootstrap samples from the common and idiosyncratic components independently, relying on the consistency of the estimated components with an over-specified factor number as reported in [Theorem 2](#).

Let  $\mathcal{T}_{s,e}^{\chi}(k)$  and  $\mathcal{T}_{s,e}^{\epsilon}(k)$  denote the test statistics  $\mathcal{T}_{s,e}$  computed on the interval  $[s, e]$  for the panel data obtained from the WT of  $\hat{\chi}_{it}^k$  and  $\hat{\epsilon}_{it}^k$ , respectively. The proposed resampling procedure aims at approximating the distributions of  $\mathcal{T}_{s,e}^{\chi}(k)$  and  $\mathcal{T}_{s,e}^{\epsilon}(k)$  under the null hypothesis of no change-point, which then can be used for the selection of  $\pi_{s,e}^{\chi}(k)$  and  $\pi_{s,e}^{\epsilon}(k)$ , the corresponding, interval-specific test criteria for  $\hat{\chi}_{it}^k$  and  $\hat{\epsilon}_{it}^k$  over  $[s, e]$ , an interval which is considered at some iteration of the DCBS algorithm.

Among the many block bootstrap procedures proposed in the literature for bootstrapping time series (see [Politis and White, 2004](#)) for an overview), stationary bootstrap (SB) proposed in [Politis and Romano \(1994\)](#) generates bootstrap samples which are stationary conditional on the observed data (see [Appendix D.4](#) for details). Based on the SB, our procedure derives  $\pi_{s,e}^{\chi}(k)$  and  $\pi_{s,e}^{\epsilon}(k)$ . Recall that  $\hat{\lambda}_{ij}$  and  $\hat{f}_{jt}$  denote the loadings and factors estimated via the capped PCA for  $j = 1, \dots, k$ , see [Section 3.1](#).

Stationary bootstrap algorithm for factor models.

Step 1 For the common components: For each  $l \in \{1, \dots, k\}$ , produce the SB sample of  $\{\hat{f}_{lt}, t = 1, \dots, T\}$  as  $\{f_{lt}^{\bullet}, t = 1, \dots, T\}$ . Compute  $\chi_{it}^{k\bullet} = \sum_{l=1}^k \hat{\lambda}_{il} f_{lt}^{\bullet}$ .

For the idiosyncratic components: Produce the SB sample of  $\{\hat{\epsilon}_t^k = (\hat{\epsilon}_{1t}^k, \dots, \hat{\epsilon}_{nt}^k)^{\top}, t = 1, \dots, T\}$  as  $\{\epsilon_t^{k\bullet} = (\epsilon_{1t}^{k\bullet}, \dots, \epsilon_{nt}^{k\bullet})^{\top}, t = 1, \dots, T\}$ .

Step 2: Generate  $y_{it}^{\bullet}$  through transforming  $\chi_{it}^{k\bullet}$  or  $\epsilon_{it}^{k\bullet}$  using  $g_j(\cdot)$  and  $h_j(\cdot, \cdot)$  as described in [Section 3.3](#).

Step 3: Compute  $\mathcal{Y}_{s,b,e}^{k\bullet}$  on  $y_{it}^{\bullet}$  and generate the test statistic  $\mathcal{T}_{s,e}^{\bullet}$  according to [\(10\)](#).

Step 4: Repeat Steps 1–3  $R$  times. The critical value  $\pi_{s,e}^{\chi}(k)$  or  $\pi_{s,e}^{\epsilon}(k)$  for the segment  $[s, e]$  is selected as the  $(1 - \alpha)$ -quantile of the  $R$  bootstrap test statistics  $\mathcal{T}_{s,e}^{\bullet}$  at given  $\alpha \in (0, 1)$ .

The bootstrap algorithm is designed to produce  $\chi_{it}^{k\bullet}$  and  $\epsilon_{it}^{k\bullet}$  that mimic the second-order structure of  $\hat{\chi}_{it}^k$  and  $\hat{\epsilon}_{it}^k$ , respectively, when there is no change-point present, and thus approximates the distributions of the test statistics under the null hypothesis. In the algorithm, the treatment of the common and idiosyncratic components differs only in the application of SB in Step 1: since the factors estimated from the PCA are uncorrelated by construction, we generate the SB samples of  $\hat{f}_{jt}$  for each  $j$  separately, while the  $n$  elements of  $\hat{\epsilon}_t^k$  are resampled jointly in an attempt to preserve the cross-sectional dependence therein. We discuss the choice of the bootstrap size  $R$  and the level of quantile  $\alpha$  in [Section 6.2](#).

## 6.2. Selection of tuning parameters

The proposed methodology involves the choice of tuning parameters for the capped PCA, WT, DCBS algorithm and the bootstrap procedure. We here list the values used for the simulation studies ([Section 7](#)) and real data analysis ([Section 8](#)). We provide further guidance on the implementation of the proposed methodology in [Appendix D](#).

In the examples considered in [Section 7](#), we have not observed unreasonably large contributions to  $\hat{\chi}_{it}^k$  from spurious factors. As discussed in [Section 3.1](#), selecting a sufficiently large constant as  $c_w$  effectively disables the capping for the (unknown)  $r$  leading principal components to  $\hat{\chi}_{it}^k$ . For this reason, in the current implementation, we disable the capping. However, this does not necessarily mean that capping will always be of no practical use. Therefore, we recommend the data-driven choice of  $c_w = \sqrt{n} \max_{1 \leq j \leq \bar{r}} \max_{1 \leq i \leq n} |\hat{w}_{x,ij}|$ . With such  $c_w$ , the capping is enabled for only those  $\hat{w}_{x,ij}$  with  $\bar{r} + 1 \leq j \leq \bar{r}$ , where  $\bar{r}$  and  $\bar{r}$  denote the smallest and largest number of factors considered in [Section 4.3](#).

For the WT, we propose to use  $J_T^* = \lfloor C \log_2 \log^{\nu} T \rfloor$  number of finest Haar wavelets for some  $\nu \in (0, 1]$  and  $C > 0$ , in order to control for any bias in change-point estimation arising from WT. Noting that the bias increases at the rate  $2^C$  with increasing  $C$ , we recommend the choice of  $J_T^* = \lfloor \log_2 \log_2 T \rfloor$  in practice.

Although omitted from the description of the DCBS algorithm, we select a parameter  $d_T$  controlling the minimum distance between two estimated change-points. In light of [Remark 2](#) and [Theorem 3](#), we choose to use  $d_T = \lfloor \log^2 T \wedge 0.25T^{6/7} \rfloor$ . Note that we can avoid using this parameter by replacing the binary segmentation procedure with wild binary segmentation ([Fryzlewicz, 2014](#)), such that the Double CUSUM is applied over randomly drawn intervals to derive the test statistic. It is conjectured that such a procedure will place a tighter bound on the bias in estimated change-point locations, as well as bypassing the need for the parameter  $d_T$ . We leave the investigation in this direction for the future research.

Finally, for the proposed bootstrap procedure, we use the bootstrap sample size  $R = 200$  for the simulation studies and  $R = 500$  for the real data analysis. As for the level of quantile, we select  $\alpha = 0.05$ ; note that this choice does not indicate the significance level in hypothesis testing.

## 7. Simulation studies

In this section, we apply the proposed change-point detection methodology to both single and multiple change-point scenarios under factor modelling. While our methodology is designed for multiple change-point detection, Section 7.1 gives us insight into its performance in the presence of a single change-point, which is of the type, size and denseness that vary in a systematic manner. Multiple change-point scenarios are considered in Section 7.2.

### 7.1. Single change-point scenarios

The following stationary  $q$ -factor model allows for serial correlations in  $f_{jt}$  and both serial and cross-sectional correlations in  $\epsilon_{it}$ :

$$x_{it} = \sum_{j=1}^q \lambda_{ij} f_{jt} + \sqrt{\vartheta} \epsilon_{it}, \quad \text{where} \quad \vartheta = \phi \cdot \frac{q}{1 - \rho_f^2} \cdot \frac{1 - \rho^2}{1 + 2H\beta^2}, \quad (12)$$

$$f_{jt} = \rho_{f,j} f_{j,t-1} + u_{jt}, \quad u_{jt} \sim \text{iid} \mathcal{N}(0, 1), \quad (13)$$

$$\epsilon_{it} = \rho_{\epsilon,i} \epsilon_{i,t-1} + v_{it} + \beta_i \sum_{|k| \leq H_i, k \neq 0} v_{i+k,t}, \quad v_{it} \sim \text{iid} \mathcal{N}(0, 1), \quad (14)$$

and  $\lambda_{ij} \sim \text{iid} \mathcal{N}(0, 1)$ . The stationary factor model in (12) has been frequently adopted for empirical studies in the factor model literature including (Bai and Ng, 2002). Throughout we set  $q = 5$ . The parameters  $\rho_{f,j} = \rho_f - 0.05(j - 1)$  with  $\rho_f = 0.4$ , and  $\rho_{\epsilon,i} \sim \text{iid} \mathcal{U}(-\rho, \rho)$  with  $\rho = 0.5$ , determine the autocorrelations in the factors and idiosyncratic components, while  $H_i = H = \min(n/20, 10)$  and  $\beta_i$  (randomly drawn from  $\{-\beta, \beta\}$  with  $\beta = 0.2$ ), determine the cross-sectional correlations in  $\epsilon_t$ . The parameter  $\phi$ , or its inverse  $\phi^{-1}$ , is chosen from  $\{1, 1.5, 2, 2.5\}$  depending on the change-point scenario, in order to investigate the impact of the ratio between the variance of the common and idiosyncratic components, on the performance of the change-point detection methodology. We fix the number of observations at  $T = 200$  and vary the dimensionality as  $n \in \{100, 300\}$ .

A single change-point is introduced to either  $x_{it}$  or  $\epsilon_{it}$  at  $\eta_1 = [T/3] = 67$  as follows.

- (S1) Change in the loadings. For a randomly chosen index set  $S \subset \{1, \dots, n\}$ , the loadings  $\lambda_{ij}$ ,  $i \in S$  are shifted by  $\Delta_{ij} \sim \text{iid} \mathcal{N}(0, \sigma^2)$ , where  $\sigma \in \{\sqrt{2}, 1, 0.75, 0.5, 0.25\}$ .
- (S2) Change in  $\rho_{f,j}$ . The signs of the AR parameters in (13) are switched such that the autocorrelations of  $f_{jt}$  change while their variance remains the same.
- (S3) A new factor. For a randomly chosen  $S \subset \{1, \dots, n\}$ , a new factor is introduced to  $x_{it}$ ,  $i \in S$ :  $f_{q+1,t} = \rho_{f,q+1} f_{q+1,t-1} + u_{q+1,t}$  with  $u_{q+1,t} \sim \text{iid} \mathcal{N}(0, 1)$  and  $\lambda_{i,q+1} \sim \text{iid} \mathcal{N}(0, \sigma^2)$ , where  $\sigma = \sqrt{2}$ .
- (S4) Change in  $\rho_{\epsilon,i}$ . For a randomly chosen  $S \subset \{1, \dots, n\}$ , the corresponding  $\rho_{\epsilon,i}$ ,  $i \in S$  in (14) have their signs switched so that the autocorrelations of such  $\epsilon_{it}$  change while their variance remains the same.
- (S5) Change in the covariance of  $\epsilon_t$ . For a randomly chosen  $S \subset \{1, \dots, n\}$ , the bandwidth  $H_i$ ,  $i \in S$  in (14) doubles.

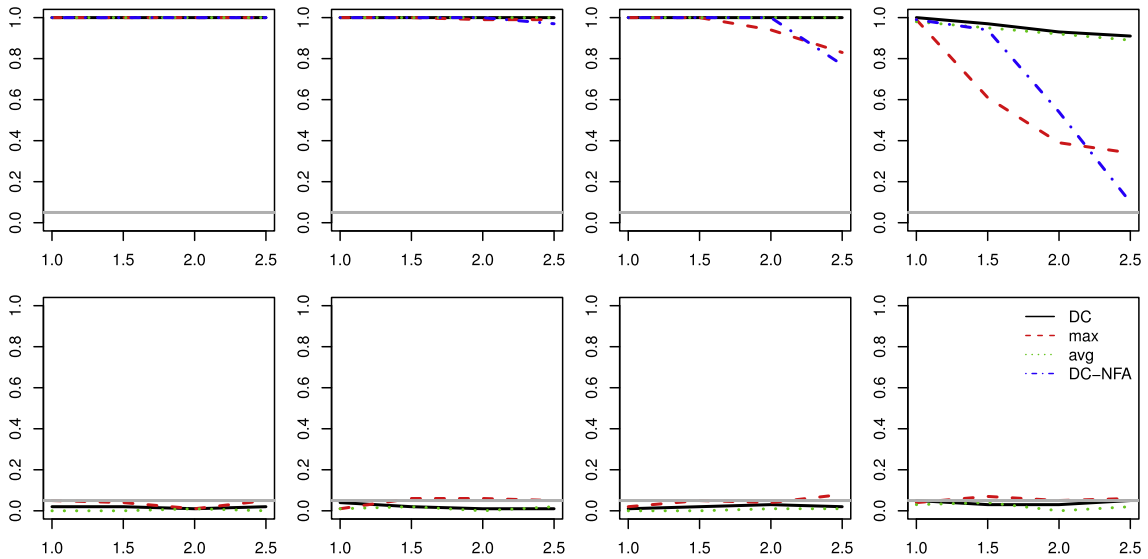
In (S1) and (S3)–(S5), the size of the index set  $S$  is controlled by the parameter  $\varrho \in \{1, 0.75, 0.5, 0.25\}$ , as  $|S| = [\varrho n]$ .

For each simulated dataset, we consider the range of possible factor numbers  $k \in \mathcal{R}$  selected as described in Section 4.3. For any given  $k$ , we estimate the common and idiosyncratic components and proceed with WT of Stage 1, to which *only the first iteration* of the DCBS algorithm is applied. In this way, we test for the existence of at most a single change-point in the common and idiosyncratic components separately and, if its presence is detected, we identify its location in time. With a slight abuse of notation, we refer to the simultaneous testing and locating procedure as the DC test, and report its detection power and accuracy in change-point estimation.

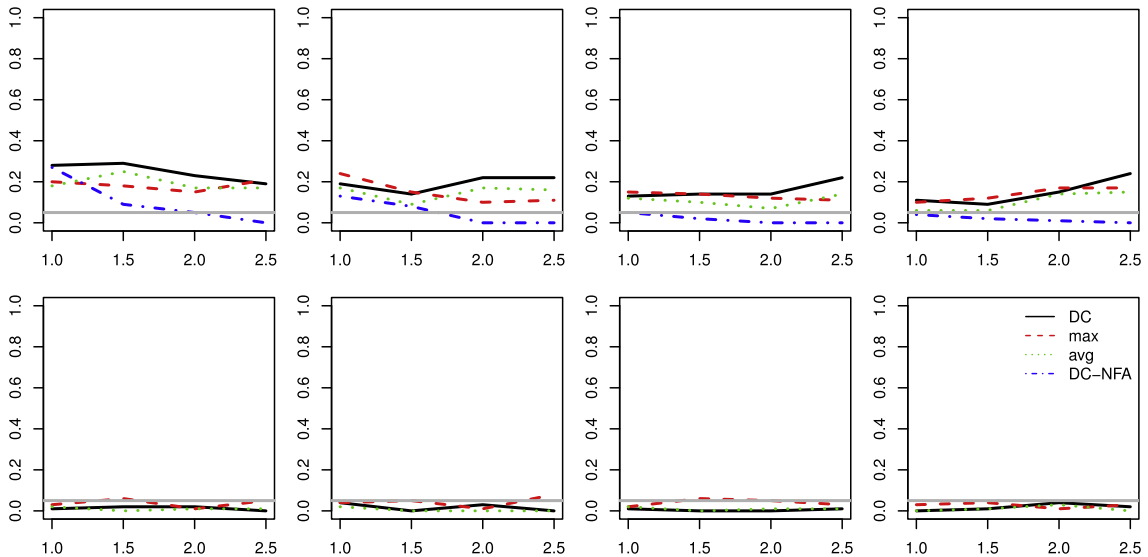
As noted in Introduction, existing methods for change-point analysis under factor modelling are not applicable to the scenarios other than (S1) and (S3). Hence, we compare the performance of the DC test to change-point tests based on two other high-dimensional CUSUM aggregation approaches, which are referred to as the MAX and AVG tests: after Stage 1, the  $N$ -dimensional CUSUMs of the WT series are aggregated via taking their point-wise maximum or average, respectively, which replaces the Double CUSUM statistics in Stage 2. In addition, we report the detection power of a variant of the DC test, where the first iteration of the DCBS algorithm is applied to the panel data consisting of the WT of  $x_{it}$ , and compare its performance against DC, MAX and AVG tests applied to the WT of  $\hat{x}_{it}^k$  under (S1)–(S3), and that of  $\hat{\epsilon}_{it}^k$  under (S4)–(S5). We include this approach, termed the DC-NFA (no factor analysis) test, in order to demonstrate the advantage in linking the factor modelling and change-point detection as proposed in our methodology.

Figs. 2–7 plot the detection power of our change-point test as well as that of MAX, AVG and DC-NFA tests under different scenarios over 100 realisations when  $n = 100$  (we only present the results from (S1) when  $\sigma \in \{\sqrt{2}, 0.02\sqrt{2}\}$ ). Since  $x_{it}$  ( $\epsilon_{it}$ ) do not contain any change under (S4)–(S5) ((S1)–(S3)), testing for a change-point in  $\hat{\epsilon}_{it}^*$  ( $\hat{x}_{it}^*$ ) in these scenarios offers insights into the size behaviour of DC, MAX and AVG tests.

Overall, change-point detection becomes more challenging as  $\sigma$  (the size of changes) or  $\varrho$  (the proportion of the coordinates with the change) decreases, and also as  $\text{Var}(x_{it})/\text{Var}(\epsilon_{it})$  decreases (with increasing  $\phi$ ) under (S1)–(S3) and



**Fig. 2.** (S1) Detection power (y-axis) of change-point tests on the common (top) and idiosyncratic (bottom) components with varying  $\phi \in \{1, 1.5, 2, 2.5\}$  (x-axis) when  $n = 100$ ,  $T = 200$ ,  $\sigma = \sqrt{2}$  and  $\varrho \in \{1, 0.75, 0.5, 0.25\}$  (left to right); horizontal grey lines indicate the significance level  $\alpha = 0.05$ .

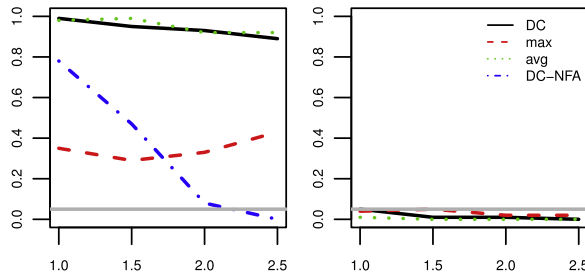


**Fig. 3.** (S1) Detection power of change-point tests when  $\sigma = 0.25\sqrt{2}$ .

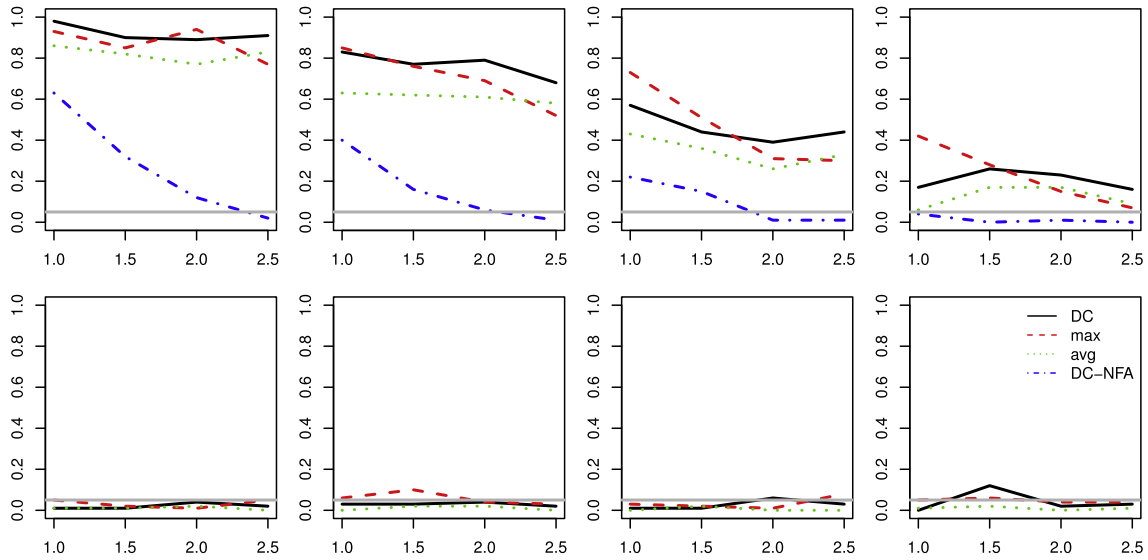
increases (with increasing  $\phi^{-1}$ ) under (S4)–(S5). In all scenarios, the DC test shows superior performance to the DC-NFA test. This confirms that the factor analysis prior to change-point analysis is an essential step in markedly improving the detection power, as well as in identifying the origins of the change-points; without a factor analysis, a change-point that appears in either of  $\chi_{it}$  or  $\epsilon_{it}$  may be ‘masked’ by the presence of the other component and thus escape detection.

The DC and AVG tests generally attain similar powers, while the MAX test tends to have considerably lower power in scenarios such as (S2), (S4) and (S5). An exception is under (S3), where the MAX test attains larger power than the others in some settings with decreasing  $\varrho$  (Fig. 5). It may be explained by the fact that the smaller  $\varrho$  is, the sparser the change-point becomes cross-sectionally, which is a setting that favours the approach taken in MAX in aggregating the high-dimensional CUSUM series (see the discussions in Section 2.1 of Cho and Fryzlewicz (2015)). Between the DC and AVG tests, the former outperforms the latter in the more challenging settings when the change is sparse cross-sectionally (with small  $\varrho$ ), particularly when the change is attributed to the introduction of a single factor to the existing five as in (S3).

Apart from (S5), the origin of the change-point is correctly identified in the sense that it is detected only from  $\hat{\chi}_{it}^*$  under (S1)–(S3) or  $\hat{\epsilon}_{it}^*$  under (S4) with power strictly above  $\alpha = 0.05$ . In (S5), we observe spillage of the change-point in  $\epsilon_{it}$ :



**Fig. 4.** (S2) Detection power of change-point tests on the common (left) and idiosyncratic (right) components with varying  $\phi \in \{1, 1.5, 2, 2.5\}$  (x-axis) when  $n = 100$  and  $T = 200$ .



**Fig. 5.** (S3) Detection power of change-point tests on the common (top) and idiosyncratic (bottom) components with varying  $\phi \in \{1, 1.5, 2, 2.5\}$  (x-axis) when  $n = 100$ ,  $T = 200$ ,  $\sigma = \sqrt{2}$  and  $\varrho \in \{1, 0.75, 0.5, 0.25\}$  (left to right).

the change-point is detected from both  $\hat{\chi}_{it}^*$  and  $\hat{\epsilon}_{it}^*$ , particularly when a large proportion of  $\epsilon_{it}$  undergoes the change of an increase in the bandwidth  $H_i$  in (14), and when  $\text{Var}(\chi_{it})/\text{Var}(\epsilon_{it})$  is small (with large  $\phi$ ), see Fig. 7. This supports the notion of a pervasive change-point being a factor, i.e., a significant break that affects the majority of the idiosyncratic components in their second-order structure, may be viewed as a *common* feature. We also note that due to the relatively large variance of common component, the detection power of the DC-NFA test behaves as that of the DC test applied to  $\hat{\chi}_{it}^*$  rather than  $\hat{\epsilon}_{it}^*$  in this scenario.

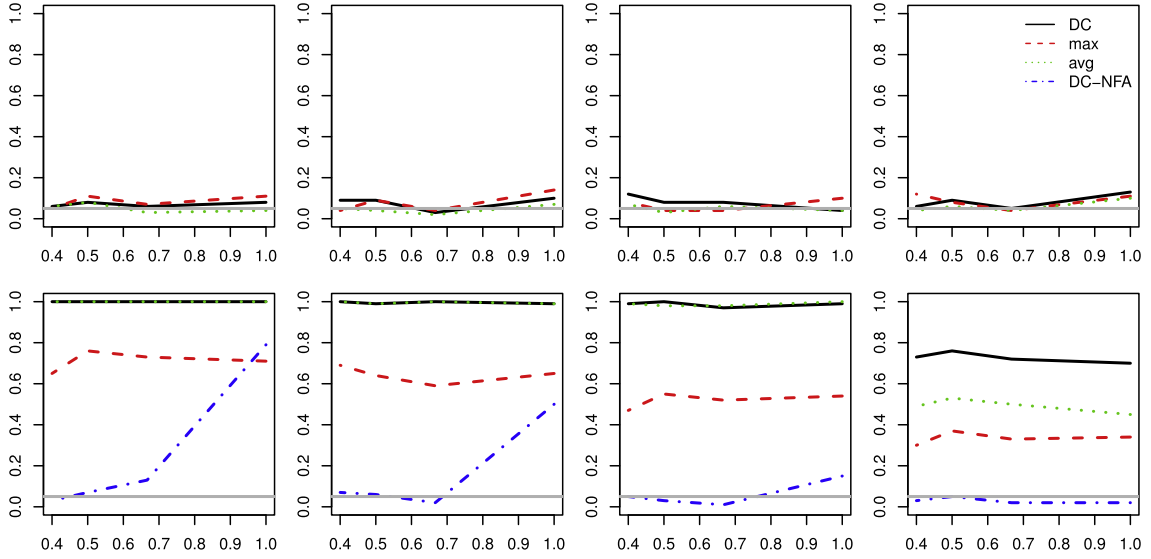
We present in a supplementary document the results on other data generating processes taken from the literature on testing for a single change-point under factor models, as well as tables and figures summarising the simulation results for (S1)–(S5) with  $n = 300$ , along with box plots of the estimated change-points  $\hat{\eta}_1$ . The performance of all tests under consideration generally improve when  $n = 300$ .

## 7.2. Multiple change-point scenarios

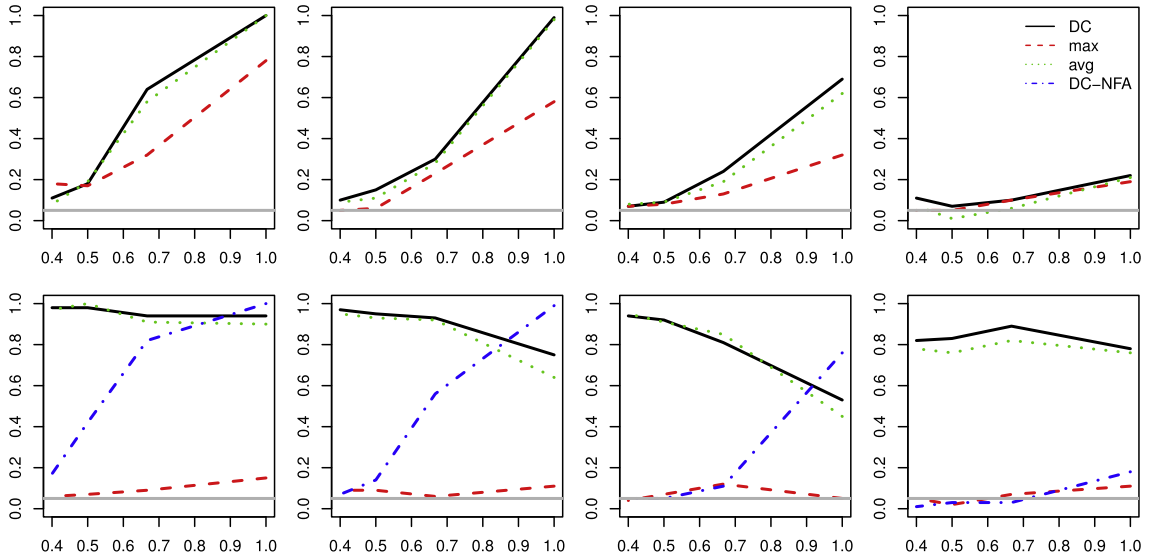
### 7.2.1. Model (M1)

This model is intended to mimic the behaviour of the Standard and Poor's 100 log-return data, analysed in Section 8.1. The information criterion of Bai and Ng (2002) returned  $q = 4$  factors for these data and, imposing stability on the loadings, common factors four and idiosyncratic components were estimated by means of PCA as  $\hat{f}_{jt}$ ,  $j = 1, \dots, q$  and  $\hat{\epsilon}_{it}$ . Estimated factors exhibit little serial correlations, show the evidence of multiple change-points in their variance and heavy tails. The same evidence holds for the estimated idiosyncratic components. Based on these observations, we adopt the following data generating model:

$$x_{it} = \chi_{it} + \sqrt{\vartheta} \epsilon_{it} = \sum_{j=1}^q \hat{\lambda}_{ij} \hat{f}_{jt} + \sqrt{\vartheta} \epsilon_{it}, \quad \text{where} \quad \vartheta = \phi \cdot \sum_{i=1}^n \widehat{\text{Var}}(\chi_{it}) \left\{ \sum_{i=1}^n \widehat{\text{Var}}(\epsilon_{it}) \right\}^{-1},$$



**Fig. 6.** (S4) Detection power of change-point tests on the common (top) and idiosyncratic (bottom) components with varying  $\phi \in \{2.5^{-1}, 2^{-1}, 1.5^{-1}, 1\}$  ( $x$ -axis) when  $n = 100$ ,  $T = 200$  and  $\varrho \in \{1, 0.75, 0.5, 0.25\}$  (left to right).



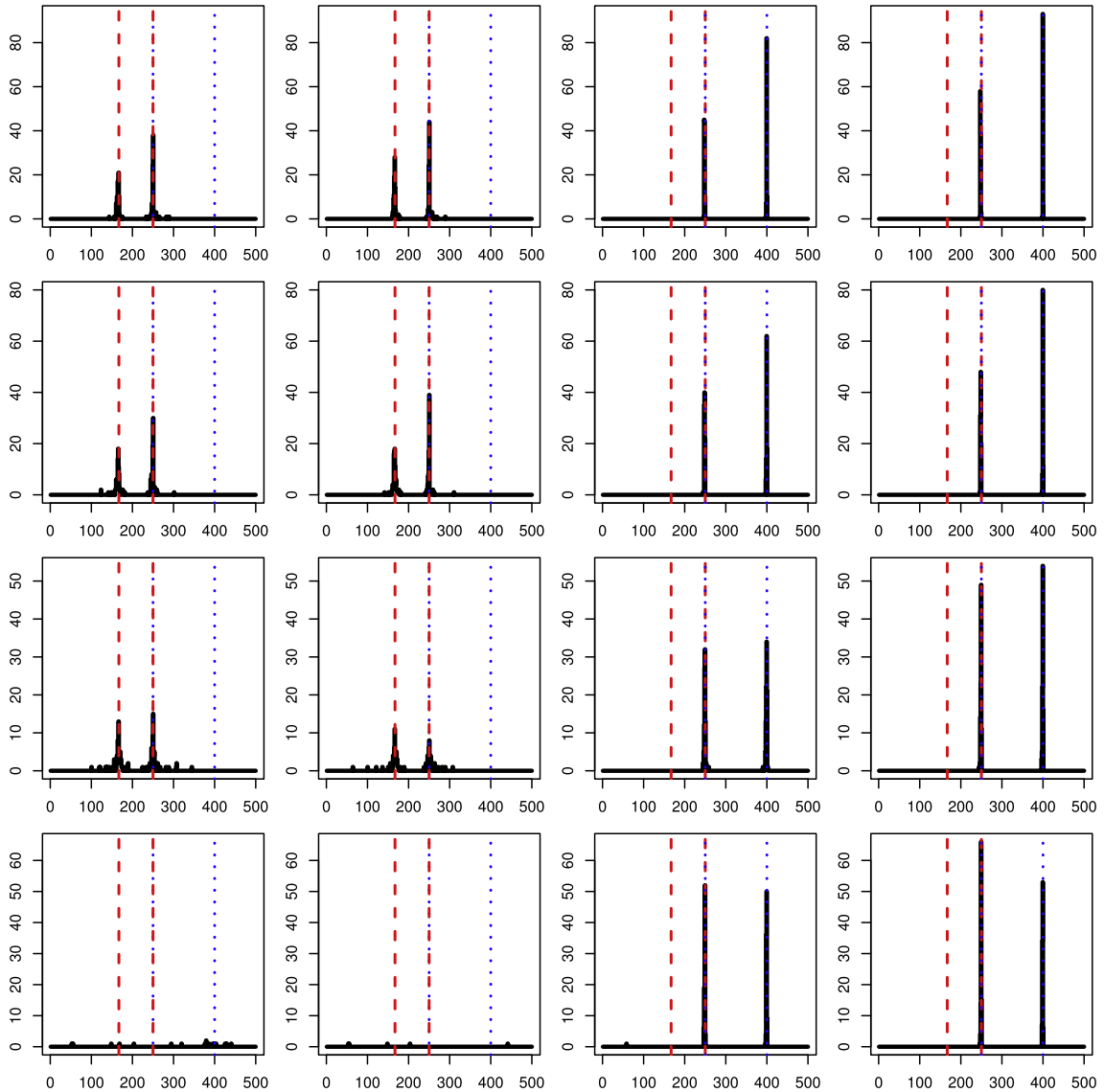
**Fig. 7.** (S5) Detection power of change-point tests on the common (top) and idiosyncratic (bottom) components with varying  $\phi \in \{2.5^{-1}, 2^{-1}, 1.5^{-1}, 1\}$  ( $x$ -axis) when  $n = 100$ ,  $T = 200$  and  $\varrho \in \{1, 0.75, 0.5, 0.25\}$  (left to right).

$$f_{jt} = \begin{cases} u_{j,t} & \text{for } 1 \leq t \leq \eta_1^x, \\ (\sigma \delta_{j,b}^f)^{g_b} \cdot u_{jt} & \text{for } \eta_b^x + 1 \leq t \leq \eta_{b+1}^x \text{ with } b \geq 1, \end{cases}$$

$$\epsilon_{it} = \begin{cases} v_{it} & \text{for } 1 \leq t \leq \eta_1^\epsilon, \\ (\sigma \delta_{i,b}^\epsilon)^{g_b} \cdot v_{it} & \text{for } \eta_b^\epsilon + 1 \leq t \leq \eta_{b+1}^\epsilon \text{ with } b \geq 1, \end{cases}$$

where  $\widehat{\text{Var}}(\cdot)$  denotes the sample variance operator,  $u_{jt}, v_{it} \sim_{\text{iid}} t_7$  and  $g_b = 1$  for even  $b$  and  $g_b = -1$  otherwise. We use the loadings estimated from the S&P100 data without capping, denoted by  $\widehat{\lambda}_{ij}$ , and  $\delta_{j,b}^f$  ( $\delta_{i,b}^\epsilon$ ) is chosen from the same dataset by contrasting pairs of intervals with visibly different  $\widehat{\text{Var}}(f_{jt})$  ( $\widehat{\text{Var}}(\epsilon_{it})$ ). The change-points in the common components are introduced to  $\text{Var}(f_{jt})$  at  $\eta_1^x = \lceil T/3 \rceil$  and  $\eta_2^x = \lceil T/2 \rceil$ , and those in the idiosyncratic components to  $\text{Var}(\epsilon_{it})$  at  $\eta_1^\epsilon = \lceil T/2 \rceil$  and  $\eta_2^\epsilon = \lceil 4T/5 \rceil$ . The magnitude of each change is controlled by  $\sigma \in \{1, 0.75, 0.5, 0.25\}$ , while  $\phi \in \{1, 1.5, 2, 2.5\}$  determines





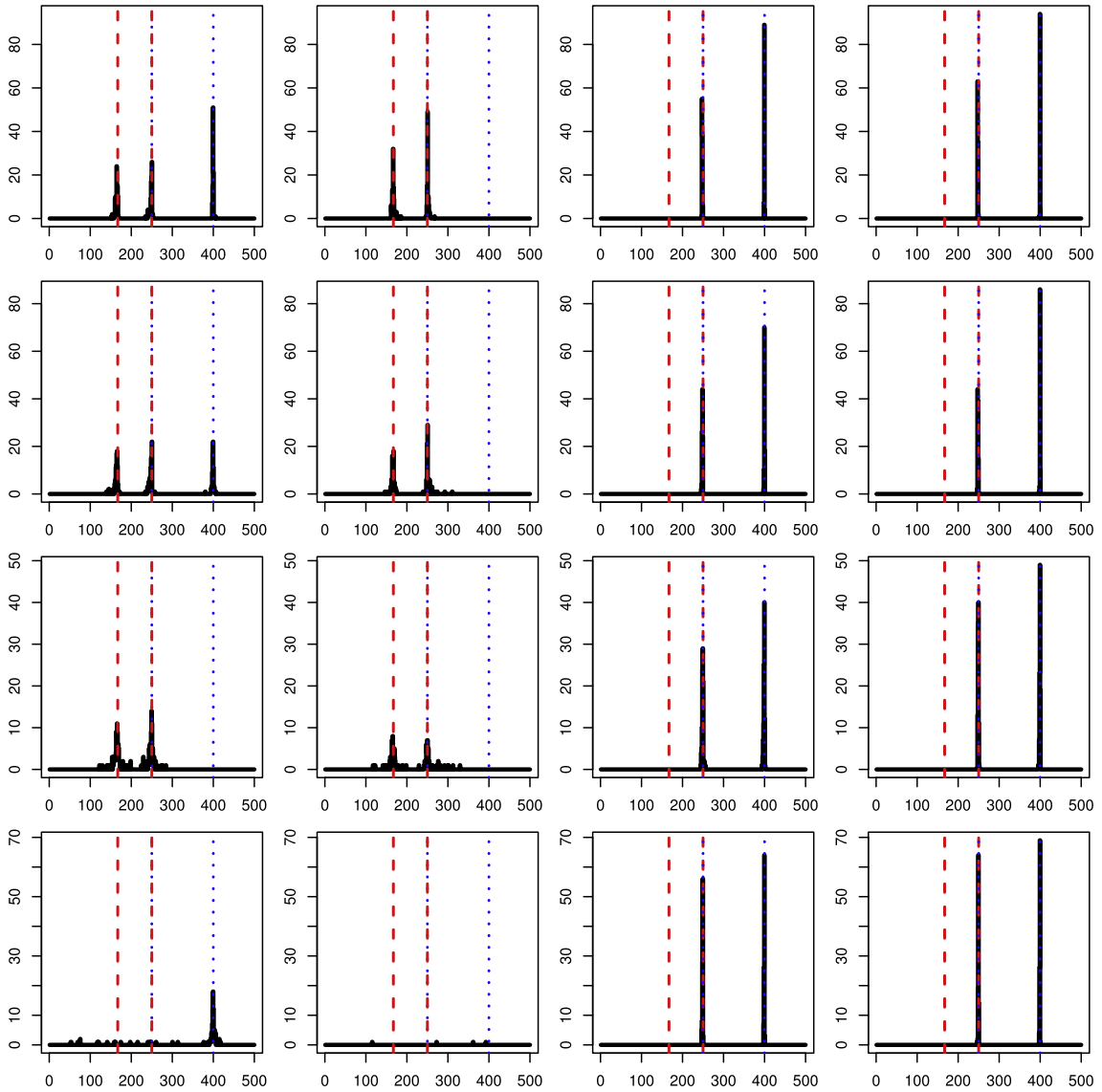
**Fig. 8.** (M1) Locations of the change-points estimated from  $\widehat{\chi}_{it}^*$ ,  $\chi_{it}$  (oracle),  $\widehat{\epsilon}_{it}^*$  and  $\epsilon_{it}$  (oracle) by the DCBS algorithm (left to right) for  $\sigma \in \{1, 0.75, 0.5, 0.25\}$  (top to bottom) when  $n = 100$ ,  $T = 500$  and  $\phi = 1$ ; vertical lines indicate the locations of the true change-points  $\eta_b^kappa$ ,  $b = 1, 2$  (dashed) and  $\eta_b^epsilon$ ,  $b = 1, 2$  (dotted); recall that  $\eta_2^kappa = \eta_1^epsilon$ .

the ratio between the variance of the common and idiosyncratic components (the larger  $\phi$ , the smaller  $\text{Var}(\chi_{it})/\text{Var}(x_{it})$  is). We fix  $T = 500$  and vary  $n \in \{100, 300\}$ .

We report the performance of our methodology over 100 realisations when  $n = 100$  in Figs. 8 and 9 for the two extreme cases with  $\phi \in \{1, 2.5\}$ . We also apply the DCBS algorithm to the WT of true common and idiosyncratic components generated under (M1) and report the corresponding results, which serve as a benchmark against which the efficacy of the PCA-based factor analysis of the proposed methodology is assessed. Additional simulation results with varying  $\phi$  and  $n$  are reported in the supplementary document.

The DCBS algorithm detects the two change-points in the idiosyncratic components equally well from  $\widehat{\epsilon}_{it}^*$  and  $\epsilon_{it}$ , regardless of the values of  $\phi$  and  $\sigma$ . On the other hand, detection of  $\eta_b^kappa$ ,  $b = 1, 2$  is highly variable with respect to these parameters. When the size of the change is large ( $\sigma \in \{1, 0.75\}$ ), the panel data generated from transforming  $\widehat{\chi}_{it}^*$  serves as good an input to the DCBS algorithm as that generated from transforming  $\chi_{it}$  in terms of translating the presence and locations of both change-points.

With decreasing  $\text{Var}(\chi_{it})/\text{Var}(x_{it})$ , the change-point  $\eta_2^epsilon$ , which appears only in the idiosyncratic components, is detected with increasing frequency as a change-point in the common components from  $\widehat{\chi}_{it}^*$ , when (a)  $\sigma \geq 0.75$  (the change in  $\epsilon_{it}$



**Fig. 9.** (M1) Locations of the estimated change-points when  $n = 100$ ,  $T = 500$  and  $\phi = 2.5$ .

is large and thus all three change-points are detected from  $\hat{\chi}_{it}^*$ , or (b)  $\sigma = 0.25$  (the changes in  $\chi_{it}$  are ignored and a single change-point is detected at  $\eta_2^c$  from  $\hat{\chi}_{it}^*$ ). This phenomenon is in line with the observation made for the model (S5) in Section 7.1, on the spillage of change-points in the idiosyncratic components over to the common components: a significant co-movement in the dependence structure of  $\epsilon_{it}$  may be regarded as being pervasive and common, and hence is captured as a change in the dependence structure of the common components by our proposed methodology.

Lastly, we note that although we impose normality on the idiosyncratic components for the theoretical development, these results show that our methodology works well even when the data exhibits some deviations of normality, such as heavy-tails.

### 7.2.2. Model (M2)

In this model, change-points are introduced as in (S1)–(S4) of Section 7.1. More specifically,

$$\lambda_{ij,t} = \begin{cases} \lambda_{ij} & \text{for } 1 \leq t \leq \eta_1^x = [T/3], \\ \lambda_{ij} + \Delta_{ij}\mathbb{I}(i \in S_1^x) & \text{for } \eta_1^x + 1 \leq t \leq T, \end{cases}$$

$$f_{jt} = \begin{cases} \rho_f f_{j,t-1} + u_{jt}, & \text{for } 1 \leq t \leq \eta_2^x = [T/2], \\ -\rho_f f_{j,t-1} + u_{jt}, & \text{for } \eta_2^x = [T/2] + 1 \leq t \leq T, \end{cases}$$

$$\epsilon_{it} = \begin{cases} \rho_{\epsilon,i} \epsilon_{i,t-1} + v_{it} + \beta_i \sum_{\substack{|k| \leq H \\ k \neq 0}} v_{i+k,t} & \text{for } 1 \leq t \leq \eta_1^\epsilon = \lfloor 3T/5 \rfloor, \\ \rho_{\epsilon,i} \{\mathbb{I}(i \notin S_1^\epsilon) - \mathbb{I}(i \in S_1^\epsilon)\} \epsilon_{i,t-1} + v_{it} + \beta_i \sum_{\substack{|k| \leq H \\ k \neq 0}} v_{i+k,t} & \text{for } \eta_1^\epsilon + 1 \leq t \leq T, \end{cases}$$

$$x_{it} = \begin{cases} \sum_{j=1}^q \lambda_{ij,t} f_{jt} + \sqrt{\vartheta} \epsilon_{it} & \text{for } 1 \leq t \leq \eta_3^x = \lfloor 4T/5 \rfloor, \\ \sum_{j=1}^q \lambda_{ij,t} f_{jt} + \sqrt{2} \lambda_{i,q+1} f_{q+1,t} \mathbb{I}(i \in S_3^x) + \sqrt{\vartheta} \epsilon_{it} & \text{for } \eta_3^x + 1 \leq t \leq T, \end{cases}$$

with  $q = 5$  and  $\lambda_{ij}, u_{jt}, v_{it} \sim \text{iid} \mathcal{N}(0, 1)$ . The parameters  $\rho_{f,j}, \rho_{\epsilon,i}, \beta_i, H$  and  $\vartheta$  are chosen identically as in Section 7.1. In summary, three change-points in the common components are introduced to the loadings ( $\eta_1^x$ ), autocorrelations of the factors ( $\eta_2^x$ ) and the number of factors ( $\eta_3^x$ ), while a single change-point in the idiosyncratic components is introduced to their serial correlations ( $\eta_1^\epsilon$ ). The cardinality of the index sets  $S_1^x, S_2^x$  and  $S_1^\epsilon$  determines the sparsity of the change-points  $\eta_1^x, \eta_2^x$  and  $\eta_1^\epsilon$ , respectively. We randomly draw each index set from  $\{1, \dots, n\}$  with its cardinality set at  $\lfloor \varrho n \rfloor$ , where  $\varrho \in \{1, 0.75, 0.5, 0.25\}$ . Also, in the case of  $\eta_1^x$ , the size of the shifts in the loadings is controlled by the parameter  $\sigma \in \sqrt{2}\{1, 0.75, 0.5, 0.25\}$ , as  $\Delta_{ij} \sim \text{iid} \mathcal{N}(0, \sigma^2)$ . Finally, we set  $\phi \in \{1, 1.5, 2, 2.5\}$ , a parameter that features in  $\vartheta$ , in order to investigate the impact of  $\text{Var}(\chi_{it})/\text{Var}(x_{it})$  on the performance of the change-point detection methodology. We fix the number of observations at  $T = 500$  and the dimensionality at  $n = 100$ .

We report the performance of our methodology over 100 realisations in Figs. 10 and 11 for the two extreme cases when  $\phi \in \{1, 2.5\}$  with  $\sigma = 0.75\sqrt{2}$ . Also, we include the results from applying the DCBS algorithm to the WT of true common and idiosyncratic components generated under (M2) as a benchmark case. Additional simulation results with varying  $\phi$  and  $\sigma$  are reported in the supplementary document.

In accordance with the observations made under single change-point scenarios (Section 7.1), detecting change-points in the common components,  $\eta_1^x$  and  $\eta_3^x$  in particular, becomes more challenging as they grow sparse cross-sectionally (with decreasing  $\varrho$ ) and as  $\text{Var}(\chi_{it})/\text{Var}(x_{it})$  decreases (with increasing  $\phi$ ). For the settings considered here, the DCBS algorithm applied to WT of  $\hat{\epsilon}_{it}^*$  performs as well as that applied to the WT of the true  $\epsilon_{it}$ , regardless of the model parameters  $\phi$  and  $\varrho$ . Not surprisingly, as the break in the loadings grows weaker with decreasing  $\sigma$ , the detection rate of  $\eta_1^x$  deteriorates, especially when  $\sigma = 0.25\sqrt{2}$ . Comparing the performance of the DCBS algorithm applied to  $\hat{\chi}_{it}^*$  and  $\chi_{it}$ , the gap is not so striking in the detection of the change-point  $\eta_2^x$  in the autocorrelations of the factors. As for  $\eta_1^x$  and  $\eta_3^x$ , provided that the breaks in the loadings and the number of factors are moderately dense ( $\varrho \geq 0.5$ ), and the magnitude of the former reasonably large ( $\sigma \geq 0.5\sqrt{2}$ ), we can expect the common components estimated via PCA to recover the both change-points as those in their second-order structure, for a range of  $\phi$ .

## 8. Real data analysis

### 8.1. S&P100 stock returns

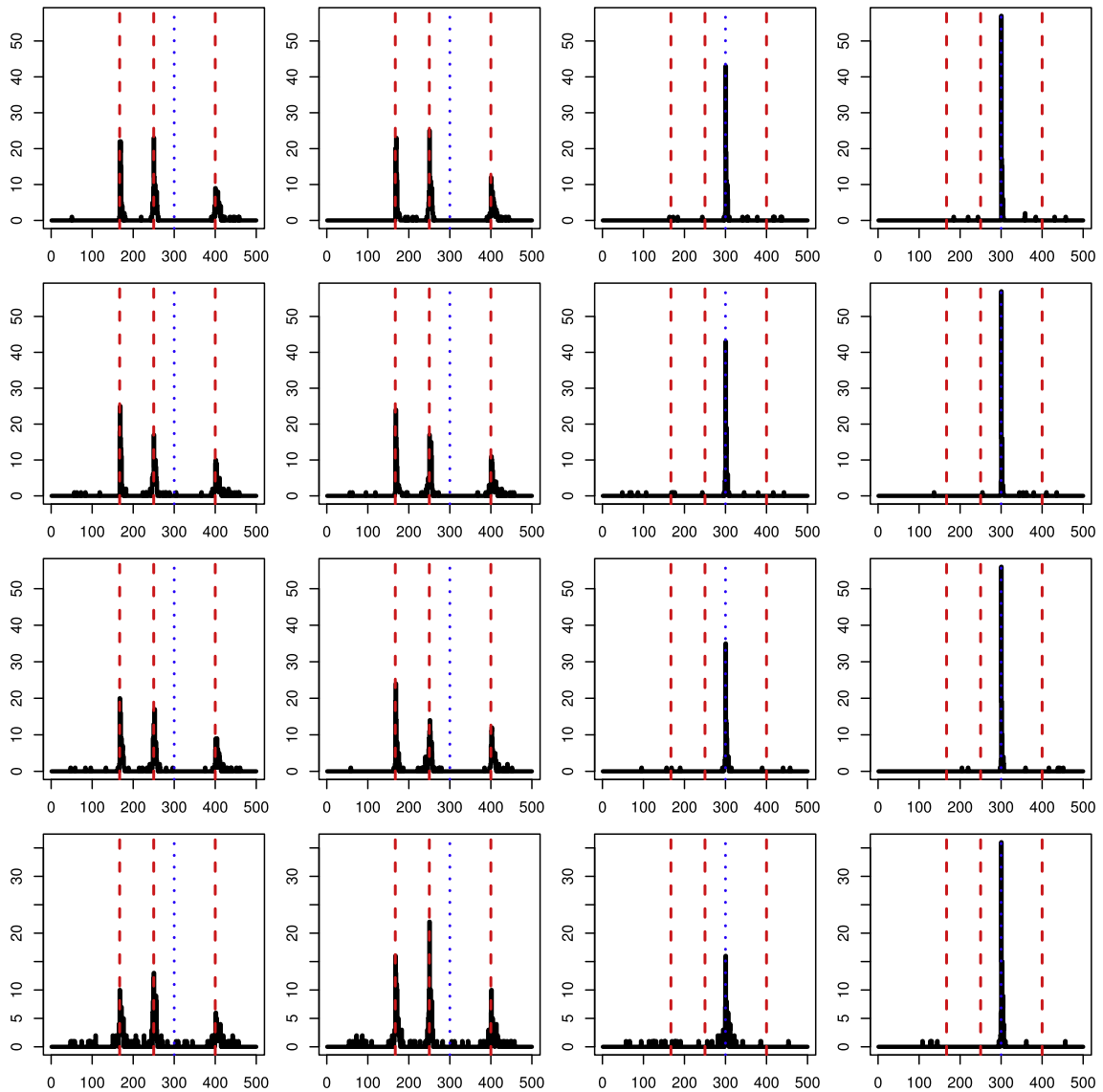
In this section, we perform change-point analysis on log returns of the daily closing values of the stocks composing the Standard and Poor's 100 (S&P100) index, observed between 4 January 2000 and 10 August 2016 ( $n = 88$  and  $T = 4177$ ). The dataset is available from Yahoo Finance. With  $\hat{\tau} = 4$  returned by the criterion in (11), the set of factor number candidates is chosen as  $\mathcal{R} = \{4, \dots, 20\}$ . Also, a constraint is imposed so that no two change-points are detected within the period of 20 working days. The maximum number of change-points for the common components is attained with  $k^* = 8$  ( $\hat{B}_x = 15$ ), and we obtain  $\hat{B}^x(k) \subset \hat{B}^x(k^*)$  for all  $k \in \mathcal{R} \setminus \{8\}$ . Table 1 reports the change-points estimated from  $\hat{\chi}_{it}^*$  and  $\hat{\epsilon}_{it}^*$ , as well as their order of detection (represented by the level index  $u$  of the nodes corresponding to the estimated change-points in the DCBS algorithm), and Fig. 12 plots two representative daily log return series from the dataset along with  $\hat{\eta}_b^x$ .

Most of the change-points we find are in a neighbourhood of the events that characterise the financial market (some of which are not exactly dated). In particular:

1. the burst of the dot-com bubble which took place between March 2000 and October 2002;
2. the start of the second Iraq war in late March 2003;
3. Lehman Brothers bankruptcy in September 2008;
4. the first and second stages of the Greek and EU sovereign debt crisis in the summers of 2011 and 2015, respectively.

By way of investigating the validity of  $\hat{\eta}_b^x$ ,  $b = 1, \dots, \hat{B}_x$ , we computed the following quantities over each segment defined by two neighbouring change-points,  $[\hat{\eta}_b^x + 1, \hat{\eta}_{b+1}^x]$ :

$$k_b(c) = \min \left\{ 1 \leq k \leq q_b : \frac{\sum_{j=1}^k \hat{\mu}_{x,j}^b}{\sum_{j=1}^{q_b} \hat{\mu}_{x,j}^b} > c \right\} \text{ for some } c \in (0, 1),$$

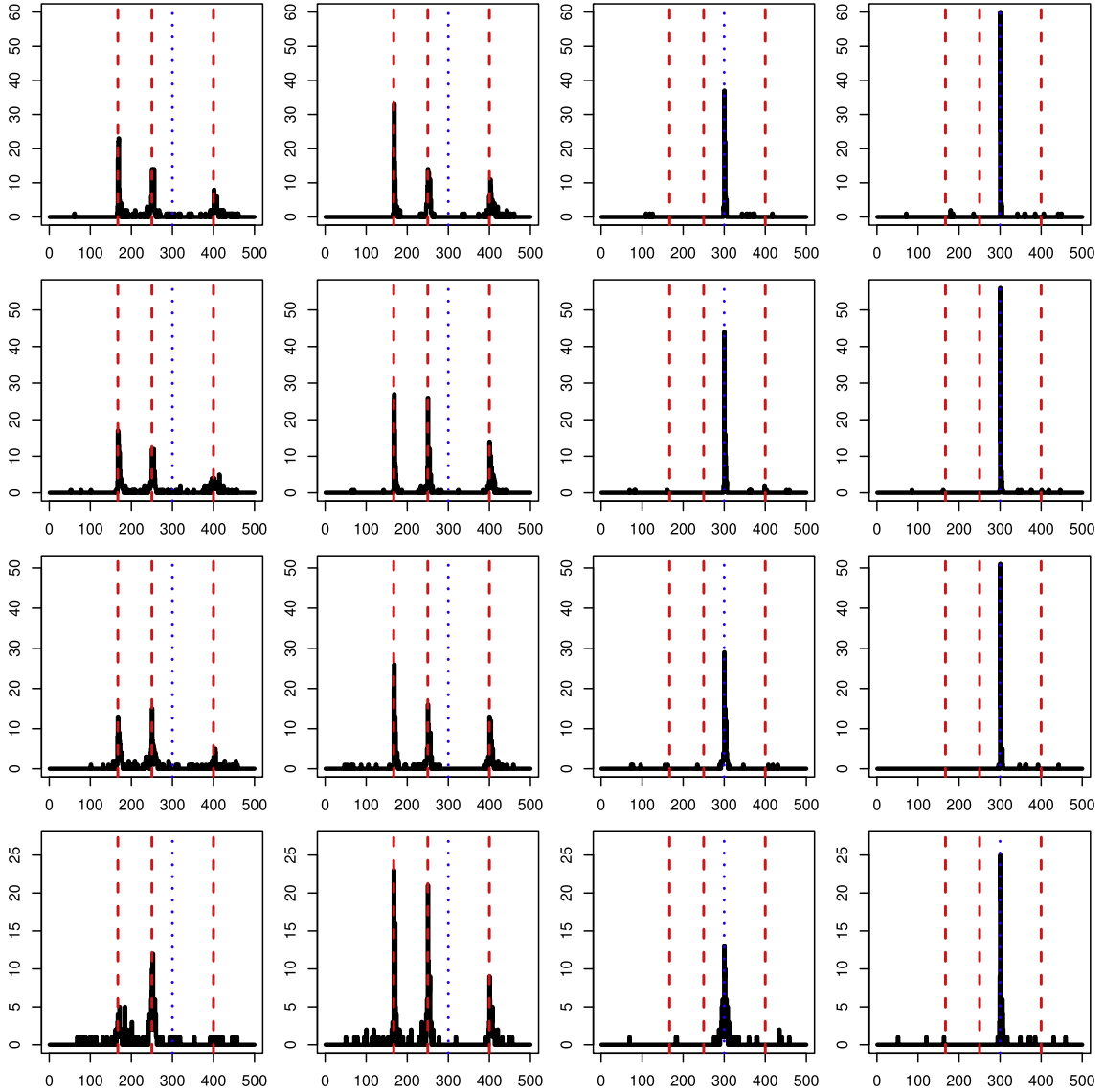


**Fig. 10.** (M2) Locations of the change-points estimated from  $\hat{\chi}_{it}^*$ ,  $\chi_{it}$  (oracle),  $\hat{\epsilon}_{it}^*$  and  $\epsilon_{it}$  (oracle) by the DCBS algorithm (left to right) for  $\rho \in \{1, 0.75, 0.5, 0.25\}$  (top to bottom) when  $n = 100$ ,  $T = 500$ ,  $\sigma = 0.75\sqrt{2}$  and  $\phi = 1$ ; vertical lines indicate the locations of the true change-points  $\eta_b^x$ ,  $b = 1, 2, 3$  (dashed) and  $\eta_b^\epsilon$  (dotted).

**Table 1**

S & P100 data: change-points estimated from the common and idiosyncratic components and their order of detection.

$\hat{\eta}_b^x$	43	89	197	331	613	656	816	1895
order	3	2	5	4	3	4	1	2
$\hat{\eta}_b^x$	2186	2249	2357	2397	2914	3020	3913	
order	4	5	6	3	5	4	5	
$\hat{\eta}_b^\epsilon$	85	181	206	268	336	631	652	735
order	3	4	5	2	4	3	4	1
$\hat{\eta}_b^\epsilon$	914	1957	2184	2210	2253	2354	2537	3911
order	4	5	3	5	6	4	2	3



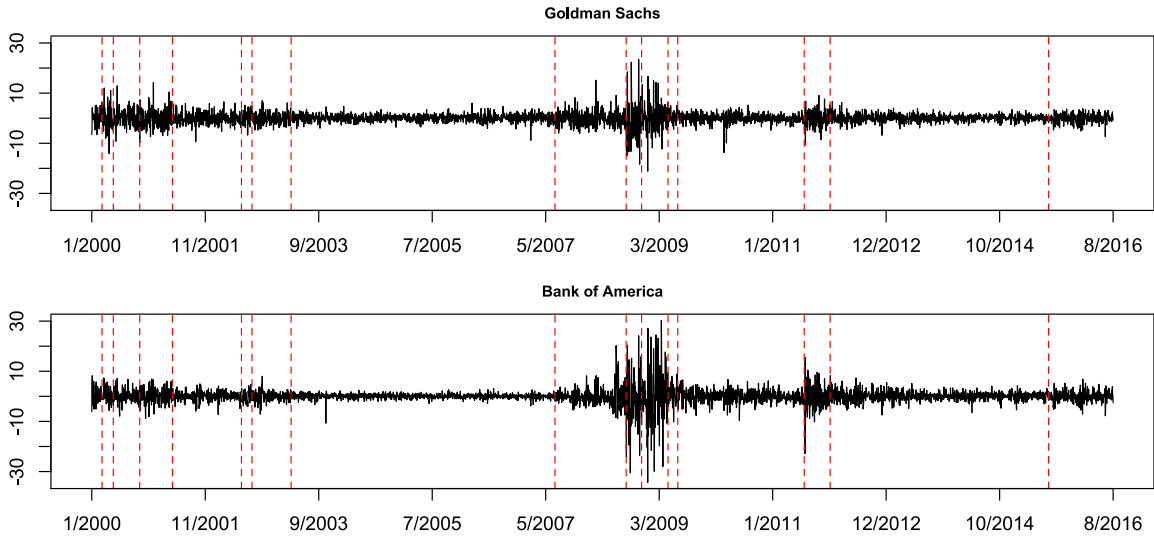
**Fig. 11.** (M2) Locations of the estimated change-points when  $n = 100$ ,  $T = 500$ ,  $\sigma = 0.75\sqrt{2}$  and  $\phi = 2.5$ .

where  $\hat{\mu}_{x,j}^b$  denotes the  $j$ th largest eigenvalue of  $\hat{\Gamma}_x^b$ , and  $q_b = (n-1) \wedge (\hat{\eta}_{b+1}^x - \hat{\eta}_b^x - 1)$ . In short,  $k_b(c)$  is the minimum number of eigenvalues required so that the proportion of the variance of  $\mathbf{x}_t$ ,  $t \in [\hat{\eta}_b^x + 1, \hat{\eta}_{b+1}^x]$  accounted for by  $\hat{\mu}_{x,j}^b$ ,  $j = 1, \dots, k_b(c)$  exceeds a given  $c$ . Varying  $c \in \{0.5, 0.55, \dots, 0.95\}$ , we plot  $k_b(c)$  over the  $\hat{B}_x + 1$  segments in Fig. 13. We observe that over long stretches of stationarity, greater numbers of eigenvalues are required to account for the same proportion of variance, compared to shorter intervals which all tend to be characterised by high volatility. This finding is in accordance with the observation made in Li et al. (2017), that a small number of factors drive the majority of the cross-sectional correlations during the periods of high volatility.

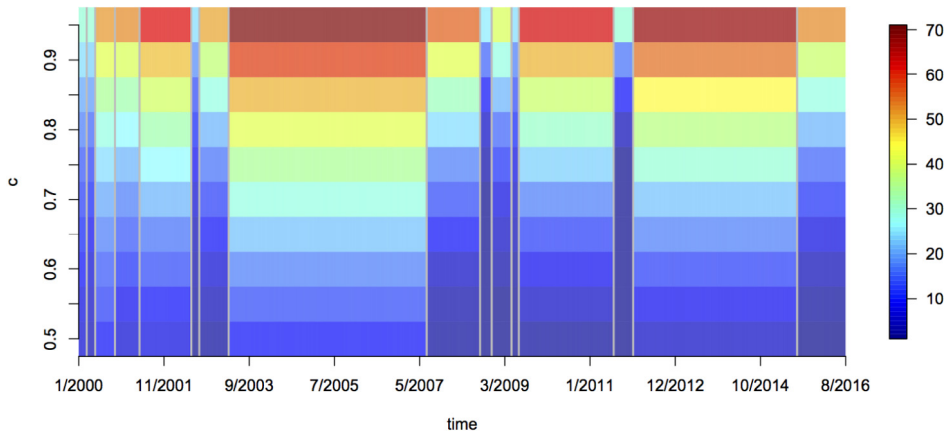
## 8.2. US macroeconomic data

We analyse the US representative macroeconomic dataset of 101 time series, collected quarterly between 1960:Q2 and 2012:Q3 ( $T = 210$ ), for change-points. Similar datasets have been analysed frequently in the factor model literature, for example, in Stock and Watson (2002). The dataset is available from the St. Louis Federal Reserve Bank website (<https://fred.stlouisfed.org/>). We impose a restriction in applying the DCBS algorithm so that no two change-points are detected within three quarters in analysing the quarterly observations. Applying the information criterion of Bai and Ng (2002),





**Fig. 12.** S&P100 data: daily log returns of Goldman Sachs (top) and Bank of America (bottom) between 4 January 2000 and 10 August 2016, along with  $\hat{\eta}_b^x$ ,  $b = 1, \dots, \hat{B}_x$  (vertical broken lines).



**Fig. 13.** S&P100 data:  $k_b(c)$  for each  $[\hat{\eta}_b^x + 1, \hat{\eta}_{b+1}^x]$ ,  $b = 1, \dots, \hat{B}_x$  according to the colour legend in the right; x-axis denotes the time and y-axis denotes  $c \in \{0.5, 0.55, \dots, 0.95\}$ .

$\hat{r} = 8$  is returned so we choose  $\mathcal{R} = \{8, \dots, 20\}$ . All  $k \geq 14$  lead to the identical change-point estimates for the common component with  $\hat{B}_x = 5$  so that we select  $k^* = 20$ . We also obtain  $\hat{B}^x(k) \subset \hat{B}^x(k^*)$  for all  $k < 14$ . Table 2 reports the change-points estimated from  $\hat{\chi}_{it}^*$  and  $\hat{\epsilon}_{it}^*$ , and we plot two representative series from the dataset, gross domestic product (GDP) growth rate and consumer price inflation (CPI), along with  $\hat{\eta}_b^x$  in Fig. 14.

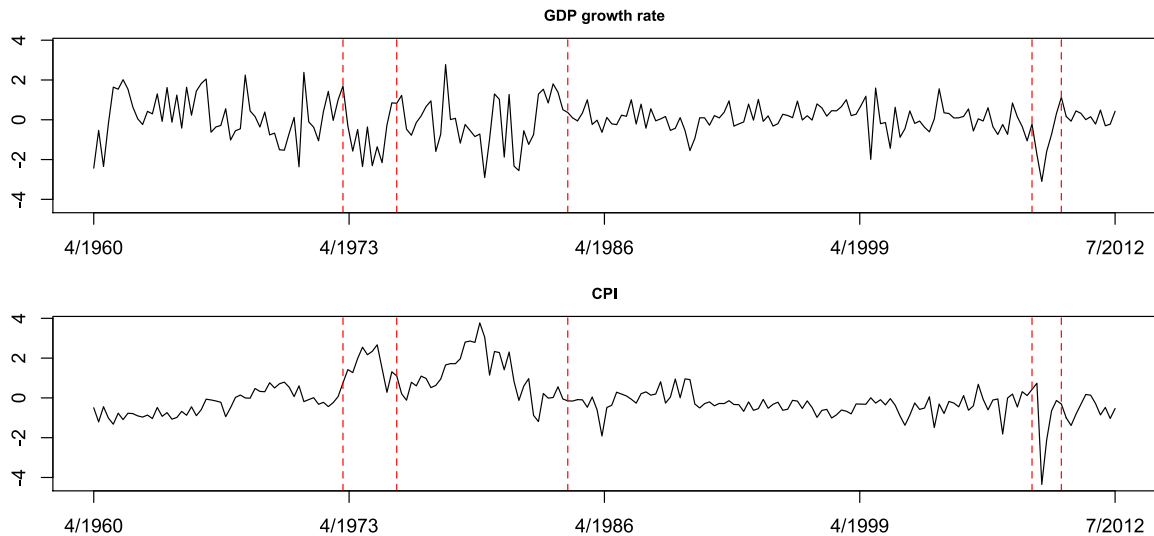
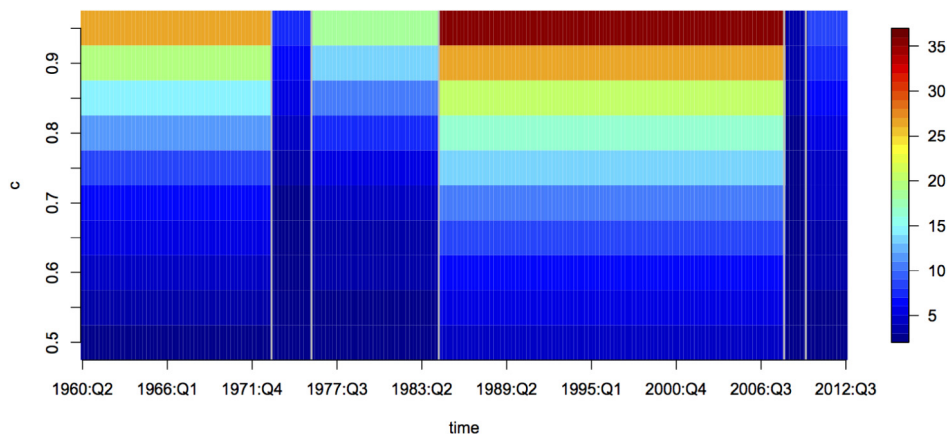
According to the change-points detected, the observations are divided into periods corresponding to different economic regimes characterised by high or low volatility. In particular, we highlight the following regimes (recessions are dated by the National Bureau of Economic Research, <http://www.nber.org/cycles.html>):

1. early 1970s to early 1980s marked by two major economic recessions, which were characterised by high inflation due to the oil crisis and the level of interest rates;
2. the so-called Great Moderation period which, according to our analysis, started in late 1983 and was characterised by low volatility of most economic indicators as a result of the implementation of new monetary policies, see also [Stock and Watson \(2003\)](#);
3. the period of the financial crisis that took place between 2007 and 2009 and corresponds to the most recent (as of 2018) economic recession, with record low levels of GDP growth and inflation and associated high volatility;
4. the post-2009 years corresponding to the slow recovery of the US economy.

**Table 2**

Macroeconomic data: change-points estimated from the common and idiosyncratic components and their order of detection.

$\hat{\eta}_b^x$	48	60	95	190	196	
	1972:Q1	1975:Q1	1983:Q4	2007:Q3	2009:Q1	
order	2	3	1	2	3	
$\hat{\eta}_b^c$	49	58	92	189	194	200
	1972:Q2	1974:Q3	1983:Q1	2007:Q2	2008:Q3	2010:Q4
order	2	3	1	2	3	4

**Fig. 14.** Macroeconomic data: GDP growth rate (top) and CPI (bottom) between 1960:Q2 and 2012:Q3, along with  $\hat{\eta}_b^x$ ,  $b = 1, \dots, \hat{B}_x$  (vertical broken lines).**Fig. 15.** Macroeconomic data:  $k_b(c)$  for each  $[\hat{\eta}_b^x + 1, \hat{\eta}_{b+1}^x]$ ,  $b = 1, \dots, \hat{B}_x$  according to the colour legend in the right; x-axis denotes the time and y-axis denotes  $c \in \{0.5, 0.55, \dots, 0.95\}$ .

Cheng et al. (2016) performed change-point analysis on a similar set of macroeconomic and financial indicators, observed monthly rather than quarterly, over a shorter span of period between January 1985 and January 2013. Their focus was on verifying the existence of a structural break corresponding to the beginning of the financial crisis, which they estimated to be in December 2007, a date which is close to our  $\hat{\eta}_4^x$  (considering that we analyse quarterly observations). As in Section 8.1, we perform a post-change-point analysis by plotting  $k_b(c)$  computed on each segment defined by  $\hat{\eta}_b^x$ , see Fig. 15, where we make similar observations about the contrast between the number of factors required over long stretches of stationarity and short intervals of volatility.

## 9. Conclusions

We have provided the first comprehensive treatment of high-dimensional time series factor models with multiple change-points in their second-order structure. We have proposed an estimation approach based on the capped PCA and wavelet transformations, first separating common and idiosyncratic components and then performing multiple change-point analysis on the levels of the transformed data. The number and locations of change-points are estimated consistently as  $n, T \rightarrow \infty$  for both the common and idiosyncratic components. Our methodology is robust to the over-specification of the number of factors which, in the presence of multiple change-points, may not be accurately estimated by standard methods. Post change-point detection, we have proved the consistency of the common components estimated via PCA on each stationary segment.

An extensive numerical study has shown the good practical performance of our method and demonstrated that factor analysis prior to change-point detection improves the detectability of change-points. Two applications involving economic data have shown that we are able to pick up most of the structural changes in the economy, such as the recent financial crisis (2008–2009), economic recessions (mid 1970s and late 2000s) or changes in the monetary policy regime (the start of the so-called Great Moderation in early 1980s). Our method is implemented in the R package `factorcpt`, available from CRAN.

## Acknowledgments

Haeran Cho's work was supported by the Engineering and Physical Sciences Research Council grant no. EP/N024435/1. Piotr Fryzlewicz's work was supported by the Engineering and Physical Sciences Research Council grant no. EP/L014246/1. We thank the Editor, Associate Editor and two referees for very helpful comments which led to a substantial improvement of this paper. All the codes used in simulation studies and real data analysis are available as the R package `factorcpt` on CRAN, and all the real datasets are available publicly online (repositories provided in the text).

## Appendix A. An example of piecewise stationary factor model

We illustrate with an example that the piecewise stationary factor model (1) under Assumption 1 admits all possible change-point scenarios that arise under factor modelling for the common components.

Let  $g_{1t}$  be an AR(1) process with white noise innovations  $u_t \sim (0, 1)$ , and let  $g_{2t}$  be a white noise process. Suppose that the common components,  $\chi_{it}$ , for all  $i = 1, \dots, n$ , are written as

$$\chi_{it} = \begin{cases} a_i g_{1t} = a_i(\alpha g_{1,t-1} + u_t) & \text{for } \eta_0^x + 1 = 1 \leq t \leq \eta_1^x, \\ a_i g_{1t} + b_i g_{2t} = a_i(\alpha g_{1,t-1} + u_t) + b_i g_{2t} & \text{for } \eta_1^x + 1 \leq t \leq \eta_2^x, \\ c_i g_{1t} + d_i g_{2t} = c_i(\alpha g_{1,t-1} + u_t) + d_i g_{2t} & \text{for } \eta_2^x + 1 \leq t \leq \eta_3^x, \\ c_i g_{1t} + d_i g_{2t} = c_i(\beta g_{1,t-1} + u_t) + d_i g_{2t} & \text{for } \eta_3^x + 1 \leq t \leq \eta_4^x = T, \end{cases} \quad (15)$$

for some  $|\alpha|, |\beta| < 1$ . In (15),  $\chi_{it}$  begins with a single factor  $g_{1t}$  for  $t \leq \eta_1^x$ , to which three change-points are introduced: (a) appearance of a new factor  $g_{2t}$  at  $t = \eta_1^x + 1$ , (b) changes in both loadings at  $t = \eta_2^x + 1$ , and (c) changes in the autocorrelation structure of  $g_{1t}$  at  $t = \eta_3^x + 1$ . We can re-write (15) into the piecewise stationary factor model in (1) with constant loadings  $\lambda_i = (a_i, b_i, c_i, d_i)^\top$ , and a factor vector  $\mathbf{f}_t$  of dimension  $r = 4$  and defined over the four segments  $[\eta_b^x + 1, \eta_{b+1}^x]$ ,  $b = 0, \dots, 3$  as

$$\mathbf{f}_t = \begin{cases} (g_{1t}, 0, 0, 0)^\top = (\alpha g_{1,t-1} + u_t, 0, 0, 0)^\top & \text{for } 1 \leq t \leq \eta_1^x, \\ (g_{1t}, g_{2t}, 0, 0)^\top = (\alpha g_{1,t-1} + u_t, g_{2t}, 0, 0)^\top & \text{for } \eta_1^x + 1 \leq t \leq \eta_2^x, \\ (0, 0, g_{1t}, g_{2t})^\top = (0, 0, \alpha g_{1,t-1} + u_t, g_{2t})^\top & \text{for } \eta_2^x + 1 \leq t \leq \eta_3^x, \\ (0, 0, g_{1t}, g_{2t})^\top = (0, 0, \beta g_{1,t-1} + u_t, g_{2t})^\top & \text{for } \eta_3^x + 1 \leq t \leq T. \end{cases} \quad (16)$$

The following four comments help in understanding the properties of model (1).

1. Note that  $\mathbf{f}_t$  in (16) meets condition Assumption 1(i) with

$$\mathbf{f}_t^0 = (g_{1t}^{(1)}, 0, 0, 0)^\top, \quad \mathbf{f}_t^1 = (g_{1t}^{(1)}, g_{2t}, 0, 0)^\top, \quad \mathbf{f}_t^2 = (0, 0, g_{1t}^{(1)}, g_{2t})^\top, \quad \mathbf{f}_t^3 = (0, 0, g_{1t}^{(2)}, g_{2t})^\top$$

where  $g_{1t}^{(1)} = \alpha g_{1,t-1} + u_t$  and  $g_{1t}^{(2)} = \beta g_{1,t-1} + u_t$  are stationary AR(1) processes. Indeed, it is clear that  $\mathbf{f}_t = \mathbf{f}_t^b$  for  $0 \leq b \leq 2$  and Assumption 1(i) trivially holds in those segments. Then, at  $\eta \equiv \eta_2^x$ , the AR parameter switches from  $\alpha$  to  $\beta$  and for  $t \geq \eta + 1$ ,

$$\begin{aligned} g_{1t} &= \beta g_{1,t-1} + u_t = \beta(\beta g_{1,t-2} + u_{t-1}) + u_t = \dots \\ &= \beta^{t-\eta-1} g_{1,\eta+1} + \beta^{t-\eta-2} u_{\eta+2} + \dots + \beta u_{t-1} + u_t, \\ g_{1t}^{(2)} &= \beta g_{1,t-1}^{(2)} + u_t = \beta^{t-\eta-1} g_{1,\eta+1}^{(2)} + \beta^{t-\eta-2} u_{\eta+2} + \dots + \beta u_{t-1} + u_t. \end{aligned}$$

Since  $|\alpha|, |\beta| < 1$ , both  $g_{1t}$  and  $g_{1t}^{(2)}$  have finite variance. Therefore, by selecting  $\rho_f = \beta^2 \in [0, 1)$ , we have

$$\mathbb{E}(g_{1t} - g_{1t}^{(2)})^2 = \beta^{2(t-\eta-1)} \mathbb{E}(g_{1,\eta+1} - g_{1,\eta+1}^{(2)})^2 \leq 2\beta^{2(t-\eta-1)} [\mathbb{E}(g_{1,\eta+1}^2) + \mathbb{E}\{(g_{1,\eta+1}^{(2)})^2\}] = O(\rho_f^{t-\eta}),$$

and [Assumption 1](#)(i) holds.

2. Recall that the number of factors in (1) satisfies  $r \geq r_b$ : in this example we have a single factor prior to  $\eta_1^x$  ( $r_0 = 1$ ) and then two factors in each of the subsequent segments ( $r_1 = r_2 = r_3 = 2$ ), whereas  $r = 4$ .
3. The representation in (1) is not unique. We can for example re-write (16) with constant loadings  $\lambda_i = (a_i, b_i, c_i - a_i, d_i - b_i)^\top$  and  $\mathbf{f}_t = (g_{1t}, g_{2t}, g_{1t}, g_{2t})^\top$  for  $\eta_1^x + 1 \leq t \leq T$ .
4. Model (1) can also be written as a piecewise stationary version of the dynamic factor model introduced in [Forni and Lippi \(2001\)](#) and [Hallin and Lippi \(2013\)](#). We can for example re-write (16) as the piecewise stationary version of a dynamic factor model with  $r = 5$  and constant dynamic loadings  $\lambda_i(L) = (a_i(1 - \alpha L)^{-1}, b_i, c_i(1 - \alpha L)^{-1}, d_i, c_i(1 - \beta L)^{-1})^\top$  ( $L$  denoting the lag operator), and factors defined as

$$\mathbf{f}_t^0 = (u_t, 0, 0, 0, 0)^\top, \quad \mathbf{f}_t^1 = (u_t, g_{2t}, 0, 0, 0)^\top, \quad \mathbf{f}_t^2 = (0, 0, u_t, g_{2t}, 0)^\top, \quad \mathbf{f}_t^3 = (0, 0, 0, g_{2t}, u_t)^\top.$$

A change in the autocorrelation of the factors can therefore be equivalently represented by a change in the dynamic loadings.

## Appendix B. Proofs

### B.1. Preliminary results

We denote by  $\varphi_i$  the  $n$ -dimensional vector with one as its  $i$ th element and zero elsewhere.

#### Lemma 1.

- (i)  $n^{-1} \|\widehat{\Gamma}_x - \Gamma_x\| = O_p\left(\sqrt{\frac{\log n}{T}} \vee \frac{1}{n}\right)$ ;
- (ii)  $n^{-1/2} \|\varphi_i^\top (\widehat{\Gamma}_x - \Gamma_x)\| = O_p\left(\sqrt{\frac{\log n}{T}} \vee \frac{1}{\sqrt{n}}\right)$ .

**Proof.** Under [Assumptions 2–5](#), [Lemmas A.3](#) and [B.1 \(ii\)](#) of [Fan et al. \(2011a\)](#) show that

$$\begin{aligned} \max_{1 \leq i, i' \leq n} \left| \frac{1}{T} \sum_{t=1}^T x_{it} x_{i't} - \mathbb{E}\left(\frac{1}{T} \sum_{t=1}^T x_{it} x_{i't}\right) \right| &\leq \max_{1 \leq j, j' \leq r} r^2 \bar{\lambda}^2 \left| \frac{1}{T} \sum_{t=1}^T f_{jt} f_{j't} - \mathbb{E}\left(\frac{1}{T} \sum_{t=1}^T f_{jt} f_{j't}\right) \right| \\ &+ \max_{1 \leq i, i' \leq n} \left| \frac{1}{T} \sum_{t=1}^T \epsilon_{it} \epsilon_{i't} - \mathbb{E}\left(\frac{1}{T} \sum_{t=1}^T \epsilon_{it} \epsilon_{i't}\right) \right| + 2 \max_{\substack{1 \leq j \leq r \\ 1 \leq i \leq n}} r \bar{\lambda} \left| \frac{1}{T} \sum_{t=1}^T f_{jt} \epsilon_{it} \right| = O_p\left(\sqrt{\frac{\log n}{T}}\right). \end{aligned} \quad (17)$$

Therefore,

$$\frac{1}{n} \|\widehat{\Gamma}_x - \Gamma_x\| \leq \frac{1}{n} \|\widehat{\Gamma}_x - \Gamma_x\|_F + \frac{1}{n} \|\Gamma_\epsilon\| = O_p\left(\sqrt{\frac{\log n}{T}} \vee \frac{1}{n}\right), \quad (18)$$

which follows from (17), [Assumption 5](#)(i) and the observation under [Assumption 4](#)(i) that  $\mu_{\epsilon,1} = \|\Gamma_\epsilon\| < \infty$  for any  $n$ , see also the result in (C2). This proves part (i).

For part (ii), now we deal with an  $n$ -dimensional vector and not a matrix. Hence, using the same approach as in part (i),

$$\frac{1}{\sqrt{n}} \|\varphi_i^\top (\widehat{\Gamma}_x - \Gamma_x)\| \leq \frac{1}{\sqrt{n}} \|\varphi_i^\top (\widehat{\Gamma}_x - \Gamma_x)\| + \frac{1}{\sqrt{n}} \|\Gamma_\epsilon\| = O_p\left(\sqrt{\frac{\log n}{T}} \vee \frac{1}{\sqrt{n}}\right),$$

which completes the proof.  $\square$

**Lemma 2.** Let  $r \times r$  diagonal matrices  $\widehat{\mathbf{M}}_x$  and  $\mathbf{M}_x$  have the  $r$  largest eigenvalues of  $\widehat{\Gamma}_x$  and of  $\Gamma_x$  in the decreasing order as the diagonal elements, respectively. Then,

$$\left\| \left(\frac{\widehat{\mathbf{M}}_x}{n}\right)^{-1} - \left(\frac{\mathbf{M}_x}{n}\right)^{-1} \right\| = O_p\left(\sqrt{\frac{\log n}{T}} \vee \frac{1}{n}\right).$$

**Proof.** As a consequence of [Lemma 1](#)(i) and Weyl's inequality,  $\widehat{\mu}_{x,j}$  satisfy

$$\frac{1}{n} |\widehat{\mu}_{x,j} - \mu_{x,j}| \leq \frac{1}{n} \|\widehat{\Gamma}_x - \Gamma_x\| = O_p\left(\sqrt{\frac{\log n}{T}} \vee \frac{1}{n}\right), \quad j = 1, \dots, r. \quad (19)$$

From (19), there exists  $\underline{c}_r \in (0, \infty)$  such that  $\mu_{\chi,r}/n \geq \underline{c}_r$  and thus  $\widehat{\mu}_{\chi,r}/n \geq \underline{c}_r + O_p(\sqrt{\frac{\log n}{T}} \vee \frac{1}{n})$ , which implies that the matrix  $n^{-1}\mathbf{M}_\chi$  is invertible and the inverse of  $n^{-1}\widehat{\mathbf{M}}_\chi$  exists with probability tending to one as  $n, T \rightarrow \infty$ . Therefore,

$$\left\| \left( \frac{\mathbf{M}_\chi}{n} \right)^{-1} \right\| = \frac{n}{\mu_{\chi,r}} = o(1), \quad \left\| \left( \frac{\widehat{\mathbf{M}}_\chi}{n} \right)^{-1} \right\| = \frac{n}{\widehat{\mu}_{\chi,r}} = O_p(1).$$

Finally,

$$\begin{aligned} \left\| \left( \frac{\widehat{\mathbf{M}}_\chi}{n} \right)^{-1} - \left( \frac{\mathbf{M}_\chi}{n} \right)^{-1} \right\| &\leq \sqrt{\sum_{j=1}^r \left( \frac{n}{\widehat{\mu}_{\chi,j}} - \frac{n}{\mu_{\chi,j}} \right)^2} \leq \sum_{j=1}^r n \left| \frac{\widehat{\mu}_{\chi,j} - \mu_{\chi,j}}{\widehat{\mu}_{\chi,j} \mu_{\chi,j}} \right| \\ &\leq \frac{r \max_{1 \leq j \leq r} |\widehat{\mu}_{\chi,j} - \mu_{\chi,j}|}{n \underline{c}_r^2 + O_p(n \sqrt{\frac{\log n}{T}} \vee 1)} = O_p \left( \sqrt{\frac{\log n}{T}} \vee \frac{1}{n} \right). \quad \square \end{aligned}$$

**Lemma 3.** Denote the  $n$ -dimensional normalised eigenvectors corresponding to the  $j$ th largest eigenvalues of  $\widehat{\Gamma}_\chi$  and  $\Gamma_\chi$ , by  $\widehat{\mathbf{w}}_{\chi,j}$  and  $\mathbf{w}_{\chi,j}$ , respectively. We further define the  $n \times r$  matrices  $\widehat{\mathbf{W}}_\chi = [\widehat{\mathbf{w}}_{\chi,1}, \dots, \widehat{\mathbf{w}}_{\chi,r}]$  and  $\mathbf{W}_\chi = [\mathbf{w}_{\chi,1}, \dots, \mathbf{w}_{\chi,r}]$ . Then, there exists an orthonormal  $r \times r$ -matrix  $\mathbf{S}$  such that

- (i)  $\|\widehat{\mathbf{W}}_\chi - \mathbf{W}_\chi \mathbf{S}\| = O_p \left( \sqrt{\frac{\log n}{T}} \vee \frac{1}{n} \right)$ ;
- (ii)  $\sqrt{n} \|\boldsymbol{\varphi}_i^\top (\widehat{\mathbf{W}}_\chi - \mathbf{W}_\chi \mathbf{S})\| = O_p \left( \sqrt{\frac{\log n}{T}} \vee \frac{1}{\sqrt{n}} \right)$ .

**Proof.** From Theorem 2 in Yu et al. (2015), which is a generalisation of the sin  $\theta$  theorem in Davis and Kahan (1970), we have

$$\|\widehat{\mathbf{W}}_\chi - \mathbf{W}_\chi \mathbf{S}\| \leq \frac{2^{3/2} \sqrt{r} \|\widehat{\Gamma}_\chi - \Gamma_\chi\|}{\min(\mu_{\chi,0} - \mu_{\chi,1}, \mu_{\chi,r} - \mu_{\chi,r+1})}, \quad (20)$$

where  $\mu_{\chi,0} = \infty$  and  $\mu_{\chi,r+1} = 0$ . From (19), the denominator of (20) is bounded from the below by  $\underline{c}_r n$  and thus part (i) follows immediately from Lemma 1(i).

For part (ii), evoking Corollary 1 of Yu et al. (2015) and noticing that  $\Gamma_\chi$  has distinct eigenvalues given in (C1), we can further show that  $\mathbf{S}$  is a diagonal matrix with entries  $\pm 1$ . Then, using part (i) above, Lemma 1 (ii) and 2, and the fact that  $\|\mathbf{W}_\chi \mathbf{S}\| = 1$  and  $\|\boldsymbol{\varphi}_i^\top \Gamma_\chi\| = O(\sqrt{n})$ , we have

$$\begin{aligned} \sqrt{n} \|\boldsymbol{\varphi}_i^\top (\widehat{\mathbf{W}}_\chi - \mathbf{W}_\chi \mathbf{S})\| &= \frac{1}{\sqrt{n}} \left\| \boldsymbol{\varphi}_i^\top \left\{ \widehat{\Gamma}_\chi \widehat{\mathbf{W}}_\chi \left( \frac{\widehat{\mathbf{M}}_\chi}{n} \right)^{-1} - \Gamma_\chi \mathbf{W}_\chi \mathbf{S} \left( \frac{\mathbf{M}_\chi}{n} \right)^{-1} \right\} \right\| \\ &\leq \frac{1}{\sqrt{n}} \|\boldsymbol{\varphi}_i^\top (\widehat{\Gamma}_\chi - \Gamma_\chi)\| \left\| \left( \frac{\mathbf{M}_\chi}{n} \right)^{-1} \right\| + \frac{1}{\sqrt{n}} \|\boldsymbol{\varphi}_i^\top \Gamma_\chi\| \left\| \left( \frac{\widehat{\mathbf{M}}_\chi}{n} \right)^{-1} - \left( \frac{\mathbf{M}_\chi}{n} \right)^{-1} \right\| \\ &+ \frac{1}{\sqrt{n}} \|\boldsymbol{\varphi}_i^\top \Gamma_\chi\| \left\| \left( \frac{\mathbf{M}_\chi}{n} \right)^{-1} \right\| \|\widehat{\mathbf{W}}_\chi - \mathbf{W}_\chi \mathbf{S}\| + o_p \left( \sqrt{\frac{\log n}{T}} \vee \frac{1}{\sqrt{n}} \right) = O_p \left( \sqrt{\frac{\log n}{T}} \vee \frac{1}{\sqrt{n}} \right). \quad \square \end{aligned}$$

**Lemma 4.** For a fixed  $\theta \geq 1 + (\beta_f^{-1} \vee 1/2)$ , we have

$$\max_{1 \leq s \leq e \leq T} \frac{1}{\sqrt{e-s+1}} \left\| \sum_{t=s}^e \mathbf{x}_t \right\| = O_p(\sqrt{n \log^\theta T}).$$

**Proof.** Note that

$$\frac{1}{\sqrt{e-s+1}} \left\| \sum_{t=s}^e \mathbf{x}_t \right\| \leq \frac{1}{\sqrt{e-s+1}} \left\| \sum_{t=s}^e \mathbf{x}_t \right\| + \frac{1}{\sqrt{e-s+1}} \left\| \sum_{t=s}^e \boldsymbol{\epsilon}_t \right\|,$$

where

$$\max_{1 \leq s \leq e \leq T} \frac{1}{\sqrt{e-s+1}} \left\| \sum_{t=s}^e \mathbf{x}_t \right\| \leq r \bar{\lambda} \sqrt{n} \max_{1 \leq j \leq r} \max_{1 \leq s \leq e \leq T} \frac{1}{\sqrt{e-s+1}} \left| \sum_{t=s}^e f_{jt} \right|,$$

and

$$\max_{1 \leq s \leq e \leq T} \frac{1}{\sqrt{e-s+1}} \left\| \sum_{t=s}^e \boldsymbol{\epsilon}_t \right\| \leq \sqrt{n} \max_{1 \leq i \leq n} \max_{1 \leq s \leq e \leq T} \frac{1}{\sqrt{e-s+1}} \left| \sum_{t=s}^e \epsilon_{it} \right|.$$



When  $\sqrt{e-s+1} \leq \log T$ , under [Assumptions 2](#) (ii), [4](#) (ii) and [7](#),

$$\max_{1 \leq j \leq r} \max_{1 \leq s \leq e \leq T} \frac{1}{\sqrt{e-s+1}} \left| \sum_{t=s}^e f_{jt} \right| \leq \log T \max_{1 \leq j \leq r} \max_{1 \leq t \leq T} |f_{jt}| = O_p(\log^{1+1/\beta_f} T),$$

$$\max_{1 \leq i \leq n} \max_{1 \leq s \leq e \leq T} \frac{1}{\sqrt{e-s+1}} \left| \sum_{t=s}^e \epsilon_{it} \right| \leq \log T \max_{1 \leq i \leq n} \max_{1 \leq t \leq T} |\epsilon_{it}| = O_p(\log^{3/2} T).$$

When  $\sqrt{e-s+1} > \log T$ , under [Assumptions 2](#) (ii), [4](#) (ii) and [5](#) (ii), the exponential inequality given in Theorem 1.4 of [Bosq \(1998\)](#) is applicable. More specifically, setting  $q = \lfloor (e-s+1)/(C_0 \log T) \rfloor$  and  $k = 3$  in the statement of that theorem, we have

$$\mathbb{P} \left( \frac{1}{\sqrt{e-s+1}} \left| \sum_{t=s}^e f_{jt} \right| > C_1 \log T \right) \leq C_2 \log T \exp \left( -\frac{C_3 C_1^2}{C_0} \log T \right) + \frac{C_4 T^{3/2}}{\log T} \exp \{ -C_5 (C_0 \log^\beta T)^{6/7} \}$$

and similarly,

$$\mathbb{P} \left( \frac{1}{\sqrt{e-s+1}} \left| \sum_{t=s}^e \epsilon_{it} \right| > C_1 \log T \right) \leq C'_2 \log T \exp \left( -\frac{C'_3 C_1^2}{C_0} \log T \right) + \frac{C'_4 T^{3/2}}{\log T} \exp \{ -C'_5 (C_0 \log^\beta T)^{6/7} \}$$

for some fixed  $C_0, C_1, C_2, \dots, C_5, C'_2, \dots, C'_5 > 0$  dependent on  $\beta_f, c_f, \beta, c_\alpha$  and  $\kappa$ . Applying Bonferroni correction,

$$\begin{aligned} & \mathbb{P} \left( \max_{1 \leq j \leq r} \max_{\substack{1 \leq s \leq e \leq T \\ \sqrt{e-s+1} > \log T}} \frac{1}{\sqrt{e-s+1}} \left| \sum_{t=s}^e f_{jt} \right| > C_1 \log T \right) \\ & \leq r T^2 [C_2 \log T \exp(-C_3 C_0^{-1} C_1^2 \log T) + C_4 T^{3/2} / (\log T) \exp \{ -C_5 (C_0 \log^\beta T)^{6/7} \}] \rightarrow 0, \\ & \mathbb{P} \left( \max_{1 \leq i \leq n} \max_{\substack{1 \leq s \leq e \leq T \\ \sqrt{e-s+1} > \log T}} \frac{1}{\sqrt{e-s+1}} \left| \sum_{t=s}^e \epsilon_{it} \right| > C_1 \log T \right) \\ & \leq n T^2 [C'_2 \log T \exp(-C'_3 C_0^{-1} C_1^2 \log T) + C'_4 T^{3/2} / (\log T) \exp \{ -C'_5 (C_0 \log^\beta T)^{6/7} \}] \rightarrow 0 \end{aligned}$$

as  $T \rightarrow \infty$  for large enough  $C_1$ , which completes the proof.  $\square$

## B.2. Proof of [Theorem 1](#)

Note that under the normalisation adopted in [Assumption 2](#)(i),  $\hat{\chi}_{it} = \varphi_i^\top \hat{\mathbf{W}}_x \hat{\mathbf{W}}_x^\top \mathbf{x}_t$  and  $\chi_{it} = \varphi_i^\top \mathbf{W}_x \mathbf{W}_x^\top \mathbf{x}_t$ . Hence,

$$\begin{aligned} \max_{1 \leq i \leq n} \max_{1 \leq t \leq T} |\hat{\chi}_{it} - \chi_{it}| & \leq \max_{1 \leq i \leq n} \max_{1 \leq t \leq T} |\varphi_i^\top \hat{\mathbf{W}}_x \hat{\mathbf{W}}_x^\top \mathbf{x}_t - \varphi_i^\top \mathbf{W}_x \mathbf{W}_x^\top \mathbf{x}_t| \\ & \quad + \max_{1 \leq i \leq n} \max_{1 \leq t \leq T} |\varphi_i^\top \mathbf{W}_x \mathbf{W}_x^\top \epsilon_t| = I + II. \end{aligned} \quad (21)$$

For  $I$ , we have

$$\begin{aligned} I & \leq \max_{1 \leq i \leq n} \max_{1 \leq t \leq T} |\varphi_i^\top \hat{\mathbf{W}}_x \hat{\mathbf{W}}_x^\top \mathbf{x}_t - \varphi_i^\top \mathbf{W}_x \hat{\mathbf{S}}_x^\top \mathbf{x}_t| + \max_{1 \leq i \leq n} \max_{1 \leq t \leq T} |\varphi_i^\top \mathbf{W}_x \hat{\mathbf{S}}_x^\top \mathbf{x}_t - \varphi_i^\top \mathbf{W}_x \mathbf{W}_x^\top \mathbf{x}_t| \\ & \leq \max_{1 \leq i \leq n} \|\varphi_i^\top (\hat{\mathbf{W}}_x - \mathbf{W}_x \mathbf{S})\| \|\hat{\mathbf{W}}_x\| \max_{1 \leq t \leq T} \|\mathbf{x}_t\| + \max_{1 \leq i \leq n} \|\varphi_i^\top \mathbf{W}_x\| \|\hat{\mathbf{W}}_x \mathbf{S} - \mathbf{W}_x\| \max_{1 \leq t \leq T} \|\mathbf{x}_t\| \\ & = O_p \left\{ \left( \sqrt{\frac{\log n}{T}} \vee \frac{1}{\sqrt{n}} \right) \log^\theta T \right\}, \end{aligned} \quad (22)$$

from [Lemma 3](#)(i)–(ii), [Lemma 4](#) and the result in (C1) lead to

$$\max_{1 \leq i \leq n} \|\varphi_i^\top \mathbf{W}_x\| \leq \max_{1 \leq i \leq n} \|\varphi_i^\top \Gamma_x\| \|\mathbf{W}_x\| \|\mathbf{M}_x^{-1}\| = O \left( \frac{1}{\sqrt{n}} \right). \quad (23)$$

As for  $II$ , due to normalisation of the eigenvectors, we invoke [Assumption 4](#)(i):

$$\mathbb{E}(\|\mathbf{W}_x^\top \epsilon_t\|^2) = \sum_{j=1}^r \mathbb{E}\{(\mathbf{w}_{x,j}^\top \epsilon_t)^2\} = \sum_{j=1}^r \sum_{i,i'=1}^n w_{x,ij} w_{x,i'j} \mathbb{E}(\epsilon_{it} \epsilon_{i't}) < r C_\epsilon. \quad (24)$$

Thereby, due to [Assumption 4](#)(ii) and Bonferroni correction,  $\max_{1 \leq t \leq T} \|\mathbf{W}_x^\top \epsilon_t\| = O_p(\sqrt{\log T})$  and using (23)

$$\max_{1 \leq i \leq n} \max_{1 \leq t \leq T} |\varphi_i^\top \mathbf{W}_x \mathbf{W}_x^\top \epsilon_t| \leq \max_{1 \leq i \leq n} \|\varphi_i^\top \mathbf{W}_x\| \max_{1 \leq t \leq T} \|\mathbf{W}_x^\top \epsilon_t\| = O_p \left( \sqrt{\frac{\log T}{n}} \right). \quad (25)$$

Substituting (22) and (25) into (21) completes the proof.  $\square$

### B.3. Proof of Theorem 2

Note that

$$\hat{\chi}_{it}^k = \sum_{j=1}^r \tilde{w}_{x,ij} \tilde{\mathbf{w}}_{x,j}^\top \mathbf{x}_t + \sum_{j=r+1}^k \tilde{w}_{x,ij} \tilde{\mathbf{w}}_{x,j}^\top \mathbf{x}_t.$$

Recall (23), from which

$$\max_{1 \leq i \leq n} \|\varphi_i^\top \hat{\mathbf{W}}_x\| \leq \max_{1 \leq i \leq n} \|\varphi_i^\top \mathbf{W}_x\| + \max_{1 \leq i \leq n} \|\varphi_i^\top (\hat{\mathbf{W}}_x - \mathbf{W}_x \mathbf{S})\| = O_p\left(\frac{1}{\sqrt{n}}\right)$$

due to Lemma 3 (ii), and therefore  $\max_{1 \leq j \leq r} \max_{1 \leq i \leq n} \hat{w}_{x,ij} = O_p(1/\sqrt{n})$ . Therefore, for  $c_w$  chosen large enough, we have  $\hat{\chi}_{it}^r = \sum_{j=1}^r \tilde{w}_{x,ij} \tilde{\mathbf{w}}_{x,j}^\top \mathbf{x}_t = \sum_{j=1}^r \hat{w}_{x,ij} \hat{\mathbf{w}}_{x,j}^\top \mathbf{x}_t = \hat{\chi}_{it}$  for all  $i$  and  $t$  with probability tending to one, and we prove this theorem conditioning on such an event; once this is done, it then implies the unconditional arguments.

Consider the scaled partial sums

$$\begin{aligned} & \max_{1 \leq i \leq n} \max_{1 \leq s \leq e \leq T} \frac{1}{\sqrt{e-s+1}} \left| \sum_{t=s}^e (\hat{\chi}_{it}^k - \chi_{it}) \right| \\ & \leq \max_{1 \leq i \leq n} \max_{1 \leq s \leq e \leq T} \frac{1}{\sqrt{e-s+1}} \left| \sum_{t=s}^e (\hat{\chi}_{it}^k - \hat{\chi}_{it}) \right| + \max_{1 \leq i \leq n} \max_{1 \leq s \leq e \leq T} \frac{1}{\sqrt{e-s+1}} \left| \sum_{t=s}^e (\hat{\chi}_{it} - \chi_{it}) \right| = I + II. \end{aligned} \quad (26)$$

Starting from II,

$$\begin{aligned} II &= \max_{1 \leq i \leq n} \max_{1 \leq s \leq e \leq T} \frac{1}{\sqrt{e-s+1}} \left| \sum_{t=s}^e \varphi_i^\top (\hat{\mathbf{W}}_x \hat{\mathbf{W}}_x^\top \mathbf{x}_t - \mathbf{W}_x \mathbf{W}_x^\top \mathbf{x}_t) \right| \\ &\leq \max_{1 \leq i \leq n} \max_{1 \leq s \leq e \leq T} \frac{1}{\sqrt{e-s+1}} \left| \sum_{t=s}^e \varphi_i^\top (\hat{\mathbf{W}}_x \hat{\mathbf{W}}_x^\top - \mathbf{W}_x \mathbf{W}_x^\top) \mathbf{x}_t \right| + \\ &\quad \max_{1 \leq i \leq n} \max_{1 \leq s \leq e \leq T} \frac{1}{\sqrt{e-s+1}} \left| \sum_{t=s}^e \varphi_i^\top \mathbf{W}_x \mathbf{W}_x^\top \epsilon_t \right| = III + IV \end{aligned}$$

where, following (22),

$$\begin{aligned} III &\leq \max_{1 \leq i \leq n} \|\varphi_i^\top (\hat{\mathbf{W}}_x - \mathbf{W}_x \mathbf{S})\| \|\hat{\mathbf{W}}_x\| \max_{1 \leq s \leq e \leq T} \frac{1}{\sqrt{e-s+1}} \left\| \sum_{t=s}^e \mathbf{x}_t \right\| \\ &\quad + \max_{1 \leq i \leq n} \|\varphi_i^\top \mathbf{W}_x\| \|\hat{\mathbf{W}}_x \mathbf{S} - \mathbf{W}_x\| \max_{1 \leq s \leq e \leq T} \frac{1}{\sqrt{e-s+1}} \left\| \sum_{t=s}^e \mathbf{x}_t \right\| = O_p \left\{ \left( \sqrt{\frac{\log n}{T}} \vee \frac{1}{\sqrt{n}} \right) \log^\theta T \right\}, \end{aligned}$$

from Lemma 3(i)–(ii) and Lemma 4. For IV, we first invoke Assumption 4(i) as in (24):

$$\max_{1 \leq s \leq e \leq T} \frac{1}{e-s+1} \mathbb{E} \left( \left\| \sum_{t=s}^e \mathbf{W}_x^\top \epsilon_t \right\|^2 \right) = \sum_{j=1}^r \max_{1 \leq s \leq e \leq T} \frac{1}{e-s+1} \sum_{t, t'=s}^e \sum_{i, i'=1}^n w_{x,ij} w_{x,i'j} \mathbb{E}(\epsilon_{it} \epsilon_{i't'}) < r C_\epsilon.$$

Hence, under Assumption 4 (ii),  $(e-s+1)^{-1/2} \sum_{t=s}^e \mathbf{W}_x^\top \epsilon_t$  is an  $r$ -vector of zero-mean normally distributed random variables with finite variance, and thus

$$\begin{aligned} & \max_{1 \leq s \leq e \leq T} \frac{1}{\sqrt{e-s+1}} \left\| \sum_{t=s}^e \mathbf{W}_x^\top \epsilon_t \right\| = O_p(\sqrt{\log T}), \quad \text{and} \\ IV &\leq \max_{1 \leq i \leq n} \|\varphi_i^\top \mathbf{W}_x\| \max_{1 \leq s \leq e \leq T} \frac{1}{\sqrt{e-s+1}} \left\| \sum_{t=s}^e \mathbf{W}_x^\top \epsilon_t \right\| = O_p \left( \sqrt{\frac{\log T}{n}} \right). \end{aligned}$$

Therefore, we have

$$\max_{1 \leq i \leq n} \max_{1 \leq s \leq e \leq T} \frac{1}{\sqrt{e-s+1}} \left| \sum_{t=s}^e (\hat{\chi}_{it} - \chi_{it}) \right| = O_p \left\{ \left( \sqrt{\frac{\log n}{T}} \vee \frac{1}{\sqrt{n}} \right) \log^\theta T \right\},$$

which proves (i). Next,

$$\begin{aligned} I &\leq \max_{1 \leq i \leq n} \sum_{j=r+1}^k |\tilde{w}_{x,ij}| \max_{1 \leq s \leq e \leq T} \frac{1}{\sqrt{e-s+1}} \left| \sum_{t=s}^e \tilde{\mathbf{w}}_{x,j}^T \mathbf{x}_t \right| \\ &= \frac{(k-r)c_w}{\sqrt{n}} \|\tilde{\mathbf{w}}_{x,j}\| \max_{1 \leq s \leq e \leq T} \frac{1}{\sqrt{e-s+1}} \left\| \sum_{t=s}^e \mathbf{x}_t \right\| = O_p(\log^\theta T), \end{aligned}$$

thanks to Lemma 4, thus proving (ii).  $\square$

#### B.4. Proof of Proposition 1

$g_j(\hat{\chi}_{it}^k)$  and  $h_j(\hat{\chi}_{it}^k, \hat{\chi}_{i't}^k)$  admit the following decompositions

$$\begin{aligned} g_j(\hat{\chi}_{it}^k) &= \mathbb{E}\{g_j(\chi_{it}^{\beta(t)})\} + [\mathbb{E}\{g_j(\chi_{it})\} - \mathbb{E}\{g_j(\chi_{it}^{\beta(t)})\}] + [g_j(\chi_{it}) - \mathbb{E}\{g_j(\chi_{it})\}] + \{g_j(\hat{\chi}_{it}^k) - g_j(\chi_{it})\} \\ &= \mathbb{E}\{g_j(\chi_{it}^{\beta(t)})\} + I + II + III, \\ h_j(\hat{\chi}_{it}^k, \hat{\chi}_{i't}^k) &= \mathbb{E}\{h_j(\chi_{it}^{\beta(t)}, \chi_{i't}^{\beta(t)})\} + [\mathbb{E}\{h_j(\chi_{it}, \chi_{i't})\} - \mathbb{E}\{h_j(\chi_{it}^{\beta(t)}, \chi_{i't}^{\beta(t)})\}] \\ &\quad + [h_j(\chi_{it}, \chi_{i't}) - \mathbb{E}\{h_j(\chi_{it}, \chi_{i't})\}] + \{h_j(\hat{\chi}_{it}^k, \hat{\chi}_{i't}^k) - h_j(\chi_{it}, \chi_{i't})\} \\ &= \mathbb{E}\{h_j(\chi_{it}^{\beta(t)}, \chi_{i't}^{\beta(t)})\} + IV + V + VI. \end{aligned}$$

By definition,  $\mathbb{E}\{g_j(\chi_{it}^{\beta(t)})\}$  and  $\mathbb{E}\{h_j(\chi_{it}^{\beta(t)}, \chi_{i't}^{\beta(t)})\}$  are piecewise constant with their change-points belonging to  $\mathcal{B}^X$ . Moreover, under Assumption 1, all change-points in  $\mathcal{I}_X^{\beta(t)}(\tau)$ ,  $|\tau| \leq \bar{\tau}_X$ , i.e., all  $\eta_b^X \in \mathcal{B}^X$ , appear as change-points in the panel  $\{\mathbb{E}\{g_j(\chi_{it}^{\beta(t)})\}, 1 \leq i \leq n, \mathbb{E}\{h_j(\chi_{it}^{\beta(t)}, \chi_{i't}^{\beta(t)})\}, 1 \leq i < i' \leq n\}$  for  $j \geq -J_T^*$ , with the choice of  $J_T^* = \lfloor \log_2 \log^v T \rfloor$  as discussed in Section 3.2.

Next, we turn our attention to scaled sums of  $I$  and  $IV$  over any given interval  $[s, e]$ , which is bounded as in the following Lemma 5 (see Appendix B.4.1 for a proof).

**Lemma 5.** Suppose that the conditions of Theorem 2 are met. At scales  $j \geq -J_T^*$ ,

$$\begin{aligned} \max_{1 \leq s < e \leq T} \frac{1}{\sqrt{e-s+1}} \left\{ \max_{1 \leq i \leq n} \left| \sum_{t=s}^e \mathbb{E}\{g_j(\chi_{it})\} - \mathbb{E}\{g_j(\chi_{it}^{\beta(t)})\} \right| \vee \right. \\ \left. \max_{1 \leq i < i' \leq n} \left| \sum_{t=s}^e \mathbb{E}\{h_j(\chi_{it}, \chi_{i't})\} - \mathbb{E}\{h_j(\chi_{it}^{\beta(t)}, \chi_{i't}^{\beta(t)})\} \right| \right\} = O(\log^v T). \end{aligned}$$

For bounding  $III$  and  $VI$ , we investigate the behaviour of  $\xi_j(\hat{\chi}_{it}^k) = g_j(\hat{\chi}_{it}^k) - g_j(\chi_{it})$  and  $\zeta_j(\hat{\chi}_{it}^k, \hat{\chi}_{i't}^k) = h_j(\hat{\chi}_{it}^k, \hat{\chi}_{i't}^k) - h_j(\chi_{it}, \chi_{i't})$ , the errors arising from replacing the unobservable  $\chi_{it}$  by its estimate  $\hat{\chi}_{it}^k$ , in Lemma 6 (see Appendix B.4.2 for a proof).

**Lemma 6.** Suppose that the conditions of Theorem 2 are met. At scales  $j \geq -J_T^*$  and  $k \geq r$ ,

$$\max_{1 \leq s < e \leq T} \frac{1}{\sqrt{e-s+1}} \left\{ \max_{1 \leq i \leq n} \left| \sum_{t=s}^e \xi_j(\hat{\chi}_{it}^k) \right| \vee \max_{1 \leq i < i' \leq n} \left| \sum_{t=s}^e \zeta_j(\hat{\chi}_{it}^k, \hat{\chi}_{i't}^k) \right| \right\} = O_p(\log^{\theta+v} T).$$

Finally, the scaled partial sums of  $II$  and  $V$  are handled by Lemma 7 (see Appendix B.4.3 for a proof).

**Lemma 7.** Suppose that the conditions of Theorem 2 are met. At scales  $j \geq -J_T^*$ ,

$$\begin{aligned} \max_{1 \leq s < e \leq T} \frac{1}{\sqrt{e-s+1}} \left\{ \max_{1 \leq i \leq n} \left| \sum_{t=s}^e g_j(\chi_{it}) - \mathbb{E}\{g_j(\chi_{it})\} \right| \right. \\ \left. \vee \max_{1 \leq i < i' \leq n} \left| \sum_{t=s}^e h_j(\chi_{it}, \chi_{i't}) - \mathbb{E}\{h_j(\chi_{it}, \chi_{i't})\} \right| \right\} = O_p(\log^{\theta+v} T). \end{aligned}$$

From Lemmas 5–7 Proposition 1 follows.  $\square$

##### B.4.1. Proof of Lemma 5

Let  $\tau = e - s + 1$  when there is no confusion. Note that

$$\begin{aligned} \frac{1}{\sqrt{\tau}} \left| \sum_{t=s}^e \mathbb{E}\{g_j(\chi_{it}) - g_j(\chi_{it}^{\beta(t)})\} \right| &\leq \sum_{h=0}^{\mathcal{L}_j-1} \frac{1}{\sqrt{\tau}} \left| \sum_{t=s}^e \mathbb{E}\{g_j(\chi_{it}^{\beta(t)}, \dots, \chi_{i,t-h+1}^{\beta(t)}, \chi_{i,t-h}, \dots, \chi_{i,t-\mathcal{L}_j+1}^{\beta(t)}) \right. \\ &\quad \left. - g_j(\chi_{it}^{\beta(t)}, \dots, \chi_{i,t-h}^{\beta(t)}, \chi_{i,t-h-1}, \dots, \chi_{i,t-\mathcal{L}_j+1}^{\beta(t)}) \right|. \end{aligned} \quad (27)$$

Restricting our attention to the first summand in the RHS of (27) (when  $h = 0$ ), by the definition of  $g_j$ ,

$$\begin{aligned} & \frac{1}{\sqrt{\tau}} \left| \mathbb{E} \left\{ \underbrace{\sum_{t=s}^e \sum_{l=0}^{\mathcal{L}_j-1} \psi_{j,l} \chi_{i,t-l}}_{w_{it}} - \underbrace{|\psi_{j,0} \chi_{it}^{\beta(t)} + \sum_{l=1}^{\mathcal{L}_j-1} \psi_{j,l} \chi_{i,t-l}|}_{v_{it}} \right\} \right| \leq \frac{1}{\sqrt{\tau}} \left| \sum_{t: w_{it} \cdot v_{it} \geq 0} \mathbb{E}(w_{it} - v_{it}) \right| \\ & + \frac{1}{\sqrt{\tau}} \left| \sum_{\substack{t: w_{it} \cdot v_{it} < 0 \\ |w_{it}| \geq |v_{it}|}} \mathbb{E}(w_{it} + v_{it}) \right| + \frac{1}{\sqrt{\tau}} \left| \sum_{\substack{t: w_{it} \cdot v_{it} < 0 \\ |w_{it}| < |v_{it}|}} \mathbb{E}(w_{it} + v_{it}) \right| = I + II + III. \end{aligned}$$

Starting from  $I$ , let  $b_1 = \beta(s)$  and  $b_2 = \beta(e) + 1$ . Then,

$$I^2 \leq \psi_{j,0}^2 \sum_{t=s}^e \mathbb{E}(\chi_{it} - \chi_{it}^{\beta(t)})^2 \leq \psi_{j,0}^2 \sum_{b=b_1}^{b_2} \sum_{t=s \vee (\eta_b+1)}^{e \wedge \eta_{b+1}} O(r^2 \bar{\lambda}^2 \rho_f^{t-\eta_b}) = O\left(\frac{\psi_{j,0}^2 B_X r^2 \bar{\lambda}^2 \rho_f}{1 - \rho_f}\right) = O(1)$$

uniformly in  $1 \leq s \leq e \leq T$  for any  $i = 1, \dots, n$ , from Assumptions 1(i) and 3. As for  $II$ , note that for all  $t$  satisfying  $w_{it} \cdot v_{it} < 0$  and  $|w_{it}| \geq |v_{it}|$ , we have  $w_{it} + v_{it} = c_t(w_{it} - v_{it})$  for some  $c_t \in [0, 1)$ , hence we can similarly show that  $II = O(1)$  and the same arguments apply to  $III$ . Then, (27) involves summation of such summands over  $h = 0, \dots, \mathcal{L}_j - 1$  and thus is bounded by  $O(\log^\nu T)$ . Similar arguments can be employed to derive a bound on the scaled partial sums of  $\mathbb{E}\{h_j(\chi_{it}, \chi_{it}^{\beta(t)}) - \mathbb{E}\{h_j(\chi_{it}^{\beta(t)}, \chi_{it}^{\beta(t)})\}$ , which concludes the proof.  $\square$

#### B.4.2. Proof of Lemma 6

The proof of Theorem 2 indicates that

$$\max_{1 \leq i \leq n} \max_{1 \leq t \leq T} \frac{1}{\sqrt{|I|}} \left| \sum_{t \in I} (\hat{\chi}_{it}^k - \chi_{it}) \right| = O_p(\log^\theta T). \quad (28)$$

Let  $\tau = e - s + 1$  when there is no confusion. Adopting the similar arguments as in Appendix B.4.1, we need to derive a bound on the following:

$$\begin{aligned} & \frac{1}{\sqrt{\tau}} \left| \sum_{t=s}^e \sum_{l=0}^{\mathcal{L}_j-1} \psi_{j,l} \hat{\chi}_{i,t-l}^k - |\psi_{j,0} \chi_{it} + \sum_{l=1}^{\mathcal{L}_j-1} \psi_{j,l} \hat{\chi}_{i,t-l}^k| \right| \leq \frac{1}{\sqrt{\tau}} \left| \sum_{t: w_{it} \cdot v_{it} \geq 0} (w_{it} - v_{it}) \right| \\ & + \frac{1}{\sqrt{\tau}} \left| \sum_{\substack{t: w_{it} \cdot v_{it} < 0 \\ |w_{it}| \geq |v_{it}|}} (w_{it} + v_{it}) \right| + \frac{1}{\sqrt{\tau}} \left| \sum_{\substack{t: w_{it} \cdot v_{it} < 0 \\ |w_{it}| < |v_{it}|}} (w_{it} + v_{it}) \right| = I + II + III. \end{aligned}$$

Since  $w_{it} - v_{it} = \psi_{j,0}(\hat{\chi}_{it}^k - \chi_{it})$ , we have  $I = O_p(\log^\theta T)$  from (28). As for  $II$ , note that for all  $t$  satisfying  $w_{it} \cdot v_{it} < 0$  and  $|w_{it}| \geq |v_{it}|$ , we have  $w_{it} + v_{it} = c_t(w_{it} - v_{it})$  for some  $c_t \in [0, 1)$ , which as for  $I$  it leads to  $II = O_p(\log^\theta T)$ . Similar arguments apply to  $III$ . Then, similarly to (27),  $\tau^{-1/2} |\sum_{t=s}^e \zeta_j(\hat{\chi}_{it}^k)|$  involves summation of  $\mathcal{L}_j$  such summands, and thus is  $O_p(\log^{\theta+\nu} T)$ . Analogous arguments can be applied to bound the scaled partial sums of  $\zeta_j(\hat{\chi}_{it}^k, \hat{\chi}_{it}^k)$ .  $\square$

#### B.4.3. Proof of Lemma 7

The proof of Lemma 4 implies that

$$\max_{1 \leq i \leq n} \max_{1 \leq t \leq T} \frac{1}{\sqrt{|I|}} \left| \sum_{t \in I} \{\chi_{it} - \mathbb{E}(\chi_{it})\} \right| = O_p(\log^\theta T)$$

(with  $\mathbb{E}(\chi_{it}) = 0$ ), from which Lemma 7 follows since  $g_j(\chi_{it})$  and  $h_j(\chi_{it}, \chi_{it}^{\beta(t)})$  involve at most  $O(2^{J_T^*}) = O(\log^\nu T)$  (lagged) terms of  $\chi_{it}$ ; for details of the arguments, see the proof of Lemma 6 in Appendix B.4.2.  $\square$

#### B.5. Proof of Theorem 3

The DC operator in this paper is identical to  $\mathcal{D}_m^\varphi(\cdot)$  defined in Cho (2016) with  $\varphi = 1/2$ . Let the additive panel data considered therein be  $y'_{\ell t} = z'_{\ell t} + \varepsilon'_{\ell t}$ . Then Assumptions 8 and 9 along with Assumption 6 imposed on the change-points in  $z_{\ell t}$ , are sufficient for the conditions imposed on  $z'_{\ell t}$ . On the other hand, their noise term satisfies  $\mathbb{E}(\varepsilon'_{\ell t}) = 0$ , while it is generally expected that  $\mathbb{E}(\varepsilon_{\ell t}) \neq 0$ . Assuming that  $\varepsilon'_{\ell t}$  is strong mixing with bounded moments, it was shown that  $(e - s + 1)^{-1/2} |\sum_{t=s}^e \varepsilon'_{\ell t}| \leq \log T$  uniformly in  $\ell \in \{1, \dots, N\}$  and  $1 \leq s < e \leq T$  (their Lemma 1), which is comparable to

the bound of  $\log^{\theta+\nu} T$  on  $(e-s+1)^{-1/2} |\sum_{t=s}^e \varepsilon_{\ell t}|$  as shown in [Propositions 1](#) and [4](#). This enables us to directly employ the arguments used in the proofs of [Theorem 3.3](#) of [Cho \(2016\)](#) for the proof of [Theorem 3](#).  $\square$

### B.6. Proof of [Proposition 2](#)

In order to prove [Proposition 2](#), we first introduce the following lemmas (see [Appendix B.6.1–B.6.3](#) for the proofs).

**Lemma 8.** Suppose that all the conditions in [Theorem 3](#) hold. Let  $\mathbf{I}_x^b = (\eta_{b+1}^x - \eta_b^x)^{-1} \mathbb{E}(\sum_{t=\eta_b^x+1}^{\eta_{b+1}^x} \mathbf{x}_t \mathbf{x}_t^\top)$ , and define  $\widehat{\mathbf{I}}_x^b$  analogously as  $\widehat{\mathbf{I}}_x^b$ . Then, for all  $b = 0, \dots, \widehat{B}_x$ ,

$$\frac{1}{n} \|\widehat{\mathbf{I}}_x^b - \mathbf{I}_x^b\|_F = O_p \left( \frac{\omega_{n,T} \log^{\theta-1/2} T}{T} \vee \sqrt{\frac{\log n}{T}} \right), \quad (29)$$

$$\frac{1}{n} \|\widehat{\mathbf{I}}_x^b - \mathbf{I}_x^b\|_F = O_p \left( \frac{\omega_{n,T} \log^{\theta-1/2} T}{T} \vee \sqrt{\frac{\log n}{T}} \right). \quad (30)$$

**Lemma 9.** Suppose that all the conditions in [Theorem 3](#) hold. Let  $\mathbf{w}_x^b = [\mathbf{w}_{x,1}^b, \dots, \mathbf{w}_{x,r_b}^b]$ , with  $\mathbf{w}_{x,j}^b$  denoting the normalised eigenvectors corresponding to  $\mu_{x,j}^b$ , the  $j$ th largest eigenvalues of  $\mathbf{I}_x^b$ . For some  $0 \leq k \leq r_b - 1$ , let  $\widehat{\mathbf{V}} = [\widehat{\mathbf{w}}_{x,k+1}^b, \dots, \widehat{\mathbf{w}}_{x,r_b}^b]$ . Then, there exists an orthonormal  $r_b \times (r_b - k)$  matrix  $\widetilde{\mathbf{S}}$  such that

$$\|\widehat{\mathbf{V}} - \mathbf{w}_x^b \widetilde{\mathbf{S}}\| = O_p \left( \sqrt{\frac{\log n}{T}} \vee \frac{1}{n} \right).$$

**Lemma 10.** Suppose that all the conditions in [Theorem 3](#) and [Assumption 10](#) hold. For a fixed  $k > r$ , let  $\widehat{\mathbf{V}} = [\widehat{\mathbf{w}}_{x,r_b+1}^b, \dots, \widehat{\mathbf{w}}_{x,k}^b]$ . Then,

$$\|\widehat{\mathbf{V}}^\top \mathbf{A}_b\| = O_p \left( \sqrt{\frac{n \log n}{T}} \vee \frac{1}{\sqrt{n}} \right).$$

From [Theorem 3](#), we have  $P(C_{n,T}) \rightarrow 1$  where

$$C_{n,T} = \left\{ \widehat{B}_x = B_x; \max_{1 \leq b \leq \widehat{B}_x} |\widehat{\eta}_b^x - \eta_b^x| < c_1 \omega_{n,T} \right\}$$

for some  $c_1 > 0$  and  $\omega_{n,T} = J_T^* \min_{1 \leq b \leq B_x} \delta_b^{-2} \log^{2\theta+2\nu} T$ . We first show that

$$P\{r_b = \arg \min_{1 \leq k \leq \widehat{r}} V_b(k)\} \geq P\{r_b = \arg \min_{1 \leq k \leq \widehat{r}} V_b(k) | C_{n,T}\} P(C_{n,T}) \rightarrow 1, \quad (31)$$

where, denoting by  $\widehat{\mathbf{w}}_{j:l}^b = [\widehat{\mathbf{w}}_{x,j}^b, \dots, \widehat{\mathbf{w}}_{x,l}^b]$  for  $1 \leq j \leq l \leq n$ ,

$$V_b(k) = \frac{1}{n(\widehat{\eta}_{b+1}^x - \widehat{\eta}_b^x)} \sum_{t=\widehat{\eta}_b^x+1}^{\widehat{\eta}_{b+1}^x} \|\widehat{\mathbf{w}}_{1:k}^b (\widehat{\mathbf{w}}_{1:k}^b)^\top \mathbf{x}_t\|^2 + k p(n, T).$$

Thanks to [Theorem 3](#), it is sufficient to show that  $P\{r_b = \arg \min_{1 \leq k \leq \widehat{r}} V_b(k) | C_{n,T}\} \rightarrow 1$  for the proof of (31). Firstly, let  $k > r_b$ . Due to the orthonormality of  $\widehat{\mathbf{w}}_{x,j}^b$ ,

$$V_b(k) - V_b(r_b) = \frac{1}{n(\widehat{\eta}_{b+1}^x - \widehat{\eta}_b^x)} \sum_{t=\widehat{\eta}_b^x+1}^{\widehat{\eta}_{b+1}^x} \|\widehat{\mathbf{w}}^\top \mathbf{x}_t\|^2 + (k - r_b) p(n, T)$$

where  $\widehat{\mathbf{V}} = \widehat{\mathbf{w}}_{(r_b+1):k}^b$ . Note that

$$\begin{aligned} \frac{1}{n(\widehat{\eta}_{b+1}^x - \widehat{\eta}_b^x)} \sum_{t=\widehat{\eta}_b^x+1}^{\widehat{\eta}_{b+1}^x} \|\widehat{\mathbf{w}}^\top \mathbf{x}_t\|^2 &\leq \frac{2}{n(\widehat{\eta}_{b+1}^x - \widehat{\eta}_b^x)} \sum_{t=\widehat{\eta}_b^x+1}^{\widehat{\eta}_{b+1}^x} \|\widehat{\mathbf{w}}^\top \mathbf{x}_t\|^2 \\ &\quad + \frac{2}{n(\widehat{\eta}_{b+1}^x - \widehat{\eta}_b^x)} \sum_{t=\widehat{\eta}_b^x+1}^{\widehat{\eta}_{b+1}^x} \|\widehat{\mathbf{w}}^\top \boldsymbol{\epsilon}_t\|^2 = I + II. \end{aligned}$$

Then, using [Lemma 10](#),

$$\begin{aligned} I &\leq \frac{2}{n(\hat{\eta}_{b+1}^x - \hat{\eta}_b^x)} \sum_{t=\hat{\eta}_b^x+1}^{\hat{\eta}_{b+1}^x} \|\hat{\mathbf{V}}\|^2 \|\hat{\mathbf{V}}^\top \mathbf{A}_b\|^2 \|\mathbf{f}_t\|^2 = \frac{2}{n} \cdot O_p\left(\frac{n \log n}{T} \vee \frac{1}{n}\right) \cdot O_p(\log^{2/\beta_f} T) \\ &= O_p\left\{\left(\frac{\log n}{T} \vee \frac{1}{n^2}\right) \log^{2/\beta_f} T\right\}, \end{aligned}$$

due to [Assumptions 2 \(ii\)](#), [5 \(ii\)](#) and [6](#). Also,

$$\begin{aligned} II &= \frac{2}{n} \text{tr} \left[ \hat{\mathbf{W}}^\top \frac{1}{|I_b^x|} \sum_{t \in I_b^x} \left\{ \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^\top - \mathbb{E}(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^\top) \right\} \right] + \frac{2}{n} \text{tr} \left\{ \hat{\mathbf{W}}^\top \frac{1}{|I_b^x|} \sum_{t \in I_b^x} \mathbb{E}(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^\top) \right\} \\ &\leq \frac{2(k - r_b)}{|I_b^x| n} \left\| \sum_{t \in I_b^x} \left\{ \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^\top - \mathbb{E}(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^\top) \right\} \right\| + \frac{2(k - r_b)}{|I_b^x| n} \left\| \sum_{t \in I_b^x} \mathbb{E}(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^\top) \right\| = O_p\left(\sqrt{\frac{\log n}{T}}\right) \end{aligned}$$

invoking [Assumptions 4–5](#) and Lemma A.3 of [Fan et al. \(2011a\)](#). Hence, under the conditions imposed on  $p(n, T)$ , we conclude that  $V_b(k) > V_b(r_b)$  for any fixed  $k > r_b$  with probability tending to one as  $n, T \rightarrow \infty$ .

Next, let  $k < r_b$ . Recalling the definition of  $\mathbf{I}_\chi^b$ , denote the  $r_b \times r_b$  diagonal matrix with  $\mu_{\chi,j}^b$ ,  $j = 1, \dots, r_b$  in its diagonal by  $\mathbf{M}_\chi^b$ . Note that

$$V_b(k) - V_b(r_b) = \frac{1}{n(\hat{\eta}_{b+1}^x - \hat{\eta}_b^x)} \sum_{t=\hat{\eta}_b^x+1}^{\hat{\eta}_{b+1}^x} \|\hat{\mathbf{W}}^\top \mathbf{x}_t\|^2 + (r_b - k)p(n, T)$$

where  $\hat{\mathbf{V}} = \hat{\mathbf{W}}_{(k+1):r_b}^b$ . Further,

$$\begin{aligned} \frac{1}{n(\hat{\eta}_{b+1}^x - \hat{\eta}_b^x)} \sum_{t=\hat{\eta}_b^x+1}^{\hat{\eta}_{b+1}^x} \|\hat{\mathbf{W}}^\top \mathbf{x}_t\|^2 &= \frac{1}{n(\hat{\eta}_{b+1}^x - \hat{\eta}_b^x)} \sum_{t=\hat{\eta}_b^x+1}^{\hat{\eta}_{b+1}^x} \|\hat{\mathbf{W}}^\top \chi_t\|^2 \\ &+ \frac{2}{n(\hat{\eta}_{b+1}^x - \hat{\eta}_b^x)} \sum_{t=\hat{\eta}_b^x+1}^{\hat{\eta}_{b+1}^x} \chi_t^\top \hat{\mathbf{W}}^\top \boldsymbol{\epsilon}_t + \frac{1}{n(\hat{\eta}_{b+1}^x - \hat{\eta}_b^x)} \sum_{t=\hat{\eta}_b^x+1}^{\hat{\eta}_{b+1}^x} \|\hat{\mathbf{W}}^\top \boldsymbol{\epsilon}_t\|^2 = III + IV + V. \end{aligned}$$

Then, we can bound  $V = O_p(\sqrt{\log n/T})$  similarly as  $II$ . Also, thanks to [Lemma 9](#), there exists an  $r_b \times (r_b - k)$  matrix  $\tilde{\mathbf{S}}$  with orthonormal columns so that

$$\|\hat{\mathbf{W}}^\top - \mathbf{W}_\chi^b \tilde{\mathbf{S}} \tilde{\mathbf{S}}^\top (\mathbf{W}_\chi^b)^\top\| \leq \|\hat{\mathbf{V}}(\hat{\mathbf{V}} - \mathbf{W}_\chi^b \tilde{\mathbf{S}})^\top\| + \|(\hat{\mathbf{V}} - \mathbf{W}_\chi^b \tilde{\mathbf{S}}) \mathbf{W}_\chi^b \tilde{\mathbf{S}}^\top\| = O_p\left(\sqrt{\frac{\log n}{T}} \vee \frac{1}{n}\right). \quad (32)$$

Note that

$$\begin{aligned} III &\geq \overbrace{\frac{1}{n(\hat{\eta}_{b+1}^x - \hat{\eta}_b^x)} \sum_{t=\hat{\eta}_b^x+1}^{\hat{\eta}_{b+1}^x} \|\mathbf{W}_\chi^b \tilde{\mathbf{S}} \tilde{\mathbf{S}}^\top (\mathbf{W}_\chi^b)^\top \chi_t\|^2}^{VI} \\ &- \overbrace{\frac{2}{n(\hat{\eta}_{b+1}^x - \hat{\eta}_b^x)} \sum_{t=\hat{\eta}_b^x+1}^{\hat{\eta}_{b+1}^x} \|\mathbf{W}_\chi^b \tilde{\mathbf{S}} \tilde{\mathbf{S}}^\top (\mathbf{W}_\chi^b)^\top \chi_t\| \|\{\hat{\mathbf{V}}^\top - \mathbf{W}_\chi^b \tilde{\mathbf{S}} \tilde{\mathbf{S}}^\top (\mathbf{W}_\chi^b)^\top\} \chi_t\|}^{VII} \\ &+ \overbrace{\frac{1}{n(\hat{\eta}_{b+1}^x - \hat{\eta}_b^x)} \sum_{t=\hat{\eta}_b^x+1}^{\hat{\eta}_{b+1}^x} \|\{\hat{\mathbf{V}}^\top - \mathbf{W}_\chi^b \tilde{\mathbf{S}} \tilde{\mathbf{S}}^\top (\mathbf{W}_\chi^b)^\top\} \chi_t\|^2}^{VIII}. \end{aligned}$$



Then,

$$\begin{aligned} VI &= \frac{1}{n(\hat{\eta}_{b+1}^x - \hat{\eta}_b^x)} \sum_{t=\hat{\eta}_b^x+1}^{\hat{\eta}_{b+1}^x} \mathbf{x}_t^\top \mathbf{W}_x^b \widetilde{\mathbf{S}} \mathbf{S}^\top (\mathbf{W}_x^b)^\top \mathbf{x}_t = \frac{1}{n} \text{tr}(\mathbf{W}_x^b \widetilde{\mathbf{S}} \mathbf{S}^\top (\mathbf{W}_x^b)^\top \hat{\Gamma}_x^b) \\ &= \frac{1}{n} \text{tr}(\mathbf{W}_x^b \widetilde{\mathbf{S}} \mathbf{S}^\top (\mathbf{W}_x^b)^\top \Gamma_x^b) + \frac{1}{n} \text{tr}\{\mathbf{W}_x^b \widetilde{\mathbf{S}} \mathbf{S}^\top (\mathbf{W}_x^b)^\top (\hat{\Gamma}_x^b - \Gamma_x^b)\} \\ &\leq \frac{1}{n} \text{tr}\{\mathbf{W}_x^b \widetilde{\mathbf{S}} \mathbf{S}^\top (\mathbf{W}_x^b)^\top \Gamma_x^b\} + \|\mathbf{W}_x^b \widetilde{\mathbf{S}} \mathbf{S}^\top (\mathbf{W}_x^b)^\top\|_F \cdot \frac{1}{n} \|\hat{\Gamma}_x^b - \Gamma_x^b\|_F \\ &= \frac{1}{n} \text{tr}(\mathbf{W}_x^b \widetilde{\mathbf{S}} \mathbf{S}^\top (\mathbf{W}_x^b)^\top \Gamma_x^b) + O_p\left(\sqrt{\frac{\log n}{T}}\right) = \frac{1}{n} \text{tr}(\widetilde{\mathbf{S}} \mathbf{S}^\top \mathbf{M}_x^b) + O_p\left(\sqrt{\frac{\log n}{T}}\right), \end{aligned}$$

which follows from Lemma 8 and that  $\widetilde{\mathbf{S}} \mathbf{S}^\top$  is a rank  $r_b - k$  projection matrix, and hence  $VI > 0$ . Also, using (32) and Assumption 2 (ii),

$$\begin{aligned} \|\widehat{\mathbf{V}} \mathbf{V}^\top - \mathbf{W}_x^b \widetilde{\mathbf{S}} \mathbf{S}^\top (\mathbf{W}_x^b)^\top\| \mathbf{x}_t &= O_p\left\{\left(\sqrt{\frac{n \log n}{T}} \vee \frac{1}{\sqrt{n}}\right) \log^{1/\beta_f} T\right\}, \quad \text{and} \\ \|\mathbf{W}_x^b \widetilde{\mathbf{S}} \mathbf{S}^\top (\mathbf{W}_x^b)^\top \mathbf{x}_t\| &= O_p(\sqrt{n} \log^{1/\beta_f} T) \end{aligned}$$

uniformly in  $t$ , and therefore  $VII = O_p\{(\sqrt{\log n/T} \vee 1/n) \log^{2/\beta_f} T\}$ . Besides,

$$VIII \leq \frac{1}{n(\hat{\eta}_{b+1}^x - \hat{\eta}_b^x)} \sum_{t=\hat{\eta}_b^x+1}^{\hat{\eta}_{b+1}^x} \|\widehat{\mathbf{V}} \mathbf{V}^\top - \mathbf{W}_x^b \widetilde{\mathbf{S}} \mathbf{S}^\top (\mathbf{W}_x^b)^\top\|^2 \|\mathbf{x}_t\|^2 = O_p\left\{\left(\frac{\log n}{T} \vee \frac{1}{n^2}\right) \log^{2/\beta_f} T\right\}.$$

Combining the bounds on VI, VII and VIII, we conclude that III is bounded away from zero with probability tending to one. Finally, under Assumptions 2 (ii), 4 (ii) and 5, Lemma B.1 (ii) of Fan et al. (2011a) leads to

$$IV = \frac{2}{n|I_b^x|} \text{tr}\left(\widehat{\mathbf{V}} \mathbf{V}^\top \sum_{t \in I_b^x} \mathbf{x}_t \epsilon_t^\top\right) \leq \frac{2(r_b - k)}{n|I_b^x|} \left\| \sum_{t \in I_b^x} \mathbf{x}_t \epsilon_t^\top \right\| = O_p\left(\sqrt{\frac{\log n}{T}}\right),$$

which leads to  $V_b(k) > V_b(r_b)$  with probability converging to one. Having shown that  $V_b(k)$  is minimised at  $r_b$ , we proved (31). Then we can adopt the arguments used in the proof of Corollary 1 in Bai and Ng (2002) verbatim to complete the proof.  $\square$

#### B.6.1. Proof of Lemma 8

For a given  $b$ , without loss of generality, assume that  $\hat{\eta}_b \leq \eta_b < \hat{\eta}_{b+1} \leq \eta_{b+1}$ . Define  $\check{\Gamma}_x^b = (\eta_{b+1}^x - \eta_b^x)^{-1} \sum_{t=\eta_b^x+1}^{\eta_{b+1}^x} \mathbf{x}_t \mathbf{x}_t^\top$  and let  $\check{\gamma}_{x,ii'}^b = [\check{\Gamma}_x^b]_{i,i'}$  and  $\hat{\gamma}_{x,ii'}^b = [\hat{\Gamma}_x^b]_{i,i'}$ . Then,

$$\|\hat{\Gamma}_x^b - \Gamma_x^b\|_F^2 = \sum_{i,i'=1}^n |\hat{\gamma}_{x,ii'}^b - \gamma_{x,ii'}^b|^2 \leq 2 \sum_{i,i'=1}^n |\hat{\gamma}_{x,ii'}^b - \check{\gamma}_{x,ii'}^b|^2 + 2 \sum_{i,i'=1}^n |\check{\gamma}_{x,ii'}^b - \gamma_{x,ii'}^b|^2 = I + II.$$

Under Assumption 6, Lemmas A.3 and B.1 (ii) of Fan et al. (2011a) can be adopted to show that  $II = O_p(n^2 \log n/T)$ . Recall the definition of  $C_{n,T}$  from the proof of Proposition 2. Then, for some  $c_2 > 0$ , consider

$$\mathbb{P}\left(\max_{1 \leq i, i' \leq n} |\hat{\gamma}_{x,ii'}^b - \check{\gamma}_{x,ii'}^b| > \frac{c_2 \omega_{n,T}}{T}\right) \leq \mathbb{P}\left(\max_{1 \leq i, i' \leq n} |\hat{\gamma}_{x,ii'}^b - \check{\gamma}_{x,ii'}^b| > \frac{c_2 \omega_{n,T}}{T} \cap C_{n,T}\right) + \mathbb{P}(C_{n,T}^c), \quad (33)$$

where the second probability in the RHS of (33) tends to zero as  $n, T \rightarrow \infty$ . Note that

$$\begin{aligned} |\hat{\gamma}_{x,ii'}^b - \check{\gamma}_{x,ii'}^b| &\leq \frac{1}{\hat{\eta}_{b+1}^x - \hat{\eta}_b^x} \sum_{t=\hat{\eta}_b^x+1}^{\hat{\eta}_{b+1}^x} |x_{it} x_{i't}| + \left| \frac{1}{\hat{\eta}_{b+1}^x - \hat{\eta}_b^x} - \frac{1}{\eta_{b+1}^x - \eta_b^x} \right| \sum_{t=\eta_b^x+1}^{\eta_{b+1}^x} |x_{it} x_{i't}| \\ &\quad + \frac{1}{\eta_{b+1}^x - \eta_b^x} \sum_{t=\hat{\eta}_{b+1}^x+1}^{\eta_{b+1}^x} |x_{it} x_{i't}| = III + IV + V. \end{aligned}$$

Lemma A.2 of Fan et al. (2011a) shows that the exponential tail bound carries over to  $x_{it} x_{i't}$  with parameter  $2\beta_f/(2 + \beta_f)$ , from which we derive that  $\max_{1 \leq i, i' \leq n} \max_{1 \leq t \leq T} |x_{it} x_{i't}| = O_p(\log^{\theta-1/2} T)$  under Assumption 7. Then, we have  $III, IV, V = O_p(\omega_{n,T} \log^{\theta-1/2} T/T)$  uniformly in  $i, i' = 1, \dots, n$  in the event of  $C_{n,T}$  under Assumption 6. Hence, the RHS of (33) tends to zero with  $n, T \rightarrow \infty$ , which leads to  $I = O_p(n^2 \omega_{n,T}^2 \log^{2\theta-1} T/T^2)$  and concludes the proof of (29). The proof of (30) follows analogously.  $\square$

### B.6.2. Proof of Lemma 9

Recalling the discussion on Assumption 9 and the definition of  $\omega_{n,T}$  in Theorem 3, we have  $\omega_{n,T} \ll T^{1/2}$  for estimating the change-points in the common components, and thus  $\sqrt{\log n/T}$  dominates the RHS of (29). Then, we apply the variation of  $\sin \theta$  theorem in Yu et al. (2015) as in Lemma 3 in combination with Lemma 8, and show that  $\|\widehat{\mathbf{V}} - \mathbf{W}_\chi^b \widetilde{\mathbf{S}}\| = O_p\left(\sqrt{\frac{\log n}{T}} \vee \frac{1}{n}\right)$ .  $\square$

### B.6.3. Proof of Lemma 10

Let  $\mathbf{W}_\chi^b = [\widehat{\mathbf{W}}_{\chi,1}^b, \dots, \widehat{\mathbf{W}}_{\chi,r_b}^b]$ . Since  $\widehat{\mathbf{V}}^\top \widehat{\mathbf{W}}_\chi^b = \mathbf{O}_{(k-r_b) \times r_b}$  and  $\|\widehat{\mathbf{V}}\| = 1$ , we have

$$\|\widehat{\mathbf{V}}^\top \mathbf{W}_\chi^b\| = \|\widehat{\mathbf{V}}^\top \mathbf{W}_\chi^b \widetilde{\mathbf{S}}\| = \|\widehat{\mathbf{V}}^\top (\widehat{\mathbf{W}}_\chi^b - \mathbf{W}_\chi^b \widetilde{\mathbf{S}})\| \leq \|\widehat{\mathbf{W}}_\chi^b - \mathbf{W}_\chi^b \widetilde{\mathbf{S}}\| = O_p\left(\sqrt{\frac{\log n}{T}} \vee \frac{1}{n}\right) \quad (34)$$

from Lemma 9. Note that  $\mathbf{I}_\chi^b = \mathbf{W}_\chi^b \mathbf{M}_\chi^b \mathbf{W}_\chi^{b\top} = \mathbf{A}_b \mathbf{I}_f^b \mathbf{A}_b^\top$ . Then,

$$\begin{aligned} \|\widehat{\mathbf{V}}^\top \mathbf{A}_b\| &= \|\widehat{\mathbf{V}}^\top \mathbf{W}_\chi^b \mathbf{M}_\chi^b \mathbf{W}_\chi^{b\top} \mathbf{A}_b (\mathbf{A}_b^\top \mathbf{A}_b)^{-1} (\mathbf{I}_f^b)^{-1}\| \leq \|\widehat{\mathbf{V}}^\top \mathbf{W}_\chi^b\| \|\mathbf{M}_\chi^b\| \|\mathbf{A}_b (\mathbf{A}_b^\top \mathbf{A}_b)^{-1}\| \|(\mathbf{I}_f^b)^{-1}\| \\ &= O_p\left\{\left(\sqrt{\frac{\log n}{T}} \vee \frac{1}{n}\right) \cdot n \cdot \frac{1}{\sqrt{n}}\right\} = O_p\left(\sqrt{\frac{n \log n}{T}} \vee \frac{1}{\sqrt{n}}\right). \end{aligned}$$

using (34) and Assumption 10.  $\square$

### B.7. Proof of Proposition 3

We can show the consistency of the PCA-based estimator of the common components within each segment (in the sense of Theorem 1), by establishing consistency of the  $r_b$  leading eigenvectors of  $\widehat{\mathbf{I}}_\chi^b$  in estimating the leading eigenvectors of  $\mathbf{I}_\chi^b$  up to a rotation. See the proof of Lemma 1(i).

Lemma 8 shows the element-wise consistency of  $\widehat{\mathbf{I}}_\chi^b$  over each  $I_b^x$  defined by the estimated change-points  $\widehat{\eta}_b^x$ ,  $b = 1, \dots, \widehat{B}_\chi$ . Recalling that  $\omega_{n,T} \ll T^{1/2}$  for estimating the change-points in the common components, we have  $\sqrt{\log n/T}$  dominate the RHS of (29). Therefore,

$$\frac{1}{n} \|\widehat{\mathbf{I}}_\chi^b - \mathbf{I}_\chi^b\| \leq \frac{1}{n} \|\widehat{\mathbf{I}}_\chi^b - \mathbf{I}_\chi^b\|_F + \frac{1}{n} \|\mathbf{I}_\epsilon^b\| = O_p\left(\sqrt{\frac{\log n}{T}} \vee \frac{1}{n}\right),$$

as  $\mu_{\epsilon,1}^b < C_\epsilon$  from Assumption 4(i). Given this result, the arguments adopted in the proof of Theorem 1 are applicable verbatim for the proof of Proposition 3 and details are therefore omitted.  $\square$

## Appendix C. WT for change-point analysis of the idiosyncratic components

The WT proposed in Section 3.3 is directly applicable to  $\widehat{\epsilon}_{it}^k$  for change-point analysis in the idiosyncratic components. For completeness, we state the corresponding arguments below.

Recall the notation  $\gamma(t) = \max\{0 \leq b \leq B_\epsilon : \eta_b^\epsilon + 1 \leq t\}$ . Then,  $g_j(\widehat{\epsilon}_{it}^k)$  and  $h_j(\widehat{\epsilon}_{it}^k, \widehat{\epsilon}_{i't}^k)$  admit the following decompositions

$$g_j(\widehat{\epsilon}_{it}^k) = \mathbb{E}\{g_j(\epsilon_{it}^{\gamma(t)})\} + [\mathbb{E}\{g_j(\epsilon_{it})\} - \mathbb{E}\{g_j(\epsilon_{it}^{\gamma(t)})\}] + [g_j(\epsilon_{it}) - \mathbb{E}\{g_j(\epsilon_{it})\}] + \xi_j(\widehat{\epsilon}_{it}^k), \quad (35)$$

$$\begin{aligned} h_j(\widehat{\epsilon}_{it}^k, \widehat{\epsilon}_{i't}^k) &= \mathbb{E}\{h_j(\epsilon_{it}^{\gamma(t)}, \epsilon_{i't}^{\gamma(t)})\} + [\mathbb{E}\{h_j(\epsilon_{it}, \epsilon_{i't})\} - \mathbb{E}\{h_j(\epsilon_{it}^{\gamma(t)}, \epsilon_{i't}^{\gamma(t)})\}] \\ &\quad + [h_j(\epsilon_{it}, \epsilon_{i't}) - \mathbb{E}\{h_j(\epsilon_{it}, \epsilon_{i't})\}] + \zeta_j(\widehat{\epsilon}_{it}^k, \widehat{\epsilon}_{i't}^k). \end{aligned} \quad (36)$$

Under Assumption 1 (iii)–(iv), all change-points in  $\mathbf{I}_\epsilon^{\gamma(t)}(\tau)$  at lags  $|\tau| \leq \bar{\tau}_\epsilon$ , namely all  $\eta_b^\epsilon \in \mathcal{B}^\epsilon$ , appear as change-points in the piecewise constant signals  $\{\mathbb{E}\{g_j(\epsilon_{it}^{\gamma(t)})\}, 1 \leq i \leq n, \mathbb{E}\{h_j(\epsilon_{it}^{\gamma(t)}, \epsilon_{i't}^{\gamma(t)})\}, 1 \leq i < i' \leq n\}$  at scales  $j \geq -J_T^*$ . Analogous to Lemmas 5–7, we establish the bound on the scaled partial terms in (35)–(36).

**Lemma 11.** Suppose that the conditions of Theorem 2 are met. At  $j \geq -J_T^*$  and  $k \geq r$ ,

(i)

$$\begin{aligned} \max_{1 \leq s \leq e \leq T} \frac{1}{\sqrt{e-s+1}} \left\{ \max_{1 \leq i \leq n} \left| \sum_{t=s}^e \mathbb{E}\{g_j(\epsilon_{it})\} - \mathbb{E}\{g_j(\epsilon_{it}^{\gamma(t)})\} \right| \vee \right. \\ \left. \max_{1 \leq i < i' \leq n} \left| \sum_{t=s}^e \mathbb{E}\{h_j(\epsilon_{it}, \epsilon_{i't})\} - \mathbb{E}\{h_j(\epsilon_{it}^{\gamma(t)}, \epsilon_{i't}^{\gamma(t)})\} \right| \right\} = O(\log^v T). \end{aligned}$$

(ii)

$$\max_{1 \leq s < e \leq T} \frac{1}{\sqrt{e-s+1}} \left\{ \max_{1 \leq i \leq n} \left| \sum_{t=s}^e \xi_j(\widehat{\epsilon}_{it}^k) \right| \vee \max_{1 \leq i < i' \leq n} \left| \sum_{t=s}^e \zeta_j(\widehat{\epsilon}_{it}^k, \widehat{\epsilon}_{i't}^k) \right| \right\} = O(\log^{\theta+\nu} T).$$

(iii)

$$\begin{aligned} & \max_{1 \leq s < e \leq T} \frac{1}{\sqrt{e-s+1}} \left[ \max_{1 \leq i \leq n} \left| \sum_{t=s}^e g_j(\epsilon_{it}) - \mathbb{E}\{g_j(\epsilon_{it})\} \right| \right. \\ & \quad \left. \vee \max_{1 \leq i < i' \leq n} \left| \sum_{t=s}^e h_j(\epsilon_{it}, \epsilon_{i't}) - \mathbb{E}\{h_j(\epsilon_{it}, \epsilon_{i't})\} \right| \right] = O_p(\log^{\theta+\nu} T). \end{aligned}$$

The proof of Lemma 11 take the analogous steps as the proofs of Lemmas 5–7 and thus is omitted. Then Proposition 4 holds for the WT of  $\widehat{\epsilon}_{it}^k$ . Again, its proof takes the identical arguments as the proof of Proposition 1 thus is omitted.

**Proposition 4.** Suppose that all the conditions in Theorem 2 hold. For some fixed  $k \geq r$  and  $J_T^* = \lfloor C \log_2 \log^{\nu} T \rfloor$ , consider the  $N = J_T^* n(n+1)/2$ -dimensional panel in (8), and denote as  $y_{\ell t}$  a generic element of the panel. Then, we have the following decomposition:

$$y_{\ell t} = z_{\ell t} + \varepsilon_{\ell t}, \quad \ell = 1, \dots, N, \quad t = 1, \dots, T.$$

- (i)  $z_{\ell t}$  are piecewise constant as the corresponding elements of (9). That is, all change-points in  $z_{\ell t}$  belong to  $\mathcal{B}^{\epsilon} = \{\eta_1^{\epsilon}, \dots, \eta_{B_{\epsilon}}^{\epsilon}\}$  and for each  $b \in \{1, \dots, B_{\epsilon}\}$ , there exists at least a single index  $\ell \in \{1, \dots, N\}$  for which  $|z_{\ell, \eta_b^{\epsilon}+1} - z_{\ell, \eta_b^{\epsilon}}| \neq 0$ .
- (ii)  $\max_{1 \leq \ell \leq N} \max_{1 \leq s < e \leq T} (e-s+1)^{-1/2} |\sum_{t=s}^e \varepsilon_{\ell t}| = O_p(\log^{\theta+\nu} T)$ .

## Appendix D. Implementation of the proposed methodology

### D.1. Generation of the additive panel data from $\widehat{\chi}_{it}^k$

In the implementation of our proposed methodology, we use the input panel data of reduced dimension  $\widetilde{N} = J_T^* n$  instead of  $N = J_T^* n(n+1)/2$ :

$$\{g_j(\widehat{\chi}_{it}^k), 1 \leq i \leq n; -J_T^* \leq j \leq -1; 1 \leq t \leq T\}. \quad (37)$$

The motivation behind such a modification is two-fold. Firstly, it substantially reduces the computational cost by speeding up not only the computation of the  $\widetilde{N} \times (e-s)$ -dimensional array of DC statistics over the segment  $[s, e]$ , but also the resampling procedure adopted in the bootstrap algorithm for threshold selection.

Besides, suppose that a change-point, say  $\eta_b^{\chi}$ , appears as jumps only in  $h_j(\chi_{it}^{\beta(t)}, \chi_{i't}^{\beta(t)})$  at the population level, i.e.,

$$\lambda_i^{\top} \left[ \sum_{|\tau| < \mathcal{L}_j} \left\{ \widetilde{f}_f^b(\tau) - \widetilde{f}_f^{b-1}(\tau) \right\} \psi_j(\tau) \right] \lambda_i = 0 \quad \text{for all } i = 1, \dots, n \text{ and } j \geq -J_T^*,$$

where  $\widetilde{f}_f^b(\tau) = \mathbb{E}\{\mathbf{f}_{t+\tau}^b(\mathbf{f}_t^b)^{\top}\}$ . Such a constraint may be viewed as an over-determined system consisting of  $J_T^* n$  equations and  $\mathcal{L}_{-J_T^*} r(r+1)/2$  variables. With  $n \rightarrow \infty$  and  $r$  fixed, there exists only a trivial solution, namely  $\widetilde{f}_f^b(\tau) - \widetilde{f}_f^{b-1}(\tau) = \mathbf{0}$  for all  $|\tau| < \mathcal{L}_{-J_T^*}$ . Therefore, it is reasonable to expect that every change-point in the second-order structure of the common components is detectable from (37).

### D.2. Choice of the number of wavelet scales

In the paper, we have investigated the DCBS algorithm applied to the panel that consists of wavelet-transformed  $\widehat{\chi}_{it}^k$  (or  $\widehat{\epsilon}_{it}^k$ ) at multiple scales simultaneously. However, it is possible to apply the algorithm sequentially:

- (a) first apply the DCBS algorithm to the panel consisting of wavelet-transformed  $\widehat{\chi}_{it}^k$  ( $\widehat{\epsilon}_{it}^k$ ) at scale  $-1$  only, and denote the set of estimated change-points by  $\widehat{\mathcal{B}}$ ;
- (b) apply the DCBS algorithm to wavelet-transformed  $\widehat{\chi}_{it}^k$  ( $\widehat{\epsilon}_{it}^k$ ) at scale  $-2$  over each segment defined by two adjacent change-points in  $\widehat{\mathcal{B}}$ , and add the estimated change-points to  $\widehat{\mathcal{B}}$ ;
- (c) repeat (b) with wavelet scales  $j = -3, -4, \dots$  until  $\widehat{\mathcal{B}}$  does not change from one scale to another.

The above procedure is motivated by the fact that the finer wavelet scales are preferable for the purpose of change-point estimation, and allows for the data-driven choice of  $J_T^*$ .

### D.3. Binary segmentation algorithm

#### D.3.1. Trimming off of the intervals for change-point analysis

In practice, we introduce an additional parameter  $d_T$  in order to ensure that the interval of interest  $[s, e]$  is of sufficient length ( $e - s + 1 > 4d_T$ ); and account for possible bias in the previously detected change-points, by trimming off the short intervals of length  $d_T$  around the previously identified change-points in Steps 1.2–1.3 of the DCBS algorithm:

$$\mathcal{T}_{s,e} = \max_{b \in [s+d_T, e-d_T]} \max_{1 \leq m \leq N} \mathcal{D}_{s,b,e}(m) \quad \text{and} \quad \hat{\eta} = \arg \max_{b \in [s+d_T, e-d_T]} \max_{1 \leq m \leq N} \mathcal{D}_{s,b,e}(m).$$

Due to the condition on the spread of change-points in [Assumption 6](#), this adjustment does not affect the theoretical consistency in the detected change-points, while ensuring that the bias in estimated change-points do not hinder the subsequent search for change-points empirically. With regards to the discussion in [Remark 2](#) and the bias presented in [Theorem 3](#), we propose to use  $d_T = \lfloor \log^2 T \wedge 0.25T^{6/7} \rfloor$ .

Though our methodology requires the selection of a multitude of parameters as listed in this section, we observe that its empirical performance is relatively robust to their choices. The parameter that exerts the most influence is  $d_T$ : smaller values of  $d_T$  tend to return a larger set of estimated change-points, some of which may be spurious as a result of bias associated with previously detected change-points. On the other hand, if  $d_T$  is set to be too large, some change-points may be left undetected due to the restrictions imposed by its choice. The default choice of  $d_T$  recommended above worked well for the simulated datasets. Also, with some prior knowledge on the dataset,  $d_T$  may be selected accordingly, as we have done in real data analysis.

#### D.3.2. Implementation of the DCBS algorithm

Steps 1–2 of the bootstrap algorithm proposed in [Section 6.1](#) generate panel data  $y_{\ell t}^*$  of the same dimensions as the original  $\{y_{\ell t}, \ell = 1, \dots, N; t = 1, \dots, T\}$ . Therefore,  $\mathcal{T}_{s,e}^*$  over different intervals  $[s_{u,v}, e_{u,v}]$  (with  $s = s_{u,v}$  and  $e = e_{u,v}$ ), which are examined at certain iterations of the DCBS algorithm, can be computed from the same bootstrap panel data  $y_{\ell t}^*$ . Since storing  $R$  copies of such high-dimensional panel data ( $N \times T$ ) is costly, we propose to apply the DCBS algorithm with the bootstrap algorithm as proposed below.

Firstly, a binary tree of a given height, say  $\hat{h}$ , is grown from  $y_{\ell t}$  with  $(s_{u,v}, e_{u,v})$  as its nodes, by omitting the testing procedure of Step 1.3 when applying the DCBS algorithm until  $u = \hat{h}$ . We impose a minimum length constraint that requires  $e_{u,v} - s_{u,v} + 1 \geq 4d_T$ ; if not, the interval  $[s_{u,v}, e_{u,v}]$  is not split further.

Let  $\mathcal{I} = \{(s_{u,v}, e_{u,v}) : 1 \leq v \leq I_u; 1 \leq u \leq \hat{h}\}$  be the collection of the nodes in the thus-generated binary tree, with  $I_u$  denoting the number of nodes at level  $u$ . Then, the following is repeated for  $m = 1, \dots, R$ : we generate the bootstrap panel data  $y_{\ell t}^{m*}$  using the bootstrap algorithm, and compute  $\mathcal{T}_{s,e}^{m*}$  for all  $(s, e) \in \mathcal{I}$ , from which the corresponding thresholds are drawn. Once it is complete, we perform Step 1.3 by setting  $(s, e) = (s_{u,v}, e_{u,v}) \in \mathcal{I}$ ; starting from  $(u, v) = (1, 1)$ , we progress in the order  $(u, v) = (1, 1), (2, 1), \dots, (2, I_2), (3, 1), \dots$ . If  $\mathcal{T}_{s,e} \leq \pi_{N,T}$  for some  $(s, e)$ , all the nodes  $(s', e')$  that are dependent on  $(s, e)$  (in the sense that  $[s', e'] \subset [s, e]$ ) are removed from  $\mathcal{I}$ , which ensures that we are indeed performing a binary segmentation procedure.  $\hat{h} = \lfloor \log_2 T/2 \rfloor$  is used in simulation studies and real data analysis, which grows a sufficiently large tree considering that we can grow a binary tree of height at most  $\lfloor \log_2 T \rfloor$ .

### D.4. Stationary bootstrap and choice of parameters

Stationary bootstrap (SB) generates bootstrap samples  $X_1^*, \dots, X_T^*$  as below. Let  $J_1, J_2, \dots$  be i.i.d. geometric random variables independent from the data, where  $P(J_1 = L) = p(1-p)^{L-1}$  with some  $p \in (0, 1)$ ,  $I_1, \dots, I_Q$  be i.i.d. uniform variables on  $\{1, \dots, T\}$ , and denote a data block by  $B(t, L) = (X_t, \dots, X_{t+L-1})$  for  $t, L \geq 1$  with a periodic extension ( $X_t = X_{t-uT}$  for some  $u \in \mathbb{Z}$  such that  $1 \leq t - uT \leq T$ ). The SB sample  $X_1^*, \dots, X_T^*$  is generated as the first  $T$  observations in the sequence  $B(I_1, J_1), \dots, B(I_Q, J_Q)$  for  $Q = \min\{q : \sum_{l=1}^q J_l \geq T\}$ .

We used the bootstrap sample size  $R = 200$  for simulation studies, and  $R = 500$  for real data analysis. Naturally, the larger bootstrap sample size is expected to return the better choice for the threshold, but we did not observe any sensitivity due to the choice of the size of bootstrap samples in our simulation studies. The binary segmentation procedure implicitly performs multiple testing, with the number of tests determined by the choice of  $\hat{h}$  described in [Appendix D.3.2](#). However, the test statistics computed at different iterations of the DCBS algorithm are correlated. Therefore, with the added difficulty arising from the hierarchy inherent in the procedure, the task of controlling for multiple testing is highly challenging. Instead, noticing that the bootstrap test statistics tend to have heavy tails, we chose to adopt the same  $\alpha = 0.05$  at all iterations of the DCBS algorithm.

In the SB,  $p$  stands for the inverse of average block length. Typically, large  $p$  leads to large bias and small variance and vice versa. For approximating the finite sample distribution of the sample mean of a univariate series  $\{X_t\}_{t=1}^T$ , [Politis and White \(2004\)](#) (accompanying corrections in [Patton et al. \(2009\)](#)) proposed to select  $p$  as  $\bar{p}^{-1} = (\hat{G}^2 / \hat{g}^2(0))^{1/3} T^{1/3}$ , where  $\hat{G} = \sum_{k=-\Lambda}^{\Lambda} \lambda(k/\Lambda) |k| \hat{R}(k)$ ,  $\hat{g}(0) = \sum_{k=-\Lambda}^{\Lambda} \lambda(k/\Lambda) \hat{R}(k)$ ,  $\hat{R}(k) = T^{-1} \sum_{t=1}^{T-|k|} (X_t - \bar{X})(X_{t+|k|} - \bar{X})$  and  $\lambda(z)$  is a trapezoidal shape symmetric taper around zero:

$$\lambda(z) = \begin{cases} 1 & \text{for } |z| \in [0, 1/2], \\ 2(1 - |z|) & \text{for } |z| \in [1/2, 1], \\ 0 & \text{otherwise.} \end{cases}$$

For the stationarity testing in multivariate (finite-dimensional) time series, Jentsch and Politis (2015) proposed to modify the above  $p$  as per the theory developed for the consistency of their stationary bootstrap, by (a) replacing  $T^{1/3}$  with  $T^{1/5}$ , and (b) taking the average of  $p^{-1}$  associated with each univariate series. Taking their approach, we choose to select  $p_j$  for individual  $\hat{f}_{jt}$ ,  $j = 1, \dots, k$  according to (a) in Step 1 for the common components, and select  $p_i$  for individual  $\hat{\epsilon}_{it}^k$ ,  $i = 1, \dots, n$  according to (a) and use  $(n^{-1} \sum_{i=1}^n p_i^{-1})^{-1}$  according to (b) in the same step for the idiosyncratic components. As for  $\Lambda$ , we plug in the automatically chosen bandwidth based on the autocorrelation structure of  $\hat{f}_{jt}$  and  $\hat{\epsilon}_{it}^k$  as suggested in Politis and White (2004).

## Appendix E. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.jeconom.2018.05.003>.

## References

- Ahn, S.C., Horenstein, A.R., 2013. Eigenvalue ratio test for the number of factors. *Econometrica* 81, 1203–1227.
- Aït-Sahalia, Y., Xiu, D., 2017. Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *J. Econometrics* 201, 384–399.
- Alessi, L., Barigozzi, M., Capasso, M., 2010. Improved penalization for determining the number of factors in approximate static factor models. *Statist. Probab. Lett.* 80, 1806–1813.
- Bai, J., 2003. Inferential theory for factor models of large dimensions. *Econometrica* 71, 135–171.
- Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. *Econometrica* 70, 191–221.
- Bai, J., Han, X., Shi, Y., 2017. Estimation and Inference of Structural Changes in High Dimensional Factor Models. Tech. Rep. <https://ssrn.com/abstract=2875193>.
- Baltagi, B.H., Kao, C., Wang, F., 2017. Identification and estimation of a large factor model with structural instability. *J. Econometrics* 197, 87–100.
- Barigozzi, M., Hallin, M., 2017. A network analysis of the volatility of high dimensional financial series. *J. R. Stat. Soc. Ser. C. Appl. Stat.* (66), 581–605.
- Barnett, I., Onnela, J.-P., 2016. Change point detection in correlation networks. *Sci. Rep.* 6.
- Bates, B.J., Plagborg-Møller, M., Stock, J.H., Watson, M.W., 2013. Consistent factor estimation in dynamic factor models with structural instability. *J. Econometrics* 177, 289–304.
- Bosq, D., 1998. *Nonparametric Statistics for Stochastic Process: Estimation and Prediction*. Springer.
- Brault, V., Ouadah, S., Sansonnet, L., Lévy-Leduc, C., 2016. Nonparametric homogeneity tests and multiple change-point estimation for analyzing large hi-c data matrices, arXiv preprint arXiv:1605.03751.
- Breitung, J., Eickmeier, S., 2011. Testing for structural breaks in dynamic factor models. *J. Econometrics* 163, 71–84.
- Brodsky, B.E., Darkhovsky, B.S., 1993. *Nonparametric Methods in Change-point Problems*. Springer.
- Chamberlain, G., Rothschild, M., 1983. Arbitrage, factor structure, and mean–variance analysis on large asset markets. *Econometrica* 51, 1281–1304.
- Chen, L., Dolado, J.J., Gonzalo, J., 2014. Detecting big structural breaks in large factor models. *J. Econometrics* 180, 30–48.
- Cheng, X., Liao, Z., Schorfheide, F., 2016. Shrinkage estimation of high-dimensional factor models with structural instabilities. *Rev. Econom. Stud.* 83, 1511–1543.
- Cho, H., 2016. Change-point detection in panel data via double CUSUM statistic. *Electron. J. Stat.* 10, 2000–2038.
- Cho, H., Barigozzi, M., Fryzlewicz, P., 2016. *factorcpt: Simultaneous change-point and factor analysis*, R package version 0.1.2.
- Cho, H., Fryzlewicz, P., 2011. Multiscale interpretation of taut string estimation and its connection to Unbalanced Haar wavelets. *Stat. Comput.* 21, 671–681.
- Cho, H., Fryzlewicz, P., 2012. Multiscale and multilevel technique for consistent segmentation of nonstationary time series. *Statist. Sinica* 22, 207–229.
- Cho, H., Fryzlewicz, P., 2015. Multiple change-point detection for high-dimensional time series via Sparsified Binary Segmentation. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 77, 475–507.
- Corradi, V., Swanson, N.R., 2014. Testing for structural stability of factor augmented forecasting models. *J. Econometrics* 182, 100–118.
- Davis, C., Kahan, W.M., 1970. The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.* 7, 1–46.
- Enikeeva, F., Harchaoui, Z., 2015. High-dimensional change-point detection with sparse alternatives, arXiv preprint, arXiv:1312.1900.
- Fan, J., Liao, Y., Mincheva, M., 2011a. High dimensional covariance matrix estimation in approximate factor models. *Ann. Statist.* 39, 3320–3356.
- Fan, J., Liao, Y., Mincheva, M., 2013. Large covariance estimation by thresholding principal orthogonal complements. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 75, 603–680.
- Fan, J., Lv, J., Qi, L., 2011b. Sparse high dimensional models in economics. *Annu. Rev. Econ.* 3, 291–317.
- Forni, M., Giannone, D., Lippi, M., Reichlin, L., 2009. Opening the black box: structural factor models versus structural vars. *Econometric Theory* 25, 1319–1347.
- Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2000. The Generalized Dynamic Factor Model: identification and estimation. *Rev. Econ. Stat.* 82, 540–554.
- Forni, M., Lippi, M., 2001. The Generalized Dynamic Factor Model: representation theory. *Econometric Theory* 17, 1113–1141.
- Fryzlewicz, P., 2014. Wild binary segmentation for multiple change-point detection. *Ann. Statist.* 42, 2243–2281.
- Fryzlewicz, P., Nason, G.P., 2006. Haar–Fisz estimation of evolutionary wavelet spectra. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 68, 611–634.
- Fryzlewicz, P., Subba Rao, S., 2014. Multiple-change-point detection for auto-regressive conditional heteroscedastic processes. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 76, 903–924.
- Gonçalves, S., Perron, B., 2014. Bootstrapping factor-augmented regression models. *J. Econometrics* 182, 156–173.
- Gonçalves, S., Perron, B., 2016. Bootstrapping factor models with cross sectional dependence, Preprint.
- Groen, J.J., Kapetanios, G., Price, S., 2013. Multivariate methods for monitoring structural change. *J. Appl. Econometrics* 28, 250–274.
- Hallin, M., Lippi, M., 2013. Factor models in high–dimensional time series: a time-domain approach. *Stochastic Process. Appl.* 123, 2678–2695.
- Han, X., Inoue, A., 2014. Tests for parameter instability in dynamic factor models. *Econometric Theory* 31, 1–36.
- Horváth, L., Hušková, M., 2012. Change-point detection in panel data. *J. Time Series Anal.* 33, 631–648.
- Jentsch, C., Politis, D.N., 2015. Covariance matrix estimation and linear process bootstrap for multivariate time series of possibly increasing dimension. *Ann. Statist.* 43, 1117–1140.
- Jirak, M., 2015. Uniform change point tests in high dimension. *Ann. Statist.* 43, 2451–2483.
- Korostelev, A., 1987. On minimax estimation of a discontinuous signal. *Theory Probab. Appl.* 32, 727–730.
- Li, J., Todorov, V., Tauchen, G., Lin, H., 2018. Rank tests at jump events. *J. Bus. Econom. Statist.* <http://dx.doi.org/10.1080/07350015.2017.1328362>. (in press).
- Ma, S., Su, L., 2016. Estimation of large dimensional factor models with an unknown number of breaks, Preprint.
- Massacci, D., 2017. Least squares estimation of large dimensional threshold factor models. *J. Econometrics* 197, 101–129.

- Merlevède, F., Peligrad, M., Rio, E., 2011. A Bernstein type inequality and moderate deviations for weakly dependent sequences. *Probab. Theory Related Fields* 151, 435–474.
- Nason, G.P., von Sachs, R., Kroisandt, G., 2000. Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 62, 271–292.
- Omranian, N., Mueller-Roeber, B., Nikoloski, Z., 2015. Segmentation of biological multivariate time-series data. *Sci. Rep.* 5.
- Onatski, A., 2010. Determining the number of factors from empirical distribution of eigenvalues. *Rev. Econ. Stat.* 92, 1004–1016.
- Onatski, A., 2015. Asymptotic analysis of the squared estimation error in misspecified factor models. *J. Econometrics* 186, 388–406.
- Patton, A., Politis, D.N., White, H., 2009. Correction to “Automatic block-length selection for the dependent bootstrap” by D. Politis and H. White. *Econometric Rev.* 28, 372–375.
- Pelger, M., 2015. Large-Dimensional Factor Modeling Based on High-Frequency Observations, Tech. Rep. Stanford University.
- Politis, D.N., Romano, J.P., 1994. The stationary bootstrap. *J. Amer. Statist. Assoc.* 89, 1303–1313.
- Politis, D.N., White, H., 2004. Automatic block-length selection for the dependent bootstrap. *Econometric Rev.* 23, 53–70.
- Schröder, A.L., Ombao, H., 2016. FreSpeD: Frequency-specific change-point detection in epileptic seizure multi-channel EEG data, Preprint.
- Stock, J.H., Watson, M., 2009. Forecasting in dynamic factor models subject to structural instability. In: J. C., N. S. (Eds.), *The Methodology and Practice of Econometrics. A Festschrift in Honour of David F. Hendry*. Oxford University Press, pp. 173–205.
- Stock, J.H., Watson, M.W., 2002. Forecasting using principal components from a large number of predictors. *J. Amer. Statist. Assoc.* 97, 1167–1179.
- Stock, J.H., Watson, M.W., 2003. Has the business cycle changed and why?. In: *NBER Macroeconomics Annual 2002*, vol. 17. MIT press, pp. 159–230.
- Sun, Z., Liu, X., Wang, L., 2016. A hybrid segmentation method for multivariate time series based on the dynamic factor model. *Stoch. Environ. Res. Risk Assess.* 31, 1–14.
- Trapani, L., 2013. On bootstrapping panel factor series. *J. Econometrics* 172, 127–141.
- Van Bellegem, S., von Sachs, R., 2008. Locally adaptive estimation of evolutionary wavelet spectra. *Ann. Statist.* 36, 1879–1924.
- Wang, T., Samworth, R.J., 2018. High dimensional change point estimation via sparse projection. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 80, 57–83.
- Yamamoto, Y., Tanaka, S., 2015. Testing for factor loading structural change under common breaks. *J. Econometrics* 189, 187–206.
- Yu, Y., Wang, T., Samworth, R.J., 2015. A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika* 102, 315–323.