Annals of the Institute of Statistical Mathematics (2021) 73:821–853
https://doi.org/10.1007/s10463-020-00773-0

Robust test for structural instability in dynamic factor models

Byungsoo Kim[1] · **Junmo Song**[2] · **Changryong Baek**[3]

Received: 9 April 2020 / Revised: 7 October 2020 / Accepted: 9 October 2020 / Published online: 2 January 2021

Abstract

In this paper, we consider a robust test for structural breaks in dynamic factor models. The proposed framework considers structural changes when the underlying high-dimensional time series is contaminated by outlying observations, which are often observed in many real applications such as fMRI, economics and finance. We propose a test based on the robust estimation of a vector autoregressive model for principal component factors using minimum density power divergence. The simulations study shows excellent finite sample performance, higher powers while achieving good sizes in all cases considered. Our method is illustrated to the resting state fMRI series to detect brain connectivity changes.

Keywords High-dimensional time series · Dynamic factor models · Minimum density power divergence · Parameter change test · Outliers

All authors contributed equally to this article.
This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korea government (NRF-2019R1F1A1057104) (C. Baek), (NRF-2019R1I1A3A01056924) (J. Song), and (NRF-2019R1C1C1004662) (B. Kim).
The online version of this article contains supplementary material.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10463-020-00773-0) contains supplementary material, which is available to authorized users.

✉ Changryong Baek
crbaek@skku.edu

Byungsoo Kim
bkim@yu.ac.kr

Junmo Song
jsong@knu.ac.kr

[1] Department of Statistics, Yeungnam University, 280 Daehak-ro, Gyeongsan, Gyeongbuk 38541, Korea

[2] Department of Statistics, Kyungpook National University, 80 Daehakro, Bukgu, Daegu 41566, Korea

[3] Department of Statistics, Sungkyunkwan University, 25-2 Sungkyunkwan-ro, Jongno-gu, Seoul 03063, Korea

1 Introduction

Over the last decade, innovations in computer science and technology have rapidly shifted our focus to high-dimensional time series (HDTS). Dynamic factor models (DFMs), among others, have grown to be among the most popular and successful models for HDTS since few latent factors can explain much higher-dimensional dynamics. DFMs have abundant applications in economics, finance, medical engineering, hydrology and related fields. See, for example, Bai and Ng (2008), Stock and Watson (2011), Baek et al. (2018))and the references therein.

DFMs are generally assumed to be time-homogeneous, that is, they undergo no structural or parameter changes. However, this assumption is no longer valid in many applications, especially in HDTS context. Indeed, there is strong evidence of structural changes, for example, in macroeconomic time series (e.g., Balke and Fomby 1994; Atkinson et al. 1997; Han and Inoue 2015) and in volatility of financial series (e.g., Lee et al. 2015; Song and Baek 2019). We focus on the problem of testing for parameter changes in DFMs. In particular, we study the situation where the data are contaminated by outlying observations. It is well known that existing naive tests for parameter changes are severely damaged by outliers (e.g., Song and Kang 2019; Song 2020). That is, outliers can lead to severe size distortions and power losses. This obviously indicates that, when atypical observations are included in a data set being suspected of having parameter changes, whether the testing results are due to genuine changes or not cannot be readily determined. It is, therefore, crucial to take into account such outlying observations in statistical tests for detecting parameter changes.

More concrete examples can be found in brain connectivity studies in neuroscience, where the scientific question of interest involves detecting and locating connectivity changes in brain regions (e.g., Cribben et al. 2012; Baek et al. 2021). Functional magnetic resonance imaging (fMRI) data measured at hundreds of voxels are used in analyses, and fMRI series are widely reported to contain multiple outlying observations due to issues in acquisition such as the anatomical alignment of images and head movement correction. Such outliers can lead to poor performance in time-series modeling (e.g., Chen and An 1998; Ledolter 1989) and can potentially result in wrong inference about brain mapping (e.g., Poldrack 2012; Magnotti and Billor 2014).

The widely used practice for handling outliers in multivariate time series is to apply univariate outlier detection methods to each component and remove the outlier effects before applying multivariate time series modeling. However, this approach cannot fully adjust for outlier effects as it ignores joint underlying dynamics such as correlation structure in the multivariate system (e.g., Tsay et al. 2000). Even with the correct detection of outliers, dropping or adjusting outliers has the potential to be dangerous depending on the area of interest, particularly with temporal correlations. In this study, we do not attempt to remove and ignore outliers, but rather reduce the impact of possible outliers in the testing procedure by using a divergence-based method.

Divergence measures such as Kullback–Leibler (KL) divergence were originally introduced as a measure of discrepancy between two probability distributions. Later, several divergences were proposed for robust estimation (e.g., Basu et al. 1998; Fujisawa and Eguchi 2006; Ghosh and Basu 2017) and were found to accommodate outlier effects successfully with little loss in efficiency. More recently, the robust estimation problem has moved to the testing area. Several authors have suggested divergence-based tests and have demonstrated their robustness and efficiency (e.g., Basu et al. 2013, 2016; Batsidis et al. 2013; Song and Kang 2019). We focus on the so-called density power (DP) divergence and a DP divergence version of the score test for parameter changes as they enjoy the merits of both the score test and the robustness, as studied in Song and Kang (2019). Besides outliers effects, a further challenge is that the dimension can possibly diverge faster than the sample size. Our strategy is to reduce the structural changes of DFMs into a finite-dimensional problem by focusing on the structural instability of several principal component factors obtained from the whole sample. This is in the spirit of Stock and Watson (2002), Breitung and Eickmeier (2011), Chen et al. (2014) and Han and Inoue (2015) to name a few. That is, we propose a robust procedure by applying a DP divergence estimation on the principal factors based on VAR(p) models. Further details are provided in Sect. 2.

The rest of this paper is organized as follows. We first explain the reasonings behind dimension reduction by clarifying models and hypotheses of interest in Sect. 2. Our proposed robust testing procedure with the supporting theoretical results is discussed in Sect. 3. Finite sample performances are extensively examined in Sect. 4 through Monte Carlo simulations. In addition, our method is illustrated to the resting state fMRI series in Sect. 5. Section 6 provides the conclusion, and all proofs are given in Sect. 7.

2 Preliminaries

Consider a DFM of q-vector time series $\{X_n\}_{t=1}^n$ given by

$$X_t = \Lambda_t Y_t + \varepsilon_t,$$

where Λ_t is a loading matrix of dimension $q \times r$, $\{Y_t\}_{t=1}^n$ is an r-vector time series of latent r factors, and $\{\varepsilon_t\}_{t=1}^n$ are idiosyncratic disturbances which are assumed to be uncorrelated with the factors $\{Y_t\}_{t=1}^n$. The factors $\{Y_t\}_{t=1}^n$ are typically assumed to follow a vector autoregressive model of order p (VAR(p)),

$$\boldsymbol{\Phi}(B)Y_t = \epsilon_t,$$

where $\boldsymbol{\Phi}(B) = I - A_1 B - \ldots - A_p B^p$ is an autoregressive matrix polynomial expressed in terms of the backshift operator B, A_i, $i = 1, \ldots, p$, are $r \times r$ coefficients matrices, and innovations $\{\epsilon_t\}_{t=1}^n$ are assumed to follow multivariate normal distribution, denoted by $\mathcal{N}(0, \Sigma)$. The factors are usually estimated through the principal component analysis (PCA), as follows: Compute the sample covariance matrix $\widehat{\Sigma}_X := n^{-1} \sum_{t=1}^n X_t X_t^T = \widehat{U} \mathrm{diag}(\hat{e}_1, \ldots, \hat{e}_q) \widehat{U}^T$, where $\hat{e}_1 \geq \hat{e}_2 \ldots \geq \hat{e}_q$ are ordered eigenvalues and $\widehat{U} = (\hat{u}_1, \ldots, \hat{u}_q)$ is $q \times q$ orthogonal matrix. Then, $r(\leq q)$ principal factors are obtained by

$$\widehat{Y}_t = \frac{1}{\sqrt{q}} \widehat{U}_r^T X_t, \quad t = 1, \dots, n,$$

where $\widehat{U}_r = (\hat{u}_1, \dots, \hat{u}_r)$. The number of factors r is crucial in finite sample performance and the information criteria of Bai and Ng (2002) is widely used in practice.

We are interested in testing structural stability in the factor loadings:

$$H_0 : \Lambda_1 = \dots = \Lambda_n$$

against the alternative hypothesis of common structural changes,

$$H_1 : \Lambda_1 = \dots = \Lambda_k \neq \Lambda_{k+1} = \dots = \Lambda_n$$

at a common date k. As has been studied in Stock and Watson (2009), Breitung and Eickmeier (2011), Chen et al. (2014), Han and Inoue (2015) and elsewhere, the size of the break should be sufficiently large to detect changes. More importantly, this problem can be reduced to that of detecting instabilities of the r latent principal component factors since the PCA implicitly imposes the restriction that the estimated factor loadings are time-homogeneous. The intuition is that a DFM with structural changes in loadings can be rewritten as another DFM with constant loading but with a larger set of factors.

For example, consider a simple DFM with one common structural change at k and factors are distinct for all dimensions. In the matrix form, this can be written as

$$\mathbb{X}_1 = \mathbb{Y}_1 \Lambda_1^T + e_1, \quad \mathbb{X}_2 = \mathbb{Y}_2 \Lambda_{k+1}^T + e_2,$$

where $\mathbb{X}_1^T = (X_1, \dots, X_k)$, $\mathbb{Y}_1^T = (Y_1, \dots, Y_k)$, $e_1^T = (e_1, \dots, e_k)$, $\mathbb{X}_2^T = (X_{k+1}, \dots, X_n)$, $\mathbb{Y}_2^T = (Y_{k+1}, \dots, Y_n)$ and $e_2^T = (e_{k+1}, \dots, e_n)$, Λ_1 and Λ_{k+1} are the loading matrices before/after the change at k, respectively. Then, this can be rewritten as

$$\begin{pmatrix} \mathbb{X}_1 \\ \mathbb{X}_2 \end{pmatrix} = \begin{pmatrix} \mathbb{Y}_1 & 0 \\ 0 & \mathbb{Y}_2 \end{pmatrix} \begin{pmatrix} \Lambda_1^T \\ \Lambda_{k+1}^T \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}. \tag{1}$$

This is another larger set of a factor model with a constant loading matrix $(\Lambda_1 \ \Lambda_{k+1})^T$. Furthermore, under mild assumptions on the approximate factor models where temporal, spatial correlation and heteroskedasticity are allowed in idiosyncratic errors, PCA factors consistently estimate the factors on the extended space (e.g., Theorem 1 of Bai 2003, Proposition 1 of Chen et al. 2014) and the information criteria of Bai and Ng (2007) consistently estimates the number of equivalent factor models (1) (e.g., Proposition 1 of Han and Inoue 2015). Therefore, principal factors can be used as good low-dimensional proxies under the alternative hypothesis once factors are obtained from the whole sample using PCA. Most of the tests for the structural instability of factor loadings are based on the instability of factors rather than the loading matrix. For instance, Chen et al. (2014) proposed a structural break test based on the first principal factor after regressing out the remaining factors. Han and Inoue (2015) proposed Wald and LM-like statistics based on the pre- and post-break subsample averages of $\widehat{Y}_t \widehat{Y}_t^T$.

However, to the best of our knowledge, no tests have been proposed that accommodate outlying observations. To address this gap, we transformed the structural instability of DFMs into parameter changes of VAR(p) models on principal factors. This turns the hypotheses of the loading matrices into parameter changes of the VAR(p) model as

H'_0 : The VAR(p) model parameter for Y_t does not change over $Y_1, \dots, Y_n$,

against

H'_1 : not H'_0.

That is, we intend to test H'_0 against H'_1, particularly in the presence of outliers. Our proposed method is based on the so-called minimum DP divergence estimator (MDPDE) of VAR(p) parameters on the principal factors. It can also be used for the robust testing of parameter changes in VAR(p) models if the data are directly used in a moderate dimension instead of factors. Hence, for simplicity and independent interest on the smaller dimension, we shall first describe the results with VAR(p) models, then we describe the results for DFM with estimated factors under suitable additional assumptions. We introduce the MDPDE for the VAR(p) parameter in Sect. 3.1, and a DP divergence-based test for parameter change in Sect. 3.2.

3 Main results

In this section, we introduce a robust test for parameter changes in DFMs based on DP divergence with factors being assumed as VAR models. We first consider VAR(p) models and then extend to DFM with factors following VAR(p) models.

3.1 Minimum DP divergence estimator for Gaussian VAR models

For the given two density functions f and g, Basu et al. (1998) introduced the DP divergence d_α defined as

$$d_\alpha(g,f) = \begin{cases} \int \left\{ f^{1+\alpha}(y) - (1+\frac{1}{\alpha})g(y)f^\alpha(y) + \frac{1}{\alpha}g^{1+\alpha}(y) \right\} \mathrm{d}y, & \alpha > 0, \\ \int g(y)\{\log g(y) - \log f(y)\}\mathrm{d}y, & \alpha = 0. \end{cases}$$

For a parametric family $\{F_\theta; \theta \in \Theta\}$ possessing densities $\{f_\theta\}$ and a distribution G with density g, Basu et al. (1998) defined the minimum DP divergence functional $T_\alpha(G)$ by the requirement that $d_\alpha(g,f_{T_\alpha(G)}) = \min_{\theta\in\Theta} d_\alpha(g,f_\theta)$. In particular, if $G = F_{\theta_0} \in \{F_\theta\}$, then $T_\alpha(F_{\theta_0}) = \theta_0$. Based on the above, given a random sample $\{Y_1, \dots, Y_n\}$ with an unknown density g, the MDPDE is defined by

$$\hat{\theta}_{\alpha,n} = \operatorname{argmin}_{\theta\in\Theta} H_{\alpha,n}(\theta),$$

where $H_{\alpha,n}(\theta) = \frac{1}{n}\sum_{t=1}^n h_{\alpha,t}(\theta)$ and

$$h_{\alpha,t}(\theta) = \begin{cases} \int f_\theta^{1+\alpha}(y)\mathrm{d}y - \left(1 + \frac{1}{\alpha}\right) f_\theta^\alpha(Y_t), & \alpha > 0, \\ -\log f_\theta(Y_t), & \alpha = 0. \end{cases}$$

Note that when $\alpha = 0$ and 1, the MDPDE is the same as the MLE and L_2-distance estimator, respectively. The tuning parameter α controls the trade-off between the robustness and asymptotic efficiency of the MDPDE. That is, if α increases, the robustness of the MDPDE improves, but the asymptotic efficiency decreases. Basu et al. (1998) demonstrated that the estimator is robust against outliers, but still retains high efficiency when the true distribution belongs to a parametric family $\{F_\theta\}$ and α is close to zero. Although they introduced the MDPDE based on the independent and identically distributed (i.i.d.) random variables, we extend their method to a multivariate time series.

To apply the above procedure to the VAR models, we prepare the conditional version of the MDPDE. Let $\{f_\theta(y|x)\}$ be a parametric family of regression models indexed by the parameter θ, and let $f_{\theta_0}(y|x)$ be the true conditional density of Y given $X = x$. Then, the conditional version of the DP divergence can be defined as

$$\begin{aligned} & d_\alpha\left(f_{\theta_0}(\cdot|x), f_\theta(\cdot|x)\right) \\ & = \begin{cases} \int \left\{ f_\theta^{1+\alpha}(y|x) - (1 + \frac{1}{\alpha}) f_{\theta_0}(y|x) f_\theta^\alpha(y|x) + \frac{1}{\alpha} f_{\theta_0}^{1+\alpha}(y|x) \right\} \mathrm{d}y, & \alpha > 0, \\ \int f_{\theta_0}(y|x) \left\{ \log f_{\theta_0}(y|x) - \log f_\theta(y|x) \right\} \mathrm{d}y, & \alpha = 0. \end{cases} \end{aligned}$$

Given the observations $(X_1, Y_1), \dots, (X_n, Y_n)$, the conditional version of the MDPDE is defined as

$$\hat{\theta}_{\alpha,n} = \operatorname{argmin}_{\theta \in \Theta} H_{\alpha,n}(\theta),$$

where $H_{\alpha,n}(\theta) = \frac{1}{n} \sum_{t=1}^n h_{\alpha,t}(\theta)$ and

$$h_{\alpha,t}(\theta) = \begin{cases} \int f_\theta^{1+\alpha}(y|X_t)\mathrm{d}y - \left(1 + \frac{1}{\alpha}\right) f_\theta^\alpha(Y_t|X_t), & \alpha > 0, \\ -\log f_\theta(Y_t|X_t), & \alpha = 0, \end{cases} \tag{2}$$

(cf. Section 2 of Lee and Song 2009).

The MDPDE for the VAR models is constructed as follows. Suppose that r-dimensional multivariate time series Y_t follows VAR(p);

$$Y_t = c + \sum_{i=1}^p A_i Y_{t-i} + \epsilon_t, \tag{3}$$

where c is a r-dimensional vector, A_i are $r \times r$ matrices for $i = 1, \dots, p$, and ϵ_t is a sequence of i.i.d. multivariate normal random vectors with a zero mean and covariance matrix Σ, which is symmetric and positive definite. For notational convenience, we denote η as the dimension of the parameters, *i.e.*, $\eta = r + r^2 p + r(r+1)/2$, and set $\theta = (\theta_1, \dots, \theta_\eta)^T = (c^T, vec(A_1)^T, \dots, vec(A_p)^T, vech(\Sigma)^T)^T \in \Theta \subset \mathbb{R}^\eta$, where $vec(A)$ transforms matrix A into a column vector by stacking all columns of A, and $vech(A)$ converts the lower triangular elements of A into a column vector.

Suppose that $Y_1, \ldots, Y_n$ are observed from (3) and let $Y_{a:b}$ denote the observations from $t = a$ to $t = b$. Then, conditional on the initial values $Y_{1-p:0}$, $Y_t|Y_{t-p:t-1}$ follows $\mathcal{N}(c + \sum_{i=1}^{p} A_i Y_{t-i}, \Sigma)$ for $t = 1, \ldots, n$. If we denote the conditional density of Y_t by $f_\theta(y|Y_{t-p:t-1})$, then from (2), we can define the MDPDE as

$$\hat{\theta}_{\alpha,n} = \operatorname{argmin}_{\theta\in\Theta} H_{\alpha,n}(\theta),$$

where $H_{\alpha,n}(\theta) = \frac{1}{n}\sum_{t=1}^{n} h_{\alpha,t}(\theta)$,

$$h_{\alpha,t}(\theta) = \begin{cases} \int f_\theta^{1+\alpha}(y|Y_{t-p:t-1})dy - \left(1+\frac{1}{\alpha}\right) f_\theta^{\alpha}(Y_t|Y_{t-p:t-1}), & \alpha > 0, \\ -\log f_\theta(Y_t|Y_{t-p:t-1}), & \alpha = 0, \end{cases}$$
$$= \begin{cases} \frac{1}{\{(2\pi)^r \det(\Sigma)\}^{\alpha/2}} \left\{ \frac{1}{(1+\alpha)^{r/2}} - \left(1+\frac{1}{\alpha}\right) \exp\left(-\frac{\alpha}{2}\epsilon_t^T \Sigma^{-1}\epsilon_t\right) \right\}, & \alpha > 0, \\ \frac{r}{2}\log 2\pi + \frac{1}{2}\log\det(\Sigma) + \frac{1}{2}\epsilon_t^T \Sigma^{-1}\epsilon_t, & \alpha = 0, \end{cases}$$

and $\epsilon_t = Y_t - c - \sum_{i=1}^{p} A_i Y_{t-i}$.

Now, we establish the consistency and asymptotic normality of the MDPDE. Throughout this paper,

$$\theta_0 = (\theta_{01}, \ldots, \theta_{0\eta})^T = (c_0^T, vec(A_{01})^T, \ldots, vec(A_{0p})^T, vech(\Sigma_0)^T)^T$$

denotes the true value of θ, and $E(\cdot)$ is taken under θ_0. To study the asymptotic properties of the MDPDE, we assume that the following conditions hold.

(A1) $\det(I_r - \sum_{i=1}^{p} A_i z^i) \neq 0$ for $|z| \leq 1$, where I_r is an $r \times r$ identity matrix.

(A2) $\theta_0 \in \Theta$ and Θ is compact. In addition, for all Σ such that $vech(\Sigma)$ belongs to Θ, there exist positive constants λ_L and λ_U satisfying

$$0 < \lambda_L \leq \lambda_{\min}(\Sigma) \leq \lambda_{\max}(\Sigma) \leq \lambda_U < \infty,$$

where $\lambda_{\min}(\Sigma)$ and $\lambda_{\max}(\Sigma)$ denote the smallest and largest eigenvalues of Σ, respectively.

(A3) θ_0 is an interior point of Θ.

Remark 1 It is well known that under **(A1)**, Y_t is weakly stationary and ergodic. However, since we assume that ϵ_t is an i.i.d. multivariate normal random vector, Y_t constitutes a Gaussian process (cf. p.16 of Lütkepohl 2005). Hence, condition **(A1)** also ensures the strict stationarity of Y_t.

Note that, regardless of the values of $Y_{1-p:0}$, $\{h_{\alpha,t}(\theta)|t \geq p+1\}$ becomes a stationary and ergodic process under **(A1)**. J_α and K_α defined in the following theorem are obtained from the stationary process.

Remark 2 Under the conditions **(A1)** and **(A2)**, it can be shown that J_α and K_α are non-singular matrices, and hence invertible. See Lemma 2.

Under the above conditions, we obtain the following asymptotic results.

Theorem 1 *Suppose that conditions* **(A1)** *and* **(A2)** *hold. Then,*

$$\hat{\theta}_{\alpha,n} \xrightarrow{a.s.} \theta_0 \quad \text{as} \quad n \to \infty.$$

Theorem 2 *Suppose that conditions* **(A1)–(A3)** *hold. Then,*

$$\sqrt{n}(\hat{\theta}_{\alpha,n} - \theta_0) \xrightarrow{d} \mathcal{N}(0, J_\alpha^{-1} K_\alpha J_\alpha^{-1}) \quad \text{as} \quad n \to \infty,$$

where

$$J_\alpha = -E\left(\frac{\partial^2 h_{\alpha,t}(\theta_0)}{\partial\theta\partial\theta^T} \right) \quad \text{and} \quad K_\alpha = E\left(\frac{\partial h_{\alpha,t}(\theta_0)}{\partial\theta} \frac{\partial h_{\alpha,t}(\theta_0)}{\partial\theta^T} \right).$$

Remark 3 The Gaussianity of VAR(p) model is needed in our framework. One may consider a quasi-MDPD estimation. However, our unreported simulations study suggests that it may require non-trivial assumptions for asymptotics. For example, the sufficient conditions for the positive definiteness of $-J_\alpha$ and K_α are not straightforward for non-Gaussian cases.

Remark 4 The robustness of an estimator can be assessed through the so-called influence function (IF), which measures the gross error sensitivity by considering contaminated distribution. See Chapter 8.3b of Hampel et al. (1986). When the IF of an estimator is bounded, the estimator is called B-robust. It corresponds to finite gross error sensitivity, meaning that the influence of outlier is limited like the trimmed sample mean. Indeed, one can show that our MDPDE for VAR(p) model is B-robust by using the results in the proof of Lemma S1 in Supplementary Material.

3.2 DP divergence-based test for parameter changes in VAR models

Here, we propose a robust test for parameter changes in VAR(p) models. Recall that, one wish to test the following hypotheses, particularly in the presence of outliers:

$$H_0' : \text{The true parameter } \theta_0 \text{ does not change over } Y_1, \dots, Y_n, \text{ against}$$
$$H_1' : \text{not } H_0'.$$

To this end, we follow the procedure by Song and Kang (2019) to introduce a test statistics. That is, we use the objective function of the MDPDE , i.e., $H_{\alpha,n}(\theta)$, to construct a robust test similar in form to a score test. According to Taylor's theorem, we have, for each $s \in [0, 1]$,

$$\frac{[ns]}{\sqrt{n}} \frac{\partial H_{\alpha,[ns]}(\hat{\theta}_{\alpha,n})}{\partial\theta} = \frac{[ns]}{\sqrt{n}} \frac{\partial H_{\alpha,[ns]}(\theta_0)}{\partial\theta} + \frac{[ns]}{n} \frac{\partial^2 H_{\alpha,[ns]}(\theta^*_{\alpha,n,s})}{\partial\theta\partial\theta^T} \sqrt{n}\left(\hat{\theta}_{\alpha,n} - \theta_0\right), \quad (4)$$

where $\theta^*_{\alpha,n,s}$ is an intermediate point between θ_0 and $\hat{\theta}_{\alpha,n}$. Here, since $\partial H_{\alpha,n}(\hat{\theta}_{\alpha,n})/\partial\theta = 0$, we have, for $s = 1$,

$$\sqrt{n}\frac{\partial H_{\alpha,n}(\theta_0)}{\partial\theta} + \frac{\partial^2 H_{\alpha,n}(\theta^*_{\alpha,n,1})}{\partial\theta\partial\theta^T}\sqrt{n}\big(\hat{\theta}_{\alpha,n} - \theta_0\big) = 0.$$

It can thus be written as

$$\sqrt{n}\big(\hat{\theta}_{\alpha,n} - \theta_0\big) = J_\alpha^{-1}\sqrt{n}\frac{\partial H_{\alpha,n}(\theta_0)}{\partial\theta} + J_\alpha^{-1}\big(B_{\alpha,n} + J_\alpha\big)\sqrt{n}\big(\hat{\theta}_{\alpha,n} - \theta_0\big),$$

where $B_{\alpha,n} = \partial^2 H_{\alpha,n}(\theta^*_{\alpha,n,1})/\partial\theta\partial\theta^T$ and J_α is given in Theorem 2. By (4), we obtain

$$\begin{aligned}\frac{[ns]}{\sqrt{n}}\frac{\partial H_{\alpha,[ns]}(\hat{\theta}_{\alpha,n})}{\partial\theta} =& \frac{[ns]}{\sqrt{n}}\frac{\partial H_{\alpha,[ns]}(\theta_0)}{\partial\theta} + \frac{[ns]}{n}\frac{\partial^2 H_{\alpha,[ns]}(\theta^*_{\alpha,n,s})}{\partial\theta\partial\theta^T}J_\alpha^{-1}\sqrt{n}\frac{\partial H_{\alpha,n}(\theta_0)}{\partial\theta} \\ &+ \frac{[ns]}{n}\frac{\partial^2 H_{\alpha,[ns]}(\theta^*_{\alpha,n,s})}{\partial\theta\partial\theta^T}J_\alpha^{-1}\big(B_{\alpha,n} + J_\alpha\big)\sqrt{n}\big(\hat{\theta}_{\alpha,n} - \theta_0\big).\end{aligned} \tag{5}$$

This is the building block for our test statistics. In the proof of Sect. 7, we show that the first two terms in the right-hand side of (5) converge weakly to a Brownian bridge, and the last term is asymptotically negligible. The results are formally stated in the following theorem.

Theorem 3 *Suppose that conditions* **(A1)–(A3)** *hold. Then, under* H_0', *we have*

$$K_\alpha^{-1/2}\frac{[ns]}{\sqrt{n}}\frac{\partial H_{\alpha,[ns]}(\hat{\theta}_{\alpha,n})}{\partial\theta} \xrightarrow{w} B^o_\eta(s) \quad \text{in } \mathbb{D}\left([0,1], \mathbb{R}^\eta\right),$$

where B^o_η *is a* η*-dimensional standard Brownian bridge, and thus,*

$$T_n^\alpha := \max_{1\le k\le n}\frac{k^2}{n}\frac{\partial H_{\alpha,k}(\hat{\theta}_{\alpha,n})}{\partial\theta^T}K_\alpha^{-1}\frac{\partial H_{\alpha,k}(\hat{\theta}_{\alpha,n})}{\partial\theta} \xrightarrow{d} \sup_{0\le s\le 1}\|B^o_\eta(s)\|_2^2.$$

Remark 5 T_n^α with $\alpha = 0$ becomes the score test for parameter changes defined by

$$T_n := \max_{1\le k\le n}\frac{k^2}{n}\frac{\partial H_{0,k}(\hat{\theta}_{0,n})}{\partial\theta^T}I^{-1}\frac{\partial H_{0,k}(\hat{\theta}_{0,n})}{\partial\theta},$$

where $\hat{\theta}_{0,n}$ is the MLE, $H_{0,k}(\theta)$ and $h_{0,t}(\theta)$ correspond to $H_{\alpha,k}(\theta)$ and $h_{\alpha,t}(\theta)$ with $\alpha = 0$, respectively, and $I = E\big(\partial^2 h_{0,t}(\theta_0)/\partial\theta\partial\theta^T\big)$. Since the score function is induced from Kullback–Leibler divergence, our test is a DP divergence version of the score test.

Remark 6 As an estimator of K_α, we can use

$$\hat{K}_\alpha = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial h_{\alpha,t}(\hat{\theta}_{\alpha,n})}{\partial\theta}\frac{\partial h_{\alpha,t}(\hat{\theta}_{\alpha,n})}{\partial\theta^T}.$$

For the consistency of the estimator, see Lemma 5.

Remark 7 As addressed in Song and Kang (2019), T_n^α is a CUSUM-type test because it can be expressed that

$$\begin{aligned}
&\frac{[ns]}{\sqrt{n}}\frac{\partial H_{\alpha,[ns]}(\hat{\theta}_{\alpha,n})}{\partial\theta} \\
&= \frac{1}{\sqrt{n}}\left\{\sum_{t=1}^{[ns]}\frac{\partial h_{\alpha,t}(\hat{\theta}_{\alpha,n})}{\partial\theta} - \frac{[ns]}{n}\sum_{t=1}^{n}\frac{\partial h_{\alpha,t}(\hat{\theta}_{\alpha,n})}{\partial\theta}\right\} \\
&= \frac{[ns]}{n}\left(1-\frac{[ns]}{n}\right)\sqrt{n}\left(\frac{1}{[ns]}\sum_{t=1}^{[ns]}\frac{\partial h_{\alpha,t}(\hat{\theta}_{\alpha,n})}{\partial\theta} - \frac{1}{n-[ns]}\sum_{t=[ns]+1}^{n}\frac{\partial h_{\alpha,t}(\hat{\theta}_{\alpha,n})}{\partial\theta}\right).
\end{aligned}$$

Hence, when the null hypothesis is rejected, one can locate a change point as the maximizer of the test statistics. That is,

$$\text{argmax}_{\,1\le k\le n}\frac{k^2}{n}\frac{\partial H_{\alpha,k}(\hat{\theta}_{\alpha,n})}{\partial\theta^T}K_\alpha^{-1}\frac{\partial H_{\alpha,k}(\hat{\theta}_{\alpha,n})}{\partial\theta} =: \text{argmax}_{\,1\le k\le n}T_{n,k}^\alpha.$$

For more details on the change-point estimation in CUSUM-type test, see, for example, Robbins et al. (2011).

Remark 8 Several authors have investigated the selection criterion of optimal α in MDPD estimation procedure. See, for example, Warwick (2005) and Durio and Isaia (2011). However, to the best of our knowledge, no in-depth studies have been conducted on systematic selection in the testing problem. Song and Kang (2019) illustrated a selecting rule in terms of forecasting performance. Since the aim of our fMRI data analysis below is not forecasting, their method may not be plausible in such type of data. In our data analysis, we do not try to select an optimal α to locate change-points. Instead, we implement the test T_n^α for several α's and incorporate the results to make a decision on change-points. That is, for each α considered, we estimate change-points and then decide the points selected in common as final change-points in the spirit of model ensemble. Here, it should be importantly noted that the empirical power of T_n^α shows a tendency to decrease with an increase in α (see Sect. 4). Since too large α can lead to a significant loss in powers, we recommend to consider α values in [0,0.5]. When one wish to make an inference based on an optimal α, one may consider the rule used in the data analyses in Song and Kang (2019) and Song (2020).

Remark 9 Binary segmentation (BS) can be used for multiple change-points detection. The BS is an iterative procedure to identify change-points by splitting the sample into two based on the previously claimed change point. See, for example, Baek and Pipiras (2014) for details.

3.3 DP divergence based test for DFM

Now, consider the DFM where factors are estimated using PCA. Since we have established results for VAR(p) process in earlier sections, the results are in fact immediate due to the blessing of dimensionality for factor models. We briefly mention required assumptions behind this extension. Detailed rate of convergence is beyond the scope of this paper due to the nonlinear structure of the objective functions in the test statistics.

Suppose that the Assumptions A–G of Bai (2003) holds. Then, Theorem 1 and Lemma B.3 of Bai (2003) imply that the estimated factors $\widehat{\mathbb{Y}}$ in a matrix form satisfy

$$\frac{1}{n}\left(\widehat{\mathbb{Y}} - \mathbb{Y}G\right)^T \widehat{\mathbb{Y}} = O_p\left(\frac{1}{\min(q,n)}\right), \tag{6}$$

where G is a suitable non-singular transformation such that loading and factors are identifiable. Therefore, if

$$\frac{\sqrt{n}}{q} \to 0, \quad n,q \to \infty, \tag{7}$$

then (6) implies that the estimation errors are negligible and we can treat factors $\{G'Y_t\}$ as *known* (see Bai 2003 p. 146). This is what blessing of dimensionality meant in DFMs. Hence, Theorem 2 and Theorem 3 assuming true factors $\{G'Y_t\}$ hold for estimated factor models under Assumptions A–G of Bai (2003) as long as (7) holds. The true parameters θ_0 corresponds to VAR(p) model of $\{G'Y_t\}$.

However, it requires the consistent estimation of the number of factors. This can be done by considering the information criteria (IC) of Bai and Ng (2002), namely,

$$\hat{r} = \operatorname{argmin}_{0\le k\le K} \{\log S(k) + kg(q,n)\}, \tag{8}$$

where $S(k)$ is the sum of squared residuals $S(k) = (qn)^{-1}\sum_{i=1}^{q}\sum_{j=1}^{n}(X_t - \widehat{\Lambda}_t\widehat{Y}_t)'(X_t - \widehat{\Lambda}_t\widehat{Y}_t)$ assuming k factors, and penalty function $g(q,\ n)$ satisfies $g(q,n) \to 0$ and $\min(q,n)g(q,n) \to \infty$ as $q,n \to \infty$. The consistency of IC criteria was provided in Bai and Ng (2002, 2007) under the null hypothesis of no parameter changes. Furthermore, Proposition 1 of Han and Inoue (2015) showed that IC criteria consistently estimate the number of factors even with possible breaks under the alternative hypothesis. However, our simulations result not reported here shows that IC criteria tends to overestimate the number of factors due to outlying observations. Hence, we slightly modified the IC criteria to accommodate outlying observations. The modification is to replace outlying observations by nearby values. For example, we identify the tth observation in ith dimension denoted by $X_{t,i}$ is outlying if

$$0.6745\left|\frac{X_{t,i} - \operatorname{med}(X_{\cdot,i})}{\operatorname{mad}(X_{\cdot,i})}\right| > 2,$$

where med($X_{\cdot,i}$) and mad($X_{\cdot,i}$) are the median and mean absolute deviation of i-th dimension. See Iglewicz and Hoaglin (1993) for more details. Then, this observation is replaced by linearly interpolating nearby observations. Trimmed mean can be used to calculate the sum of squared residuals $S(k)$ as well.

Therefore, our proposed instability test is valid for DFM and can be summarized as follows:

Step 1. Determine the number of factors $\hat{r}$ from IC (8) after accommodating outlying observations.
Step 2. Estimate PCA factors $\{\hat{Y}_t\}$ and then estimate parameter using MDPDE.
Step 3. Apply the proposed test in Sect. 3.2 to the estimated factors $\{\hat{Y}_t\}$.

All in all, our proposed DP divergence-based test can be naturally extended to DFM due to blessing of dimensionality (7) under standard assumptions on factor models (e.g., Bai 2003) together with Theorems 2 and 3 in Sect. 3.

4 Simulation study

Here, we report the finite sample performance of our proposed test. We first report the finite sample performance in a VAR model, and assessing performance in DFMs. The simulation settings are described as follows. We consider a Gaussian bivariate VAR(1) model, $Y_t = (Y_{t,1}, Y_{t,2})^T$. The performance measures are the empirical sizes and powers of the tests, obtained as the number of rejections of the null hypothesis out of 2000 replications. The empirical sizes are calculated at the nominal levels 0.05 and 0.1, where the corresponding critical values are 5.635 and 5.060, respectively, which are obtained through Monte Carlo simulations. The empirical powers are evaluated at the nominal level of 0.05, and we considered three sample sizes of 500, 1000, and 2000. As a baseline, we compare our proposed test, T_n^α with $\alpha > 0$, with the score test T_n.

The first scenario is the data free from outliers. The data generating processes (DGPs) comprise the following four models:

$$\text{Model 1:} \quad c = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad A = \begin{pmatrix} 0.1 & -0.2 \\ 0.5 & 1 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

$$\text{Model 2:} \quad c = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad A = \begin{pmatrix} 0.1 & -0.2 \\ 0.5 & 1 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

$$\text{Model 3:} \quad c = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad A = \begin{pmatrix} 0.5 & -1 \\ 0 & -0.5 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

$$\text{Model 4:} \quad c = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad A = \begin{pmatrix} 0.1 & -0.2 \\ 0.5 & 1 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

The results for the empirical sizes are presented in Table 1. Both T_n^α and T_n achieve reasonable sizes and as n increases, the values become closer to the nominal levels. Note that this observation holds for larger values of tuning parameter α, and therefore, we can verify that DP divergence-based test works well in the case of no outlying observations. The empirical powers are calculated from Model 1 by adding parameter changes at the mid-point, $t = [n/2]$. The model parameters after the change are given as

$$\text{Model 1.1:} \quad c = \begin{pmatrix} 0.2 \\ -0.2 \end{pmatrix}, \quad A = \begin{pmatrix} 0.1 & -0.2 \\ 0.5 & 1 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

$$\text{Model 1.2:} \quad c = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad A = \begin{pmatrix} 0.2 & -0.2 \\ 0.5 & 0.9 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

$$\text{Model 1.3:} \quad c = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad A = \begin{pmatrix} 0.1 & -0.2 \\ 0.5 & 1 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1.5 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

$$\text{Model 1.4:} \quad c = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad A = \begin{pmatrix} 0.1 & -0.2 \\ 0.5 & 1 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}.$$

Table 2 summarizes the empirical powers. Observe that the score test T_n generally shows higher power than T_n^α when the sample size $n = 500$, but both tests are converging to 1 as sample size increases. In addition, the power of our proposed test based on T_n^α shows a decreasing tendency as the tuning parameter α increases, which is more pronounced when the sample size is small. However, this is what expected for MDPDE and this confirms that MDPDE with large α results in a loss of efficiency. Indeed, the simulations study with no outliers shows that our proposed DP divergence-based test achieves good sizes and reasonable powers although there is a power loss in small sample with a larger tuning parameter α.

However, such a loss of empirical power is compensated when the data are contaminated by outliers. To see this, we generated contaminated data $Y_{c,t} = (Y_{c,t,1}, Y_{c,t,2})^T$ by considering

$$Y_{c,t,i} = Y_{t,i} + s \cdot P_{t,i} \cdot \text{sign}(Y_{t,i}), \quad i = 1, 2,$$

where s is a constant, and $P_{t,i}$, $i = 1, 2$ are i.i.d. Bernoulli random variables with success probability $p/2$, and sign is a sign function given by $\text{sign}(x) = 1$ if $x \geq 0$ and -1 if $x < 0$. That is, this model randomly adds an atypical value of size s to the original data with the same direction of the t-th observation. Thus, s controls the outlier size and p determines the frequency of the outlying observations. Here, we consider three cases, $(p, s) = (0.01, 10)$, $(0.01, 20)$, and $(0.05, 20)$.

Tables 3, 4 and 5 report the empirical sizes while Tables 6, 7 and 8 show empirical powers. A couple of remarks are in order. First of all, Tables 4 and 5 show that the score test T_n shows serious size distortions when the size of the jump s is large ($s = 20$) and as more frequent outlying observations appear. This is in sharp contrast with the smaller jump size $s = 10$ with less frequent outlying observations $p = 0.01$, where the size distortions are only around 1–2% for the score test as in Table 3. However, our proposed DP divergence-based test shows good sizes for moderate to larger values of α, and in fact much smaller size distortions are observed even for smaller $\alpha = 0.1$. Second, while achieving good sizes, our proposed test also gains high power in the presence of outliers. It is more clearly visible when there is a huge and frequent outlying observations. For example, in Model 1.3 in Table 6 with

Table 1 Empirical sizes of T_n and T_n^α in the case of no outliers

θ	n	T_n		T_n^α									
				$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.5$		$\alpha = 1$	
		0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1
Model 1	500	0.029	0.071	0.032	0.074	0.033	0.074	0.036	0.074	0.036	0.066	0.030	0.067
	1000	0.038	0.085	0.037	0.084	0.042	0.083	0.043	0.084	0.040	0.088	0.040	0.078
	2000	0.041	0.089	0.044	0.087	0.045	0.085	0.046	0.087	0.042	0.087	0.042	0.088
Model 2	500	0.037	0.077	0.036	0.079	0.037	0.079	0.039	0.081	0.039	0.079	0.037	0.079
	1000	0.041	0.089	0.041	0.092	0.041	0.086	0.040	0.079	0.041	0.078	0.037	0.077
	2000	0.042	0.085	0.042	0.092	0.043	0.089	0.045	0.090	0.043	0.090	0.044	0.088
Model 3	500	0.032	0.071	0.035	0.069	0.036	0.072	0.038	0.072	0.035	0.076	0.033	0.080
	1000	0.043	0.086	0.043	0.088	0.046	0.087	0.047	0.085	0.045	0.091	0.039	0.089
	2000	0.046	0.091	0.045	0.094	0.047	0.095	0.046	0.095	0.044	0.095	0.049	0.097
Model 4	500	0.032	0.073	0.037	0.074	0.039	0.075	0.037	0.078	0.036	0.074	0.037	0.078
	1000	0.035	0.078	0.036	0.085	0.035	0.083	0.034	0.086	0.035	0.080	0.034	0.078
	2000	0.050	0.097	0.049	0.104	0.053	0.104	0.054	0.103	0.052	0.103	0.049	0.097

Table 2 Empirical powers of T_n and T_n^α in the case of no outliers

θ	n	T_n	T_n^α				
			$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 1$
Model 1.1	500	0.752	0.746	0.720	0.696	0.624	0.420
	1000	0.995	0.994	0.992	0.987	0.969	0.863
	2000	1.000	1.000	1.000	1.000	1.000	0.999
Model 1.2	500	0.438	0.432	0.420	0.400	0.333	0.209
	1000	0.894	0.885	0.869	0.847	0.780	0.589
	2000	1.000	1.000	1.000	1.000	0.995	0.945
Model 1.3	500	0.589	0.624	0.609	0.565	0.467	0.303
	1000	0.964	0.970	0.962	0.945	0.886	0.673
	2000	1.000	1.000	1.000	1.000	0.998	0.969
Model 1.4	500	0.558	0.627	0.625	0.591	0.503	0.302
	1000	0.951	0.977	0.975	0.963	0.914	0.655
	2000	1.000	1.000	1.000	1.000	1.000	0.965

$n = 2000$, the power of the score test is 0.157 while our DP divergence-based tests are close to 1 for all tuning parameters considered. Although the score test shows increasing power as the sample size increases, our DP divergence-based tests show much faster convergence to 1, even with larger α value. Indeed, the tuning parameter α plays a central role as it trades off sizes and powers in MDPDE; however, it shows less size distortions and higher power than the score test regardless of the α values considered. On the whole, our findings strongly support the assertion that the proposed test is a promising candidate for detecting parameter changes in VAR models when outliers are suspected to contaminate the data.

Next, we consider DFMs to test changes in parameter with possible contamination due to outlying observations. We considered DFMs with a dimension of $q = 300$ and two factors $r = 2$. The factor loadings $\Lambda = \lambda_{jk}$ are generated from i.i.d. $\mathcal{N}(0, 1)$ random variables, but are rotated and scaled to satisfy

$$\frac{1}{r}\Lambda^T \Lambda = I_q$$

using QR decomposition. The idiosyncratic component is assumed to follow the AR(1) model with $\rho = 0.5$,

$$\varepsilon_{j,t} = 0.5\varepsilon_{j,t-1} + \xi_{j,t},$$

where $\xi_{j,t}$'s are i.i.d. $\mathcal{N}(0, 3/4)$ so that the idiosyncratic component has zero mean and unit variance. To generate factors $\{Y_t\}_{t=1}^n$, we use the same model parameter in Model 1 and Model 1.2. That is, the empirical sizes are based on Model 1, and the empirical powers are generated from Model 1 for the first half, while Model 1.2 is used for the second half. The same settings are used to generate contamination by outliers as explained above.

Table 3 Empirical sizes of T_n and T_n^α under contamination with $p = 0.01$ and $s = 10$

θ	n	T_n		T_n^α									
				$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.5$		$\alpha = 1$	
		0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1
Model 1	500	0.052	0.105	0.034	0.056	0.033	0.068	0.032	0.067	0.033	0.066	0.032	0.070
	1000	0.065	0.111	0.037	0.072	0.043	0.088	0.043	0.096	0.045	0.087	0.049	0.090
	2000	0.070	0.119	0.034	0.083	0.041	0.083	0.043	0.089	0.045	0.087	0.049	0.095
Model 2	500	0.078	0.150	0.032	0.066	0.035	0.078	0.032	0.081	0.033	0.082	0.033	0.076
	1000	0.076	0.137	0.030	0.070	0.030	0.076	0.032	0.078	0.036	0.078	0.041	0.084
	2000	0.090	0.152	0.038	0.079	0.041	0.086	0.040	0.093	0.041	0.094	0.048	0.097
Model 3	500	0.053	0.107	0.050	0.106	0.040	0.084	0.039	0.080	0.037	0.082	0.037	0.082
	1000	0.069	0.131	0.060	0.107	0.048	0.093	0.042	0.091	0.044	0.088	0.047	0.089
	2000	0.089	0.153	0.054	0.113	0.041	0.093	0.038	0.090	0.045	0.092	0.048	0.099
Model 4	500	0.067	0.124	0.030	0.066	0.032	0.069	0.031	0.070	0.029	0.070	0.032	0.077
	1000	0.083	0.143	0.032	0.079	0.033	0.074	0.035	0.076	0.038	0.078	0.044	0.088
	2000	0.086	0.147	0.044	0.093	0.052	0.094	0.050	0.094	0.040	0.097	0.042	0.098

Table 4 Empirical sizes of T_n and T_n^α under contamination with $p = 0.01$ and $s = 20$

θ	n	T_n		T_n^α									
				$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.5$		$\alpha = 1$	
		0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1
Model 1	500	0.179	0.266	0.027	0.069	0.033	0.080	0.038	0.082	0.042	0.084	0.047	0.083
	1000	0.130	0.216	0.036	0.078	0.037	0.079	0.040	0.085	0.041	0.086	0.045	0.081
	2000	0.116	0.198	0.041	0.092	0.045	0.094	0.048	0.095	0.051	0.097	0.047	0.099
Model 2	500	0.289	0.386	0.037	0.074	0.038	0.078	0.040	0.077	0.037	0.078	0.034	0.074
	1000	0.286	0.377	0.051	0.096	0.052	0.106	0.054	0.104	0.049	0.099	0.040	0.096
	2000	0.269	0.360	0.037	0.086	0.036	0.082	0.036	0.082	0.035	0.090	0.047	0.093
Model 3	500	0.097	0.160	0.039	0.083	0.038	0.083	0.040	0.083	0.041	0.085	0.040	0.084
	1000	0.087	0.156	0.055	0.102	0.051	0.108	0.051	0.105	0.051	0.103	0.047	0.091
	2000	0.081	0.152	0.040	0.094	0.045	0.093	0.048	0.095	0.047	0.098	0.045	0.097
Model 4	500	0.220	0.316	0.037	0.068	0.037	0.076	0.039	0.079	0.039	0.085	0.038	0.083
	1000	0.176	0.256	0.033	0.073	0.038	0.075	0.040	0.082	0.041	0.083	0.040	0.086
	2000	0.149	0.233	0.045	0.086	0.047	0.085	0.047	0.087	0.048	0.089	0.048	0.093

Table 5 Empirical sizes of T_n and T_n^α under contamination with $p = 0.05$ and $s = 20$

θ	n	T_n		T_n^α									
				$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.5$		$\alpha = 1$	
		0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1
Model 1	500	0.202	0.290	0.025	0.051	0.026	0.067	0.028	0.070	0.029	0.081	0.031	0.080
	1000	0.205	0.298	0.026	0.066	0.039	0.082	0.040	0.083	0.042	0.082	0.046	0.093
	2000	0.205	0.307	0.028	0.063	0.035	0.077	0.034	0.080	0.031	0.088	0.038	0.088
Model 2	500	0.419	0.507	0.021	0.055	0.027	0.072	0.027	0.073	0.032	0.076	0.036	0.076
	1000	0.404	0.514	0.020	0.056	0.036	0.075	0.041	0.078	0.039	0.082	0.039	0.084
	2000	0.429	0.527	0.027	0.061	0.043	0.086	0.045	0.084	0.045	0.091	0.046	0.087
Model 3	500	0.409	0.496	0.077	0.135	0.035	0.082	0.035	0.081	0.034	0.074	0.032	0.078
	1000	0.522	0.614	0.084	0.142	0.041	0.083	0.038	0.086	0.041	0.082	0.046	0.090
	2000	0.647	0.703	0.090	0.159	0.045	0.094	0.046	0.093	0.050	0.097	0.050	0.098
Model 4	500	0.263	0.369	0.030	0.063	0.037	0.071	0.038	0.076	0.042	0.078	0.041	0.082
	1000	0.259	0.369	0.039	0.084	0.034	0.071	0.030	0.079	0.034	0.075	0.037	0.079
	2000	0.284	0.391	0.046	0.093	0.041	0.089	0.044	0.084	0.044	0.084	0.044	0.090

Table 6 Empirical powers of T_n and T_n^α under contamination with $p = 0.01$ and $s = 10$

θ	n	T_n	T_n^α				
			$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 1$
Model 1.1	500	0.463	0.663	0.681	0.669	0.607	0.435
	1000	0.798	0.983	0.986	0.983	0.963	0.835
	2000	0.988	1.000	1.000	1.000	1.000	0.997
Model 1.2	500	0.184	0.351	0.388	0.380	0.332	0.220
	1000	0.330	0.818	0.839	0.826	0.770	0.585
	2000	0.626	0.997	0.998	0.998	0.993	0.940
Model 1.3	500	0.110	0.611	0.606	0.569	0.473	0.295
	1000	0.086	0.969	0.959	0.939	0.873	0.658
	2000	0.157	1.000	1.000	1.000	0.998	0.969
Model 1.4	500	0.222	0.610	0.617	0.592	0.495	0.268
	1000	0.398	0.964	0.962	0.951	0.895	0.657
	2000	0.707	1.000	1.000	1.000	1.000	0.968

Table 7 Empirical powers of T_n and T_n^α under contamination with $p = 0.01$ and $s = 20$

θ	n	T_n	T_n^α				
			$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 1$
Model 1.1	500	0.471	0.706	0.711	0.684	0.617	0.419
	1000	0.615	0.991	0.991	0.988	0.969	0.865
	2000	0.830	1.000	1.000	1.000	1.000	0.997
Model 1.2	500	0.313	0.398	0.415	0.396	0.335	0.227
	1000	0.333	0.833	0.854	0.825	0.750	0.549
	2000	0.495	0.999	0.998	0.995	0.990	0.944
Model 1.3	500	0.214	0.632	0.619	0.578	0.489	0.298
	1000	0.146	0.960	0.957	0.934	0.879	0.686
	2000	0.126	1.000	1.000	1.000	0.999	0.965
Model 1.4	500	0.271	0.606	0.621	0.600	0.496	0.299
	1000	0.253	0.970	0.964	0.951	0.897	0.651
	2000	0.296	1.000	1.000	1.000	1.000	0.965

Table 9 shows the empirical sizes of the DFM model. When there are no outliers, all tests seem to achieve the correct sizes; however, the score test T_n starts distorting when there are outlying observations. Note also that the size distortion of T_n appears for larger values of $s \geq 30$ compared to the usual VAR(1) model. This is consistent with Stock and Watson (2002), who showed that the estimated factors are still consistent if the size of break is sufficiently small. If we increase the contamination size, then the size distortions are more evident and the usual score test shows severe size distortions. However, our proposed method T_n^α correct sizes in all cases considered as expected. In terms of the empirical powers reported in Table 10, we observe that

Table 8 Empirical powers of T_n and T_n^α under contamination with $p = 0.05$ and $s = 20$

θ	n	T_n	T_n^α				
			$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 1$
Model 1.1	500	0.323	0.416	0.663	0.651	0.585	0.414
	1000	0.440	0.841	0.978	0.975	0.947	0.851
	2000	0.652	0.999	1.000	1.000	1.000	0.998
Model 1.2	500	0.244	0.131	0.350	0.350	0.309	0.222
	1000	0.335	0.316	0.803	0.797	0.727	0.548
	2000	0.505	0.785	0.995	0.995	0.986	0.933
Model 1.3	500	0.198	0.545	0.553	0.527	0.437	0.287
	1000	0.204	0.938	0.931	0.903	0.830	0.645
	2000	0.218	1.000	1.000	1.000	0.997	0.956
Model 1.4	500	0.186	0.443	0.552	0.538	0.462	0.293
	1000	0.218	0.878	0.944	0.923	0.873	0.658
	2000	0.239	0.998	1.000	1.000	0.996	0.958

our proposed test outperforms the score test T_n even when the data are contaminated and are quite comparable even without the outliers if the control parameter α is moderate. Again, high efficiency is observed when the jump size is large and frequent since the power converges to 1 more rapidly as the sample size increases. All in all, our simulation study shows that our proposed method based on MDPDE achieves good sizes and powers when the data are contaminated by some outliers.

5 Real applications

In statistical terms, the functional connectivity of the human brain refers to the temporal correlation between the brain regions of interest (ROI) calculated from the fMRI series. It provides important insights on the functional communication in the brain network, and the resting state fMRI (rsfMRI) data, in particular, received lots of attention in neuroscience because many studies have shown that our brain network at rest is not idle, but instead shows spontaneous connectivity changes over time (e.g., Van Den Heuvel and Pol 2010). However, the acquisition of fMRI data involves numerous challenges such as anatomical alignment of images and head motion adjustment, in turn outlying observations cannot be avoided. Furthermore, it is well-documented that brain connectivity is highly susceptible to the influence of outlying observations, hence robust inference should be used in the analysis to correctly identify change points (e.g., Poldrack 2012; Magnotti and Billor 2014). Here, we illustrate our proposed robust procedures to detect functional connectivity changes in rsfMRI data using data from the Human Connectome Project (HCP) available at www.humanconnectome.org. The data set comprises time series on 300 ROIs, and each of the 812 subjects completed four cycles of rsfMRI runs. The following provides an illustration of the results for one subject selected at random on the second run, where the data have

Table 9 Empirical sizes of T_n and T_n^{α} under DFM

Model 1	n	T_n		T_n^{α}									
				$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.5$		$\alpha = 1$	
		0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1
No outliers	500	0.027	0.067	0.033	0.074	0.037	0.073	0.036	0.076	0.043	0.081	0.041	0.081
	1000	0.038	0.090	0.046	0.091	0.046	0.090	0.045	0.087	0.039	0.093	0.042	0.086
	2000	0.042	0.085	0.040	0.083	0.037	0.087	0.056	0.095	0.048	0.091	0.048	0.084
$p = 0.01, s = 10$	500	0.024	0.068	0.033	0.083	0.032	0.083	0.038	0.083	0.030	0.082	0.046	0.083
	1000	0.041	0.080	0.036	0.084	0.037	0.083	0.040	0.087	0.042	0.085	0.042	0.085
	2000	0.054	0.113	0.045	0.091	0.043	0.089	0.052	0.088	0.046	0.095	0.047	0.092
$p = 0.01, s = 20$	500	0.056	0.118	0.024	0.061	0.027	0.063	0.030	0.063	0.027	0.066	0.033	0.068
	1000	0.052	0.112	0.038	0.086	0.040	0.086	0.038	0.088	0.041	0.094	0.046	0.093
	2000	0.056	0.116	0.036	0.081	0.040	0.078	0.041	0.087	0.043	0.092	0.046	0.096
$p = 0.01, s = 30$	500	0.161	0.227	0.038	0.077	0.040	0.092	0.040	0.076	0.039	0.079	0.038	0.092
	1000	0.117	0.186	0.037	0.086	0.042	0.094	0.044	0.092	0.052	0.092	0.045	0.095
	2000	0.096	0.161	0.050	0.101	0.046	0.097	0.045	0.096	0.042	0.098	0.036	0.089
$p = 0.05, s = 40$	500	0.240	0.342	0.033	0.073	0.033	0.082	0.032	0.077	0.033	0.082	0.036	0.082
	1000	0.246	0.336	0.041	0.087	0.043	0.090	0.045	0.086	0.042	0.091	0.043	0.093
	2000	0.252	0.351	0.041	0.088	0.045	0.093	0.043	0.098	0.046	0.093	0.043	0.093

Table 10 Empirical powers of T_n and T_n^α under DFM

Model 1 to Model 1.2	n	T_n	T_n^α				
			$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 1$
No outliers	500	0.451	0.456	0.418	0.412	0.350	0.223
	1000	0.876	0.874	0.865	0.843	0.774	0.554
	2000	0.998	0.998	0.997	0.995	0.988	0.934
$p = 0.01, s = 10$	500	0.280	0.425	0.413	0.387	0.336	0.221
	1000	0.603	0.868	0.861	0.836	0.767	0.550
	2000	0.926	0.999	0.998	0.9995	0.990	0.925
$p = 0.01, s = 20$	500	0.266	0.412	0.405	0.385	0.337	0.223
	1000	0.340	0.868	0.853	0.832	0.770	0.576
	2000	0.486	0.998	0.997	0.996	0.992	0.943
$p = 0.01, s = 30$	500	0.363	0.417	0.408	0.387	0.340	0.229
	1000	0.310	0.863	0.856	0.830	0.763	0.552
	2000	0.302	0.998	0.997	0.996	0.992	0.935
$p = 0.05, s = 40$	500	0.243	0.365	0.380	0.366	0.330	0.220
	1000	0.239	0.836	0.846	0.824	0.751	0.556
	2000	0.351	0.993	0.999	0.995	0.988	0.928

a dimension of 300×1200 since similar conclusions were drawn for other subjects and runs.

Figure 1 shows the time plot of the rsfMRI series for the four ROIs selected at random. There are clear indications of changes in connectivity. For example, there are sudden level changes around 750–800 for ROI 1 and ROI 25, and the variability appears to be different from 800 to 1000 in ROI 1 etc. In addition, there are noticeable spikes throughout the time plots; hence, it is not clear whether such changes are due to changes in brain connectivity or an artifact of outlying observations. Indeed, the time plot shows the typical features widely reported in fMRI data, namely, exhibiting both parameter instabilities and spikes. Therefore, we need to apply a robust testing method to detect connectivity changes for correct inference instead of ignoring the outliers. By assuming that the rsfMRI series follows DFM, the connectivity changes are equivalent to changes in loading matrices at certain unknown points. As discussed in Sect. 2, it can be further reduced to detect parameter changes on principal factors. Hence, we begin by estimating the number of factors to be used in the analysis. The IC criteria of Bai and Ng (2002) with penalty function $(q+n)/(qn)\log(qn/(q+n))$ and $(q+n)/(qn)\log(\min(q,n))$ selects the number of factors $\hat{r} = 5$. Figure 5 shows the estimated five principal factors along with the estimated change-points to be explained in the section below.

Next, we need to determine the order of VAR(p) model in DFM. Figure 2 shows the sample partial autocorrelation function of the first factor (left) and the sample cross-correlation between first and second factors (middle), and the third and fourth factors (right) after prewhitening using AR(1) model. This suggests that the rsfMRI series has non-negligible cross-correlations; hence, a univariate approach to handle

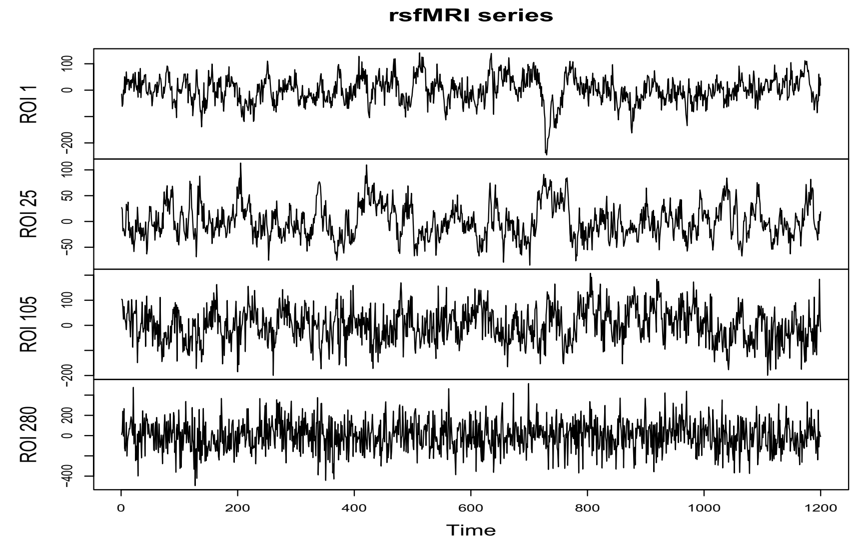

Fig. 1 Time plot for rsfMRI series

outlying observations is not suitable but joint modeling with VAR(1) is plausible. The BIC selection for the VAR order also coincides with our observations, so we perform our procedure based on the robust VAR(1) parameter estimation using MDPDE.

We have used $\alpha = 0, 0.1, 0.2, 0.3, 0.4$ and applied BS to identify multiple change points. Figure 3 illustrates the BS procedure with $\alpha = 0$. The black solid line shows the test statistics $T^{\alpha}_{n,k}$ using whole data, and the maximum value attained at $\widehat{k}_1 = 557$ is selected as the first change point since it is above the critical value 15.65. Then, the sample is split into two subsamples, before and after the first break, and the same procedure is applied. This gives $\widehat{k}_2 = 293$ and $\widehat{k}_3 = 890$ as change points. We repeat this procedure until no further change points are declared. Figure 4 shows the estimated change points according to the different tuning parameter α values. The number of change points decrease as α increases, while the change points around 292, 435 and 882 are common regardless of the α values. We take the average of the

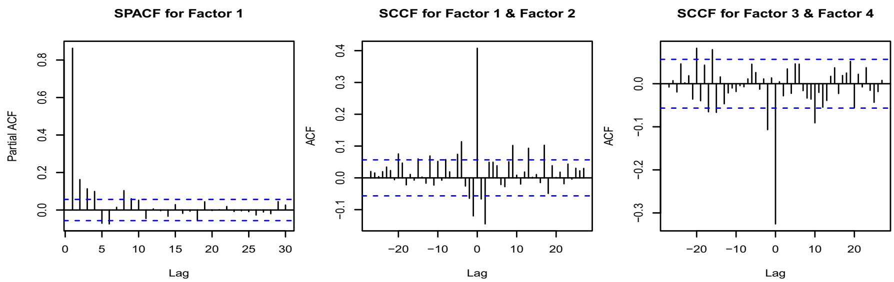

Fig. 2 The sample correlations plots for rsfMRI series

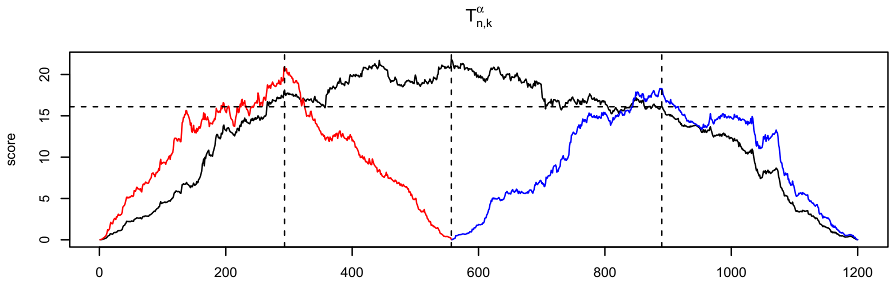

Fig. 3 Test statistics T_n^{α} and the illustration of binary segmentation with $\alpha = 0$

estimated change points to produce the final change points (see Remark 8). This also shows that the above three change points are indeed due to parameter instabilities, while the other identified change points are not as clear as them. Figure 5 shows the time plot with three common change points in vertical red lines, and Fig. 6 represents the estimated VAR(1) coefficients and estimated innovation covariance matrices $\widehat{\Sigma}$ on four regimes when $\alpha = 0.3$. We observe that both the VAR coefficients and innovation covariance matrices change over time, and the association can also change in opposite signs. All in all, this observations add one more evidence that our brain connectivity indeed changes even in the resting state and such changes are not an artifact of outlier effects (e.g., Power et al. 2015).

6 Conclusions

This paper proposed a robust test for parameter changes in a vector autoregression model, which can be adapted to DFMs for HDTS once principal factors are used. The theoretical results substantiated the validity of our method and the Monte Carlo simulations study demonstrated the outstanding performance of our method in finite samples. We also applied our method to the resting state fMRI series that typically show both connectivity changes and atypical observations. Our empirical analysis shows that brain connectivity changes even in the resting state and such changes are not an artifact of outlying observations. Our method incorporates outlier effects in the HDTS context while computationally tractable since dimension reduction is accomplished by considering several latent principal factors.

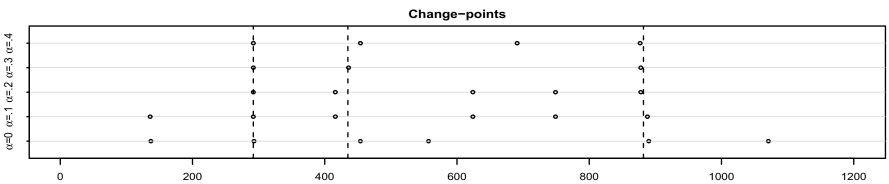

Fig. 4 Estimated change-points for $\alpha = 0, 0.1, 0.2, 0.3, 0.4$

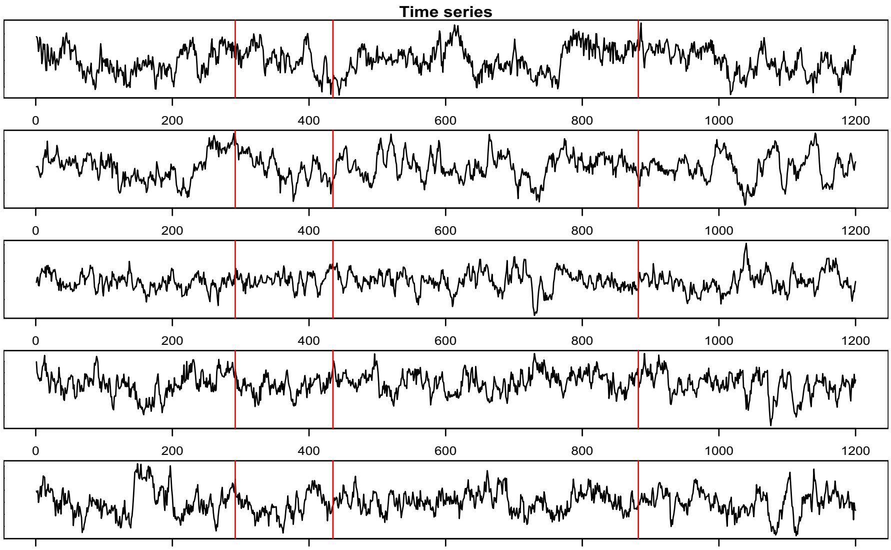

Fig. 5 Time plots of five estimated principal factors with common change points in vertical solid lines

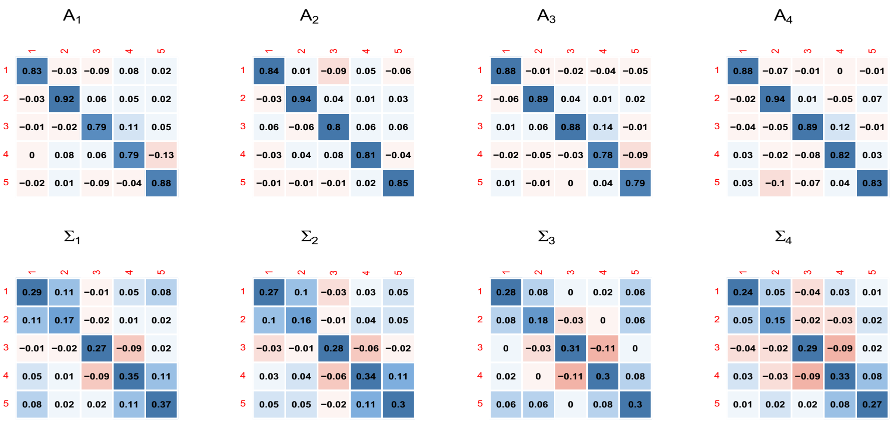

Fig. 6 Estimated VAR(1) coefficients (top) and estimated innovation covariance matrix (bottom) on each regime when $\alpha = 0.3$

7 Proofs

We provide the proofs of Theorems 1–3 for the case of $\alpha > 0$. In what follows, the symbol $\|\cdot\|$ denotes the Euclidean norm for vectors and the spectral norm for matrices.

Before proving Theorems 1–3, we prepare the following lemmas (the proofs are postponed to "Appendix").

Lemma 1 *Suppose that assumptions* **(A1)** *and* **(A2)** *hold. Then, we have that for* $t \geq p + 1$,

$$E\left(\sup_{\theta\in\Theta}\left\|\frac{\partial^2 h_{\alpha,t}(\theta)}{\partial\theta\partial\theta^T}\right\|\right) < \infty \quad \text{and} \quad E\left(\sup_{\theta\in\Theta}\left\|\frac{\partial h_{\alpha,t}(\theta)}{\partial\theta}\frac{\partial h_{\alpha,t}(\theta)}{\partial\theta^T}\right\|\right) < \infty.$$

Lemma 2 *Suppose that assumptions* **(A1)** *and* **(A2)** *hold. Then,* J_α *and* K_α *are non-singular.*

Lemma 3 *Suppose that assumptions* **(A1)** *and* **(A2)** *hold. Then, under* H_0', *we have*

$$K_\alpha^{-1/2}\frac{[ns]}{\sqrt{n}}\frac{\partial H_{\alpha,[ns]}(\theta_0)}{\partial\theta}\xrightarrow{w}B_\eta(s) \;\; in \;\; \mathbb{D}([0,1],\mathbb{R}^\eta),$$

where B_η *is a* η*-dimensional standard Brownian motion.*

Lemma 4 *Suppose that assumptions* **(A1)–(A3)** *hold. Then, under* H_0', *we have*

$$\max_{1\leq k\leq n}\frac{k}{n}\left\|\frac{\partial^2 H_{\alpha,k}(\bar{\theta}_{\alpha,n,k})}{\partial\theta\partial\theta^T}+J_\alpha\right\| = o(1) \quad a.s.,$$

where $\{\bar{\theta}_{\alpha,n,k} \mid 1 \leq k \leq n, n \geq 1\}$ *is any double array of* Θ*-valued random vectors with* $\|\bar{\theta}_{\alpha,n,k} - \theta_0\| \leq \|\hat{\theta}_{\alpha,n} - \theta_0\|$.

Lemma 5 *Suppose that assumptions* **(A1)–(A3)** *hold. Then, under* H_0', *we have*

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial h_{\alpha,t}(\hat{\theta}_{\alpha,n})}{\partial\theta}\frac{\partial h_{\alpha,t}(\hat{\theta}_{\alpha,n})}{\partial\theta^T}\xrightarrow{P}E\left(\frac{\partial h_{\alpha,t}(\theta_0)}{\partial\theta}\frac{\partial h_{\alpha,t}(\theta_0)}{\partial\theta^T}\right).$$

Proof of Theorem 1 Since $\{h_{\alpha,t}(\theta)|t \geq p+1\}$ is stationary and ergodic, and $E\big(\sup_{\theta\in\Theta}|h_{\alpha,t}(\theta)|\big) < \infty$ by Lemma S1 of the Supplementary Material, it holds that

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n}h_{\alpha,t}(\theta) - E(h_{\alpha,t}(\theta))\right|\xrightarrow{a.s.}0 \quad \text{as} \quad n\to\infty.$$

Further, it can be shown that

$$\begin{aligned}E(h_{\alpha,t}(\theta)) - E(h_{\alpha,t}(\theta_0)) &= E\big[E\big\{h_{\alpha,t}(\theta) - h_{\alpha,t}(\theta_0)|Y_{t-p:t-1}\big\}\big]\\ &= E\left[\int\left\{f_\theta^{1+\alpha}(y|Y_{t-p:t-1}) - \left(1+\frac{1}{\alpha}\right)f_\theta^\alpha(y|Y_{t-p:t-1})f_{\theta_0}(y|Y_{t-p:t-1})\right.\right.\\ &\left.\left.+\frac{1}{\alpha}f_{\theta_0}^{1+\alpha}(y|Y_{t-p:t-1})\right\}\mathrm{d}y\right] \geq 0,\end{aligned}$$

where the equality holds if and only if $\theta = \theta_0$. Thus, if $\theta \neq \theta_0$, then $E(h_{\alpha,t}(\theta)) > E(h_{\alpha,t}(\theta_0))$, which means that $E(h_{\alpha,t}(\theta))$ has a unique minimum at θ_0. Therefore, the theorem is established. □

Proof of Theorem 2 By the mean value theorem, we have

$$0 = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial h_{\alpha,t}(\theta_0)}{\partial \theta} + \left(\frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 h_{\alpha,t}(\theta^*_{\alpha,n})}{\partial \theta \partial \theta^T} \right) \sqrt{n} \big(\hat{\theta}_{\alpha,n} - \theta_0 \big),$$

where $\theta^*_{\alpha,n}$ is an intermediate point between θ_0 and $\hat{\theta}_{\alpha,n}$. By Lemma 3 with $s = 1$, we have

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial h_{\alpha,t}(\theta_0)}{\partial \theta} \xrightarrow{d} \mathcal{N}(0, K_\alpha) \quad \text{as} \quad n \to \infty. \tag{9}$$

Due to Lemma 4, we also have

$$\frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 h_{\alpha,t}(\theta^*_{\alpha,n})}{\partial \theta \partial \theta^T} \xrightarrow{a.s.} -J_\alpha \quad \text{as} \quad n \to \infty,$$

which together with (9) validates the theorem. □

Proof of Theorem 3 We first show that

$$\frac{[ns]}{\sqrt{n}} \frac{\partial H_{\alpha,[ns]}(\theta_0)}{\partial \theta} + \frac{[ns]}{n} \frac{\partial^2 H_{\alpha,[ns]}(\theta^*_{\alpha,n,s})}{\partial \theta \partial \theta^T} J_\alpha^{-1} \sqrt{n} \frac{\partial H_{\alpha,n}(\theta_0)}{\partial \theta} \xrightarrow{w} K_\alpha^{1/2} B^o_\eta(s). \tag{10}$$

Owing to Lemma 3, we have

$$\frac{[ns]}{\sqrt{n}} \frac{\partial H_{\alpha,[ns]}(\theta_0)}{\partial \theta} - \frac{[ns]}{n} \sqrt{n} \frac{\partial H_{\alpha,n}(\theta_0)}{\partial \theta} \xrightarrow{w} K_\alpha^{1/2} B^o_\eta(s) \quad \text{in } \mathbb{D}\left([0,1], \mathbb{R}^\eta\right).$$

Since $\sqrt{n}\, \partial H_{\alpha,n}(\theta_0) = O_P(1)$, it follows from Lemma 4 that

$$\begin{aligned} &\sup_{0 \le s \le 1} \frac{[ns]}{n} \left\| \frac{\partial^2 H_{\alpha,[ns]}(\theta^*_{\alpha,n,s})}{\partial \theta \partial \theta^T} J_\alpha^{-1} \sqrt{n} \frac{\partial H_{\alpha,n}(\theta_0)}{\partial \theta} + \sqrt{n} \frac{\partial H_{\alpha,n}(\theta_0)}{\partial \theta} \right\| \\ &\le \left\| J_\alpha^{-1} \sqrt{n} \frac{\partial H_{\alpha,n}(\theta_0)}{\partial \theta} \right\| \max_{1 \le k \le n} \frac{k}{n} \left\| \frac{\partial^2 H_{\alpha,k}(\theta^*_{\alpha,n,k})}{\partial \theta \partial \theta^T} + J_\alpha \right\| \\ &= o_P(1), \end{aligned}$$

where $\theta^*_{\alpha,n,k}$ denotes the one corresponding to $\theta^*_{\alpha,n,s}$ when $[ns] = k$. Hence, (10) holds.

Next, note that by Lemma 4,

$$\sup_{0 \le s \le 1} \frac{[ns]}{n} \left\| \frac{\partial^2 H_{\alpha,[ns]}(\theta^*_{\alpha,n,s})}{\partial \theta \partial \theta^T} \right\| \le \max_{1 \le k \le n} \frac{k}{n} \left\| \frac{\partial^2 H_{\alpha,k}(\theta^*_{\alpha,n,k})}{\partial \theta \partial \theta^T} + J_\alpha \right\| + \|J_\alpha\| = O_P(1)$$

and

$$\left\|B_{\alpha,n}+J_\alpha\right\| \le \max_{1\le k\le n}\frac{k}{n}\left\|\frac{\partial^2 H_{\alpha,k}(\theta^*_{\alpha,n,k})}{\partial\theta\partial\theta^T}+J_\alpha\right\| = o(1) \quad a.s.$$

Then, since $\sqrt{n}(\hat{\theta}_{\alpha,n}-\theta_0)=O_P(1)$ by Theorem 2, we have

$$\sup_{0\le s\le 1}\frac{[ns]}{n}\left\|\frac{\partial^2 H_{\alpha,[ns]}(\theta^*_{\alpha,n,s})}{\partial\theta\partial\theta^T}\left(B_{\alpha,n}+J_\alpha\right)\sqrt{n}\left(\hat{\theta}_{\alpha,n}-\theta_0\right)\right\| = o_P(1). \tag{11}$$

Therefore, the theorem is established from (5), (10), and (11). □

Acknowledgements R code is available at https://github.com/crbaek/dpd_VAR. The data were provided in part by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. We would like to thank the Associate Editor and reviewers for many useful comments and suggestions.

Appendix

Proof of Lemma 1 Under **(A1)**, there exist a constant mean vector μ and a covariance matrix Γ_0 of Y_t such that $\mu=E(Y_t)$ and $\Gamma_0=E(Y_tY_t^T)-\mu\mu^T$ for all t, which implies $E(Y_{t,i}^2)<\infty$ for all t and $i=1,\dots,r$. Hence, by Lemma S1 of the Supplementary Material and Cauchy–Schwartz inequality, we obtain

$$E\left(\sup_{\theta\in\Theta}\left|\frac{\partial^2 h_{\alpha,t}(\theta)}{\partial\theta_i\partial\theta_j}\right|\right)<\infty$$

and

$$E\left(\sup_{\theta\in\Theta}\left|\frac{\partial h_{\alpha,t}(\theta)}{\partial\theta_i}\frac{\partial h_{\alpha,t}(\theta)}{\partial\theta_j}\right|\right)\le E\left(\sup_{\theta\in\Theta}\left|\frac{\partial h_{\alpha,t}(\theta)}{\partial\theta_i}\right|\sup_{\theta\in\Theta}\left|\frac{\partial h_{\alpha,t}(\theta)}{\partial\theta_j}\right|\right)<\infty$$

for $i,j=1,\dots,\eta$. Therefore, the lemma is validated. □

Proof of Lemma 2 Due to the first part of Lemma 1, J_α is finite. Note that

$$\begin{aligned}E\left(\frac{\partial^2 h_{\alpha,t}(\theta_0)}{\partial\theta\partial\theta^T}\right) =& E\left\{E\left(\frac{\partial^2 h_{\alpha,t}(\theta_0)}{\partial\theta\partial\theta^T}\Big|Y_{t-p:t-1}\right)\right\}\\ =&(1+\alpha)E\left\{\int f_{\theta_0}^{\alpha-1}(y|Y_{t-p:t-1})\frac{\partial f_{\theta_0}(y|Y_{t-p:t-1})}{\partial\theta}\frac{\partial f_{\theta_0}(y|Y_{t-p:t-1})}{\partial\theta^T}\mathrm{d}y\right\}.\end{aligned}$$

Letting A be a $r(r+1)/2 \times r(r+1)/2$ diagonal matrix with $A_{ii} = 1$ if $i = (m-1)(r+1-m/2)+1$ for $m = 1, \dots, r$ and 2 otherwise, we can write that for $v = (v_1^T, v_2^T, v_3^T)^T \in \mathbb{R}^r \times \mathbb{R}^{r^2 p} \times \mathbb{R}^{r(r+1)/2}$,

$$\begin{aligned}
& v^T \frac{\partial f_{\theta_0}(Y_t | Y_{t-p:t-1})}{\partial \theta} \\
&= v_1^T \frac{\partial f_{\theta_0}(Y_t | Y_{t-p:t-1})}{\partial c} + v_2^T \frac{\partial f_{\theta_0}(Y_t | Y_{t-p:t-1})}{\partial vec(A_1, \dots, A_p)} + v_3^T \frac{\partial f_{\theta_0}(Y_t | Y_{t-p:t-1})}{\partial vech(\Sigma)} \\
&= f_{\theta_0}(Y_t | Y_{t-p:t-1}) \Bigg[v_1^T \Sigma_0^{-1} \epsilon_t(\theta_0) + v_2^T \Big\{ (Y_{t-1}^T, Y_{t-2}^T, \dots, Y_{t-p}^T)^T \otimes \Sigma_0^{-1} \epsilon_t(\theta_0) \Big\} \\
&\quad - \frac{1}{2} v_3^T A vech \big\{ \Sigma_0^{-1} - \Sigma_0^{-1} \epsilon_t(\theta_0) \epsilon_t(\theta_0)^T \Sigma_0^{-1} \big\} \Bigg] \\
&:= f_{\theta_0}(Y_t | Y_{t-p:t-1}) M_t(v, \theta_0),
\end{aligned}$$

where $\epsilon_t(\theta_0) = Y_t - c_0 - A_{01} Y_{t-1} - \cdots - A_{0p} Y_{t-p} \sim N(0, \Sigma_0)$ and the symbol $\otimes$ stand for Kronecker product. Thus, we have

$$\begin{aligned}
v^T(-J_\alpha)v &= (1+\alpha) E \left\{ \int f_{\theta_0}^{\alpha-1}(y | Y_{t-p:t-1}) \left(v^T \frac{\partial f_{\theta_0}(y | Y_{t-p:t-1})}{\partial \theta} \right)^2 \mathrm{d}y \right\} \\
&= (1+\alpha) E \left\{ f_{\theta_0}^{\alpha-2}(Y_t | Y_{t-p:t-1}) \left(v^T \frac{\partial f_{\theta_0}(Y_t | Y_{t-p:t-1})}{\partial \theta} \right)^2 \right\} \\
&= (1+\alpha) E \Big\{ f_{\theta_0}^{\alpha}(Y_t | Y_{t-p:t-1}) M_t(v, \theta_0)^2 \Big\} \geq 0.
\end{aligned}$$

Since $M_t(v, \theta_0)$ has a continuous distribution induced from Gaussian random variables, one can see that the equality holds only for $v = 0$. Hence, J_α is a non-singular matrix. The non-singularity of K_α can be shown by similar arguments, so we omit the proof for K_α. □

Proof of Lemma 3 Let $\{Y_t | t \in \mathbb{Z}\}$ be the strictly stationary and ergodic solution to VAR(p) model (3). Recall that $\{h_{\alpha,t}(\theta) | t = 1, \dots, n\}$ is calculated from the observations $Y_1, \dots, Y_n$ and some initial values for $Y_{-p}, \dots, Y_0$. To verify the lemma, we introduce stationary version of $\{h_{\alpha,t}(\theta) | t = 1, \dots, n\}$. Let $\{h_{\alpha,t}^o(\theta) | t \in \mathbb{Z}\}$ and $\{H_{\alpha,t}^o(\theta) | t \in \mathbb{Z}\}$ be the counterparts of $\{h_{\alpha,t}(\theta)\}$ and $\{H_{\alpha,t}(\theta)\}$, respectively, obtained by using the solution $\{Y_t | t \in \mathbb{Z}\}$. Then, $\{\partial h_{\alpha,t}^o(\theta_0)/\partial \theta\}$ is a strictly stationary and ergodic process. Further, noting that

$$\begin{aligned}
& E\left(f_{\theta_0}^{\alpha-1}(Y_t | Y_{t-p:t-1}) \frac{\partial f_{\theta_0}(Y_t | Y_{t-p:t-1})}{\partial \theta} \middle| \mathcal{F}_{t-1} \right) \\
&= \int f_{\theta_0}^{\alpha}(y | Y_{t-p:t-1}) \frac{\partial f_{\theta_0}(y | Y_{t-p:t-1})}{\partial \theta} \mathrm{d}y,
\end{aligned}$$

where $\mathcal{F}_{t-1}$ be the sigma field generated by $\{Y_{t-1}, Y_{t-2}, \dots\}$, we have

$$E\left(\frac{\partial h^o_{\alpha,t}(\theta_0)}{\partial\theta}\Big|\mathcal{F}_{t-1}\right) = (1+\alpha)E\left(\int f^{\alpha}_{\theta_0}(y|Y_{t-p:t-1})\frac{\partial f_{\theta_0}(y|Y_{t-p:t-1})}{\partial\theta}\mathrm{d}y \right.$$
$$\left. - f^{\alpha-1}_{\theta_0}(Y_t|Y_{t-p:t-1})\frac{\partial f_{\theta_0}(Y_t|Y_{t-p:t-1})}{\partial\theta}\Big|\mathcal{F}_{t-1}\right) = 0.$$

Hence, it follows from the functional limit theorem for martingales (cf. Section 18 in Billingsley 1999) that

$$\frac{[ns]}{\sqrt{n}}\frac{\partial H^o_{\alpha,[ns]}(\theta_0)}{\partial\theta} = \frac{1}{\sqrt{n}}\sum_{t=1}^{[ns]}\frac{\partial h^o_{\alpha,t}(\theta_0)}{\partial\theta}\xrightarrow{w} K^{1/2}_{\alpha}B_{\eta}(s) \ \ in \ \ \mathbb{D}([0,1],\mathbb{R}^{\eta}).$$

Since $h^o_{\alpha,t}(\theta_0) = h_{\alpha,t}(\theta_0)$ for $t \geq p+1$, one can also see that

$$\sup_{0\leq s\leq 1}\frac{[ns]}{\sqrt{n}}\left\|\frac{\partial H^o_{\alpha,[ns]}(\theta_0)}{\partial\theta} - \frac{\partial H_{\alpha,[ns]}(\theta_0)}{\partial\theta}\right\|$$
$$\leq \frac{1}{\sqrt{n}}\sum_{t=1}^{p}\left\|\frac{\partial h^o_{\alpha,t}(\theta_0)}{\partial\theta} - \frac{\partial h_{\alpha,t}(\theta_0)}{\partial\theta}\right\| = o(1) \quad a.s.,$$

which establishes the lemma. □

Proof of Lemma 4 By Lemma 1, we have

$$E\left(\sup_{\theta\in\Theta}\left\|\frac{\partial^2 h_{\alpha,t}(\theta)}{\partial\theta\partial\theta^T} - \frac{\partial^2 h_{\alpha,t}(\theta_0)}{\partial\theta\partial\theta^T}\right\|\right) < \infty.$$

Since $\partial^2 h_{\alpha,t}(\theta)/\partial\theta\partial\theta^T$ is continuous in θ, for any $\epsilon > 0$, one can take a neighborhood $N_\epsilon(\theta_0)$ such that

$$E\left(\sup_{\theta\in N_\epsilon(\theta_0)}\left\|\frac{\partial^2 h_{\alpha,t}(\theta)}{\partial\theta\partial\theta^T} - \frac{\partial^2 h_{\alpha,t}(\theta_0)}{\partial\theta\partial\theta^T}\right\|\right) < \epsilon \qquad (12)$$

by decreasing the neighborhood to the singleton θ_0. Noting that $\hat{\theta}_{\alpha,n}$ converges almost surely to θ_0, we have, for sufficiently large n,

$$\max_{1\leq k\leq n}\frac{k}{n}\left\|\frac{\partial^2 H_{\alpha,k}(\bar{\theta}_{\alpha,n,k})}{\partial\theta\partial\theta^T} + J_\alpha\right\|$$
$$\leq \max_{1\leq k\leq n}\frac{k}{n}\left\|\frac{\partial^2 H_{\alpha,k}(\bar{\theta}_{\alpha,n,k})}{\partial\theta\partial\theta^T} - \frac{\partial^2 H_{\alpha,k}(\theta_0)}{\partial\theta\partial\theta^T}\right\| + \max_{1\leq k\leq n}\frac{k}{n}\left\|\frac{\partial^2 H_{\alpha,k}(\theta_0)}{\partial\theta\partial\theta^T} + J_\alpha\right\|$$
$$\leq \frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in N_\epsilon(\theta_0)}\left\|\frac{\partial^2 h_{\alpha,t}(\theta)}{\partial\theta\partial\theta^T} - \frac{\partial^2 h_{\alpha,t}(\theta_0)}{\partial\theta\partial\theta^T}\right\| + \max_{1\leq k\leq n}\frac{k}{n}\left\|\frac{\partial^2 H_{\alpha,k}(\theta_0)}{\partial\theta\partial\theta^T} + J_\alpha\right\|$$
$$:= I_n + II_n \quad a.s.$$

Using the ergodic theorem and (12), we can see that

$$\lim_{n\to\infty} I_n = E\left(\sup_{\theta\in N_\epsilon(\theta_0)} \left\| \frac{\partial^2 h_{\alpha,t}(\theta)}{\partial\theta\partial\theta^T} - \frac{\partial^2 h_{\alpha,t}(\theta_0)}{\partial\theta\partial\theta^T} \right\| \right) < \epsilon \quad a.s.$$

Also, from the fact that $\|\partial^2 H_{\alpha,n}(\theta_0)/\partial\theta\partial\theta^T + J_\alpha\| = o(1)$ a.s., we have

$$\max_{1\le k\le\sqrt{n}} \frac{k}{n} \left\| \frac{\partial^2 H_{\alpha,k}(\theta_0)}{\partial\theta\partial\theta^T} + J_\alpha \right\| \le \frac{1}{\sqrt{n}} \sup_k \left\| \frac{\partial^2 H_{\alpha,k}(\theta_0)}{\partial\theta\partial\theta^T} + J_\alpha \right\| = o(1) \quad a.s.$$

and

$$\max_{\sqrt{n}<k\le n} \left\| \frac{\partial^2 H_{\alpha,k}(\theta_0)}{\partial\theta\partial\theta^T} + J_\alpha \right\| = o(1) \quad a.s.,$$

which yield $II_n = o(1)$ a.s. The lemma is therefore obtained. □

Proof of Lemma 5 By the second result in Lemma 1 and the continuity of $\partial h_{\alpha,t}(\theta)/\partial\theta$ in θ, we can also take a neighborhood $N_\epsilon(\theta_0)$ such that

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \sup_{\theta\in N_\epsilon(\theta_0)} \left\| \frac{\partial h_{\alpha,t}(\theta)}{\partial\theta} \frac{\partial h_{\alpha,t}(\theta)}{\partial\theta^T} - \frac{\partial h_{\alpha,t}(\theta_0)}{\partial\theta} \frac{\partial h_{\alpha,t}(\theta_0)}{\partial\theta^T} \right\| \\ &= E\left(\sup_{\theta\in N_\epsilon(\theta_0)} \left\| \frac{\partial h_{\alpha,t}(\theta)}{\partial\theta} \frac{\partial h_{\alpha,t}(\theta)}{\partial\theta^T} - \frac{\partial h_{\alpha,t}(\theta_0)}{\partial\theta} \frac{\partial h_{\alpha,t}(\theta_0)}{\partial\theta^T} \right\| \right) < \epsilon \;\; a.s. \end{aligned} \tag{13}$$

Since $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$, one can prove the lemma by using (13) and the ergodic theorem. □

References

Atkinson, A. C., Koopman, S.-J., Shephard, N. (1997). Detecting shocks: Outliers and breaks in time series. *Journal of Econometrics*, *80*(2), 387–422.

Baek, C., Pipiras, V. (2014). On distinguishing multiple changes in mean and long-range dependence using local Whittle estimation. *Electronic Journal of Statistics*, *8*, 931–964.

Baek, C., Davis, R. A., Pipiras, V. (2018). Periodic dynamic factor models: Estimation approaches and applications. *Electronic Journal of Statistics*, *12*(2), 4377–4411.

Baek, C., Gates, K. M., Leinwand, B., Pipiras, V. (2021). Two sample tests for high-dimensional autocovariances. *Computational Statistics & Data Analysis*, *153*, 107067. https://doi.org/10.1016/j.csda.2020.107067.

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

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

Bai, J., Ng, S. (2007). Determining the number of primitive shocks in factor models. *Journal of Business & Economic Statistics*, *25*(1), 52–60.

Bai, J., Ng, S. (2008). *Large dimensional factor analysis*. Delft: Now Publishers Inc.

Balke, N. S., Fomby, T. B. (1994). Large shocks, small shocks, and economic fluctuations: Outliers in macroeconomic time series. *Journal of Applied Econometrics*, *9*(2), 181–200.

Basu, A., Harris, I. R., Hjort, N. L., Jones, M. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika*, *85*(3), 549–559.

Basu, A., Mandal, A., Martin, N., Pardo, L. (2013). Testing statistical hypotheses based on the density power divergence. *Annals of the Institute of Statistical Mathematics*, *65*, 319–348.

Basu, A., Mandal, A., Martin, N., Pardo, L. (2016). Generalized wald-type tests based on minimum density power divergence estimators. *Statistics*, *50*, 1–26.

Batsidis, A., Horvàth, L., Martin, N., Pardo, L., Zografos, K. (2013). Change-point detection in multinomial data using phi-divergence test statistics. *Journal of Multivariate Analysis*, *118*, 53–66.

Billingsley, P. (1999). *Convergence of probability measures*. New York, NY: Wiley.

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

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

Chen, M., An, H. Z. (1998). A note on the stationarity and the existence of moments of the GARCH model. *Statistica Sinica*, *8*(2), 505–510.

Cribben, I., Haraldsdottir, R., Atlas, L. Y., Wager, T. D., Lindquist, M. A. (2012). Dynamic connectivity regression: Determining state-related changes in brain connectivity. *Neuroimage*, *61*(4), 907–920.

Durio, A., Isaia, E. (2011). The minimum density power divergence approach in building robust regression models. *Informatica*, *22*, 43–56.

Fujisawa, H., Eguchi, S. (2006). Robust estimation in the normal mixture model. *Journal of Statistical Planning and Inference*, *136*, 3989–4011.

Ghosh, A., Basu, A. (2017). The minimum s-divergence estimator under continuous models: the Basu–Lindsay approach. *Statistical Papers*, *58*, 341–372.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., Stahel, W. A. (1986). *Robust statistics*: *The approach based on influence functions*. New York: Wiley.

Han, X., Inoue, A. (2015). Tests for parameter instability in dynamic factor models. *Econometric Theory*, *31*(5), 1117–1152.

Iglewicz, B., Hoaglin, D. C. (1993). *How to detect and handle outliers*, Vol. 16. Milwaukee: ASQC Quality Press.

Ledolter, J. (1989). The effect of additive outliers on the forecasts from Arima models. *International Journal of Forecasting*, *5*(2), 231–240.

Lee, S., Song, J. (2009). Minimum density power divergence estimator for GARCH models. *TEST*, *18*(2), 316–341.

Lee, T., Kim, M., Baek, C. (2015). Tests for volatility shifts in GARCH against long-range dependence. *Journal of Time Series Analysis*, *36*(2), 127–153.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Berlin: Springer.

Magnotti, J. F., Billor, N. (2014). Finding multivariate outliers in fMRI time-series data. *Computers in Biology and Medicine*, *53*, 115–124.

Poldrack, R. A. (2012). The future of fMRI in cognitive neuroscience. *Neuroimage*, *62*(2), 1216–1220.

Power, J. D., Schlaggar, B. L., Petersen, S. E. (2015). Recent progress and outstanding issues in motion correction in resting state fMRI. *Neuroimage*, *105*, 536–551.

Robbins, M., Gallagher, C., Lund, R., Aue A. (2011). Mean shift testing in correlated data. *Journal of Time Series Analysis*, *32*, 498–511.

Song, J. (2020). Robust test for dispersion parameter change in discretely observed diffusion processes. *Computational Statistics & Data Analysis*, *142*, 106832.

Song, J., Baek, C. (2019). Detecting structural breaks in realized volatility. *Computational Statistics & Data Analysis*, *134*, 58–75.

Song, J., Kang, J. (2019). Test for parameter change in the presence of outliers: The density power divergence based approach. arXiv:1907.00004.

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

Stock, J. H., Watson, M. (2011). *Dynamic factor models*. Oxford: Oxford University Press.

Stock, J. H., Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, *97*(460), 1167–1179.

Tsay, R. S., Pena, D., Pankratz, A. E. (2000). Outliers in multivariate time series. *Biometrika*, *87*(4), 789–804.
Van Den Heuvel, M. P., Pol, H. E. H. (2010). Exploring the brain network: A review on resting-state fMRI functional connectivity. *European Neuropsychopharmacology*, *20*(8), 519–534.
Warwick, J. (2005). A data-based method for selecting tuning parameters in minimum distance estimators. *Computational Statistics & Data Analysis*, *48*, 571–585.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.