Article:
Chen, Likai, Wang, Weining and Wu, Wei Biao (2020) Dynamic Semiparametric Factor Model with Structural Breaks. Journal of Business and Economic Statistics. 1730857. ISSN 0735-0015

https://doi.org/10.1080/07350015.2020.1730857

Dynamic Semiparametric Factor Model with Structural Breaks*

Likai Chen[1], Weining Wang[2], and Wei Biao Wu[3]

[1]Department of Mathematics, University of Washington at St. Louis
[2]Department of Economics and Related Studies, University of York
[3]Department of Statistics, University of Chicago

Abstract

For the change-point analysis of a high-dimensional time series, we consider a semiparametric model with dynamic structural break factors. With our model, the observations are described by a few low-dimensional factors with time-invariant loading functions of the covariates. Regarding the structural break, the factors are assumed to be nonstationary and follow a vector autoregression (VAR) process with a change in the parameter values. In addition, to account for the known spatial discrepancies, we introduce discrete loading functions. We study the theoretical properties of the estimates of the loading functions and the factors. Moreover, we provide both the consistency and the asymptotic normality for making an inference on the estimated breakpoint. Importantly, our results hold for both large and small breaks in the factor dependency structure. The estimation precision is further illustrated via a simulation study. Finally, we present two empirical applications in modeling the dynamics of the minimum wage policy in China and analyzing a limit order book dataset.

Keywords: high-dimensional time series, change-point analysis, temporal and cross-sectional dependency, vector autoregressive process

*We acknowledge valuable comments from the the Associate Editor and two anonymous referees, but the errors remain on our own.

1 Introduction

Emerging situations of big data call for statistical tools to learn the intrinsic complex structure. Modeling the structural breaks in a high-dimensional time-series is the object of our study. A very popular approach is to consider high dimensional time series with factor structure, see for example, Bai and Ng (2008) and Stock and Watson (2011). In this paper, we consider a characteristic-based factor model, which is used to describe the common movement of time series with nonparametric functions of covariates as loadings. The model is known as the dynamic semiparametric factor model (DSFM) and has been studied in Park et al. (2009) for time-varying covariates. Clearly, Connor et al. (2012) and Fan et al. (2016) also consider a similar model greatly applied in asset pricing, and the only difference is that the covariates are set to be time-invariant. In addition, a sizable literature has shown the applicability of such type of models as they take full advantage of the information provided by large cross-section and time-series dimensions. For example, the DSFM has been applied by Härdle et al. (2012) in modeling and forecasting the limit order book dynamics, by Fengler et al. (2007) in describing implied volatility surface dynamics, by van Bömmel et al. (2014) in discovering the risk patterns and brain activities, by Härdle and Majer (2016) in yield curving modeling, by Trück et al. (2014) in modeling and forecasting electricity spot prices, etc.

Although the DSFM successfully reduces the dimensionality of the data and disentangles the spatial and temporal effects, the common factors are prone to exhibit structural changes over a long time span. Many applications in economics and finance need a methodology on detecting and modeling structure breaks. For example, to evaluate the heterogeneity of the minimum wage policy in China, a cross-sectional data set on the minimum wage is collected over many years. One would like to see how the overall policy changes over time, and how much heterogeneity is there for different regions.

Nevertheless, there is a vast literature on detecting structural breaks in various statistical models other than a high dimensional semiparametric modeling framework. For example, Andrews (1993) and Bai and Perron (1998) address complex regression models; Wied et al. (2012) and Preuß et al. (2015) focus on the second-order characteristics of a time series; Dette and Wied (2016) propose a new formulation of a change point hypothesis testing. Furthermore, Wu and Zhao (2007) consider inference for trend-stationary

processes, and Shao and Zhang (2010) propose new tests for change point analysis in time series. Regarding high dimensional time series change point analysis, Jirak (2015) provides a framework to test the mean change using cumulative sum (CUSUM) statistics. Notably for factor models, there are also many recent articles on estimating loading changes, such as Cheng et al. (2016), Ma and Su (2016) that use shrinkage methods for large breaks. Bai et al. (2018) focus on establishing the asymptotic distribution of the least square estimator for the coefficients of factor models. Also, there is a recent paper by Manner et al. (2019) considering the issue of testing structural breaks in factor copula models.

However, there is no literature to our knowledge on conducting change-point analysis in a semiparametric high-dimensional time series modeling framework. For the latent dynamic factors, instead of estimating breaks in the loading parameters, we target the fitting of a structural break VAR process. The contribution lies in applying the existing change-point theory for both large and small breaks (for example, as in Bai (1997)) under a semiparametric model estimation framework. This requires connecting the literature on sieve estimation and change-point theory. We show that despite the semiparametric rate of estimates, we can still establish a similar type of results as in the low-dimension case. Moreover the estimated breakpoint is not subject to an identification issue although the factors dynamics is identified up to a sign matrix. Our estimator is in nature a semiparametric one with unknown factors and the nonparametric loading functions. For the break estimation, we establish the asymptotic distribution of our break point by allowing for general temporal and cross-sectional dependence in the error terms. It allows not only large breaks but also small breaks with magnitude tending to 0. This is beneficial as sometimes breaks are indeed small, and according to e.g. Bai et al. (2018), small breaks imply randomness of the estimated breaks in the limit and characterize the uncertainty of the estimated break points in the limit. For instance, an application is regarding analyzing limit order book data, where one is interested in understanding the quantities of the asset traded in the financial market. The loading functions are functions of price levels and may not be continuous from the ask side to the bid side. Moreover, the underlying driving factors may change in their dependency structure over time.

Our model is called as a structural break DSFM (SBDSFM), as we assume that factors follow a structural break vector autoregression model (SBVAR). The SBVAR is applied to change-point analysis

for the low-dimensional time series. For example, Galvão (2006) uses it for modeling policy change effects and predicting recessions. Moreover, to incorporate the cross-sectional effect, the discontinuity in the cross-sectional dimension is modeled as a partition estimator of the loading functions. To be more specific, the nonparametric loading functions are constructed by partitioning the support of the covariates into disjoint cells, which do not become smaller with the sample size, and within each cell, the unknown regression function is approximated by a basis expansion using a fixed-order B-spline basis. Therefore, using SBDSFM allows us therefore to account for the structural breaks in both time and cross-section dimensions. We show that the model is helpful for detecting and making an inference on the latent structural change for a high-dimensional time series.

We contribute to the literature in three aspects. First, we propose a SBDSFM model with an embedded break structure, and we also study its identification scheme and estimation method. Second, we show the consistency of the semiparametric estimates. For the latent factors, a consistency and distribution theorem allows us to make inferences on the breakpoints. Third, we illustrate the good empirical performance of our SBDSFM model via simulations and empirical examples. Our paper is organized as follows. Section 2 contains the model description and estimation; Section 3 has the relevant theorems and discussions. The simulations results are shown in Section 4. Section 5 consists of the two applications. The technical details are delegated to the Appendix in the supplementary materials.

2 Model

In this section, we describe the general setup of the model. First of all, we list the necessary math notation used throughout the paper. For matrix A, we denote $|A|_F$ (resp. $|A|_2$, $|A|_\infty$, $|A|_1$) as the matrix Frobenius norm (resp. spectral norm, ∞-norm, 1-norm). For $k > 0$ and vector $v = (v_1, \ldots, v_d)^\top \in \mathbb{R}^d$, let $|v|_k = (\sum_{i=1}^d |v_i|^k)^{1/k}$ and $|v|_\infty = \max_{1\le i\le d} |v_i|$. For two positive sequences of numbers (a_n) and (b_n), we denote $a_n = O(b_n)$ or $a_n \lesssim b_n$(resp. $a_n \asymp b_n$) if there exists a positive constant C such that $a_n/b_n \le C$(resp. $1/C \le a_n/b_n \le C$) for all large n, and we denote $a_n = o(b_n)$ or $a_n \ll b_n$ (resp. $a_n \sim b_n$), if $a_n/b_n \to 0$ (resp. $a_n/b_n \to 1$). For two sequences of random variables (X_n) and (Y_n), we write $X_n = o_{\mathbb{P}}(Y_n)$, if $X_n/Y_n \to 0$ in

probability. We let $\lambda_i(\cdot)$ be the ith largest eigenvalue, $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ be the minimum and maximum eigenvalues, respectively.

2.1. Dynamic semiparametric factor models (DSFMs) with breaks. We denote $\varepsilon_t = (\varepsilon_{t,1}, \varepsilon_{t,2}, ..., \varepsilon_{t,L})^\top$, $t = 1, 2, ..., T$, as a sequence of random vectors, and the model we consider is $(N, T \to \infty)$,

$$Y_{t,j} = m_0(X_{t,j}) + \sum_{l=1}^{L} Z_{t,l} m_l(X_{t,j}) + \varepsilon_{t,j}, \ \ j = 1, 2, \ldots, N. \tag{2.1}$$

Here, under the conditions of $\mathbb{E}(\varepsilon_{t,j}|X_{t,j}) = 0, \mathbb{E}(Z_{t,l}|X_{t,j}) = 0$, $m_0(X_{t,j})$ is the conditional expectation of $\mathbb{E}(Y_{t,j}|X_{t,j})$. The terms $m_l(X_{t,j}), l \geq 1$, are taken to be generalized nonparametric loading functions and $Z_t = (Z_{t,1}, Z_{t,2}, \cdots, Z_{t,L})^\top$ can be understood as the latent/unobservable common factors.

To incorporate the temporal break, we consider a time breakpoint $\tau^\diamond$. We let $\epsilon_t = (\epsilon_{t,1}, \epsilon_{t,2}, ..., \epsilon_{t,L})$, $t = 1, ..., T$, be i.i.d. random vectors. For $A = (A_1, A_2, ..., A_M)$, we denote $\mathcal{B}(A) = A_1\mathcal{B}^1 + A_2\mathcal{B}^2 + ... + A_M\mathcal{B}^M$, where $\mathcal{B}$ is the backward shift operator. We assume that the factors Z_t satisfy

$$Z_t \ \ = \ \ \mathcal{B}(E) Z_t \mathbf{1}_{t \leq \tau^\diamond} + \mathcal{B}(\tilde{E}) Z_t \mathbf{1}_{t > \tau^\diamond} + \epsilon_t, \quad t \geq 1, \tag{2.2}$$

and $Z_0, Z_{-1}, ..., Z_{-M+1}$ are some arbitrary vectors in $\mathbb{R}^L$, where $\tau^\diamond$ is an unknown constant,

$$E = (E_1, E_2, ..., E_M), \quad \text{and} \quad \tilde{E} = (\tilde{E}_1, \tilde{E}_2, ..., \tilde{E}_M), \tag{2.3}$$

are autoregressive coefficient matrices. Here the lags for the two regimes, denoted as M_1 and M_2, can be different or unknown by setting M to be large enough and $E_i = 0$, $\tilde{E}_j = 0$ for $i > M_1$, $j > M_2$. We assume that M is a fixed constant throughout this paper.

REMARK 1. [Comparison with Park et al. (2009)] Our settings in (2.1) are different from Park et al. (2009) in two major aspects. Firstly, Z_ts follow a nonstationary SBVAR process, namely the coefficients of VAR change after an unknown breakpoint, while Park et al. (2009) assumes that Z_t follows a stationary and strong mixing process. Assuming the SBVAR model allows for a change-point analysis, and the assumption leads to new issues of identification and estimation. Secondly, Park et al. (2009) assumes $\varepsilon_{t,j}$ to be i.i.d. and the distribution to be sub-Gaussian, while we have general assumptions allowing for spatial temporal dependence, and we impose only moment assumptions on the distribution of $\varepsilon_{t,j}$. □

The number of factors L is taken to be some fixed constant throughout the paper. Similar models without breaks are adopted in many papers in the literature (see, Connor et al. (2012); Fengler et al. (2007); Härdle et al. (2012); Park et al. (2009); Brüggemann et al. (2008); etc.)

The spatial discontinuity can be handled by taking into account the discontinuous basis functions for $m_l(\cdot)$. We assume that the functions $m_l(\cdot)$ can be approximated by $\sum_{k=1}^{J} a_{l,k}\phi_k(\cdot)$ as $J \to \infty$. Here $\phi_k(\cdot)$ is taken to be $\psi_i(\cdot)\mathbf{1}_{\cdot \in R_r}$, where R_rs forms the support of $\{X_{t,j}, 1 \le t \le T, 1 \le j \le N\}$. $\psi_i(\cdot)$ is the tensor product B-spline basis, and i, r corresponds to the index k. Then, the spatial discontinuity can be presented through regions R_r. For example, R_r can be used to model the presence of discontinuities of the regional minimum wage policies in China. We define the matrix of coefficients $A \overset{\text{def}}{=} (a_{l,k})_{l,k}$ (for every factor we have chosen the same number J of basis $\{\phi_k\}$). We denote matrix $\Phi(X_t) = (\phi(X_{t,1}), \phi(X_{t,2}), ..., \phi(X_{t,N}))^\top \in \mathbb{R}^{\mathbb{N}\times J}$, where $\phi(x) = (\phi_1(x), \phi_2(x), ..., \phi_J(x))^\top$. We let the loading functions be $m(x) = (m_0(x), m_1(x), \ldots, m_L(x))^\top$.

2.2. Temporal and cross-sectional dependencies in noise sequences. We let $\eta_{t,j}, t, j \in \mathbb{Z}$ be i.i.d. random variables with $\mathbb{E}(\eta_{t,j}) = 0$ and $\mathrm{Var}(\eta_{t,j}) = 1$. We denote $\eta_t = (\eta_{t,1}, \eta_{t,2}, \ldots, \eta_{t,N})^\top$. To incorporate both temporal and spatial dependencies, we shall consider the commonly used moving average (MA) process for the noise sequence $\varepsilon_t = (\varepsilon_{t,1}, ..., \varepsilon_{t,N})^\top$,

$$\varepsilon_t = \sum_{k\ge 0} B_k \eta_{t-k}, \tag{2.4}$$

where $(B_k)_{k\ge 0}$ are matrices in $\mathbb{R}^{N\times N}$ such that ε_t is a proper random vector. If $B_k = 0$ for all $k \ge 1$, then the noise sequences are temporally independent. If $B_0 = (I - \lambda W)^{-1}$, where λ is parameter governing the spatial dependency strength and W is a fixed adjacency matrix. Then ε_t follows a spatial autoregressive process. If matrices B_k are diagonal, then the sequences are spatially independent. In the latter case, $(\varepsilon_{t,j})_{t=1}^T$ becomes a MA sequence which is independently distributed with respect to different j. If in particular $B_k = (\lambda W)^k$, $k \ge 0$, where W is a lagged spatial weight matrix characterizing the lagged spatial dependency. The process will correspond to a regression model with a lagged spatial autoregressive structure. The MA(∞) process is very widely used in practice, and it includes many important time-series

models, such as vector autoregressive moving averages (VARMA),

$$(I - \sum_{l=1}^{p} \Theta_l \mathcal{B}^l)\varepsilon_i = \varepsilon_i - \sum_{l=1}^{p} \Theta_l \varepsilon_{i-l} = \sum_{k=1}^{q} \Xi_k \eta_{i-k},$$

where Θ_l and Ξ_k are real matrices such that $\det(I - \sum_{l=1}^{p} \Theta_l z^l)$ is not zero for all $|z| \leq 1$.

It should be noted that our theoretical results work with a predetermined number of factors similar to Bai et al. (2018). In practice, we follow the suggestions in the literature of estimating breaks for factor models: the computed number of factors should be $\hat{L}$, where $\hat{L}$ is achieved by taking the estimated number, similar in type to the Bayesian information criterion (BIC) by taking the estimated number, similar in type to the Bayesian information criterion (BIC).

2.1 Estimation

With the model specified by (2.1) on hand, we can estimate A, Z_t and $\tau^\diamond$ according to the following steps.

Step 1. We arbitrarily select $(\hat{A}^0,\ \hat{Z}_t^0, 1 \leq t \leq T)$ which minimizes

$$h(A, z_1, z_2, ..., z_T) = \sum_{t=1}^{T} |Y_t - \Phi(X_t)A(1, z_t^\top)^\top|_2^2,$$

which is

$$(\hat{A}^0, \hat{Z}_t^0, 1 \leq t \leq T) \in \text{argmin}_{A, z_t, 1 \leq t \leq T} h(A, z_1, z_2, ..., z_T). \tag{2.5}$$

It is not hard to see that the minimum point of $h(A, z_1, z_2, ..., z_T)$ is not unique. More specifically, for any minimizer $(\hat{A}^0, \hat{Z}_t^0, 1 \leq t \leq T)$, we let $\hat{A}_1^0$ and $\hat{A}_2^0$ be the first and $2 : (L+1)$th columns of matrix $\hat{A}^0$ respectively. Then for any invertible matrix $D \in \mathbb{R}^{L \times L}$, $((\hat{A}_1^0, \hat{A}_2^0 D), D^{-1}\hat{Z}_t^0, 1 \leq t \leq T)$ is also a solution. However for $\hat{Z}^0 = (\hat{Z}_1^0, \hat{Z}_2^0, ..., \hat{Z}_T^0)$, the product $\hat{A}_2^0 \hat{Z}^0$ is unique. Finding $\min h(A, z_1, z_2, ..., z_T)$ is non-trivial, since it involves a fourth-order problem. In practice, one may follow a Newton-Raphson method proposed in Park et al. (2009). For identification purpose, an

additional normalization step is needed. We consider the condensed singular value decomposition (SVD) of $T^{-1/2}\hat{A}_2^0\hat{Z}^0 = \hat{U}\hat{\Lambda}\hat{V}^\top$, where $\hat{\Lambda}$ is a diagonal matrix with diagonal entities $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_L$. (Note this decomposition is not unique. Even when $\hat{\lambda}_i \neq \hat{\lambda}_j$, for any i and j, the columns in $\hat{U}$ and $\hat{V}$ may differ up to a sign matrix. We shall select one of the decompositions as $\hat{U}$ and $\hat{V}$.) We let

$$\hat{A}_1 = \hat{A}_1^0, \quad \hat{A}_2 = \hat{U}\hat{\Lambda}, \quad \hat{A} = (\hat{A}_1^0, \hat{A}_2) \quad \text{and} \quad \hat{Z} = T^{1/2}\hat{V}^\top. \tag{2.6}$$

Then $T^{-1}\sum_t \hat{Z}_t\hat{Z}_t^\top = I_L$, and $\hat{A}_2^\top\hat{A}_2 = \hat{\Lambda}^2$ is a diagonal matrix.

Step 2. We denote $H = (H_1, H_2, ..., H_M)$, and $F = (F_1, F_2, ..., F_M)$ as the coefficient matrices, and τ as the change point. We consider the loss function

$$\hat{S}(\tau, H, F) = \sum_{t=1}^{T} \left|\hat{Z}_t - \mathcal{B}(H)\hat{Z}_t\mathbf{1}_{t\leq\tau} - \mathcal{B}(F)\hat{Z}_t\mathbf{1}_{t>\tau}\right|_2^2. \tag{2.7}$$

We denote $\hat{V}_\tau = \min_{H,F}\hat{S}(\tau, H, F)$. We set the estimate of the break location to be $\hat{\tau} = \text{argmin}_\tau\hat{V}_\tau$, and the estimate of E and $\tilde{E}$ as $(\hat{H}, \hat{F}) = \text{argmin}_{H,F}\hat{S}(\hat{\tau}, H, F)$.

REMARK 2. *The detailed numerical implementation and the selection of the number of factors will be discussed in Section 4.*

3 Theoretical Results

In this section, we provide consistency results for the parameters of interest. In addition the distribution theory is provided to facilitate making inference of the estimated breakpoint location. We consider $\min(N, T) \to \infty$ asymptotically. The relative rate of N, T is discussed in Remark 6. First we list a few assumptions.

3.1 Assumptions

ASSUMPTION 3.1 (Properties of ε_t). *We assume that the noise vectors in our SBDSFM, ε_t, $1 \le t \le T$, satisfy the MA(∞) model in (2.4) with the innovation sequence $(\eta_{i,j})$ and coefficient matrices (B_i).*

Moment. Let $(\eta_{t,j})$ be i.i.d. random variables with $\mathbb{E}(\eta_{t,j}) = 0$ and $\|\eta_{t,j}\|_q = (\mathbb{E}(|\eta_{t,j}|^q))^{1/q} < \infty$ for some $q \ge 4$. Denote $\mu_q = \|\eta_{t,j}\|_q$.

Dependence strength. We assume for some constants $c_B > 0, \beta_B > 1+1/q$, we have $|B_k|_2 \le c_B(k \vee 1)^{-\beta_B}$, where we recall that $|\cdot|_2$ represents the spectral norm of a matrix, $k \ge 0$.

ASSUMPTION 3.2 (Basis function). *Assume $\{X_t, 1 \le t \le T\}$ are independent of $\{\epsilon_t\}$ and $\{\varepsilon_t\}$. Also, we assume there are l_ϕ, $u_\phi > 0$ such that with probability approaching one as $N \to \infty$, we have*

$$l_\phi \le \min_{t=1,\dots,T} \lambda_{\min}\Big(\frac{1}{N}\Phi(X_t)^\top \Phi(X_t)\Big) \le \max_{t=1,\dots,T} \lambda_{\max}\Big(\frac{1}{N}\Phi(X_t)^\top \Phi(X_t)\Big) \le u_\phi.$$

It shall be noted that the above assumption would not require the covariates $X_{t,i}$ to be i.i.d., and instead would allow weak cross-sectional dependence. Moreover, we do not restrict the temporal dependence of $X_{t,i}$, and it can be perfectly time dependent, which means that $X_{t,i}$ does not vary with respect to t. In the special case of $X_{t,i}$ being cross-sectionally i.i.d., we can prove that Assumption 3.2 easily holds when

$$c_1 \le \min_{t=1,\dots,T} \lambda_{\min}\big(\mathbb{E}\phi(X_{t,1})\phi(X_{t,1})^\top\big) \le \max_{t=1,\dots,T} \lambda_{\max}\big(\mathbb{E}\phi(X_{t,1})\phi(X_{t,1})^\top\big) \le c_2, \tag{3.1}$$

with some constants $c_1, c_2 > 0$, and some additional conditions on N and T, c.f. Lemma 5 in the supplementary material.

ASSUMPTION 3.3. *(Properties of ϵ_t, innovations of Z_t) Assume $\epsilon_t, t \in \mathbb{Z}$, are i.i.d. random vectors in $\mathbb{R}^L$ with $\mathbb{E}(\epsilon_t) = 0$ and $\max_{1\le i\le L} \|\epsilon_{0,i}\|_{q'} < \infty$, for some $q' \ge 4$. Moreover, assume (ϵ_t) are independent of (ε_t). For the covariance matrix $\Sigma_\epsilon = \mathbb{E}(\epsilon_0\epsilon_0^\top)$, assume $l_\epsilon = \lambda_{\min}(\Sigma_\epsilon) > 0$.*

Especially when $\epsilon_{t,i}$ are i.i.d. for different $1 \le i \le L$, with mean zero and variance σ^2, then $\Sigma_\epsilon = \sigma^2 I_L$ and $l_\epsilon = \sigma^2$.

ASSUMPTION 3.4. *Assume for some* $0 < c < 1/2$, *both* $\tau^\diamond/T$ *and* $(T-\tau^\diamond)/T$ *are greater than* c.

ASSUMPTION 3.5. *(AR coefficients of* Z_t*)* *Assume there exists some invertible matrix* H *such that for coefficients* $E = (E_1, E_2, ..., E_M)$, $\tilde{E} = (\tilde{E}_1, \tilde{E}_2, ..., \tilde{E}_M)$,

(i) $\sum_{i=1}^M |HE_iH^{-1}|_2, \sum_{i=1}^M |H\tilde{E}_iH^{-1}|_2 \leq \gamma_e$, *for some constant* $0 < \gamma_e < 1$.

(ii) $\delta_e = |E - \tilde{E}|_2 > 0$ *and* $\delta_e T^{1/2} \to \infty$.

Under Assumption 3.5 (i), both $I - \mathcal{B}(E)$ and $I - \mathcal{B}(\tilde{E})$ are invertible. We let

$$\mathcal{Z}_t^{(l)} = (I - \mathcal{B}(E))^{-1}\epsilon_t,\ \mathcal{Z}_t^{(r)} = (I - \mathcal{B}(\tilde{E}))^{-1}\epsilon_t \text{ and } \mathcal{Z}_t = \mathcal{Z}_t^{(l)}\mathbf{1}_{t\leq\tau^\diamond} + \mathcal{Z}_t^{(r)}\mathbf{1}_{t>\tau^\diamond}. \tag{3.2}$$

Thus $\mathcal{Z}_t^{(l)}$ (resp. $\mathcal{Z}_t^{(r)}$) is stationary and satisfies the iteration $\mathcal{Z}_t^{(l)} = \mathcal{B}(E)\mathcal{Z}_t^{(l)} + \epsilon_t$ (resp. $\mathcal{Z}_t^{(r)} = \mathcal{B}(\tilde{E})\mathcal{Z}_t^{(r)} + \epsilon_t$).

Assumption 3.5 (ii) assures the patterns of the time series before and after the change point are different and δ_e represents the magnitude of the difference. It is also an important assumption for the identification of the breaks. We allow δ_e to go to 0, however, $\delta_e T^{1/2} \to \infty$. Denote the covariance matrices

$$W_1 = \mathbb{E}\mathcal{Z}_0^{(l)}\mathcal{Z}_0^{(l)\top},\ W_2 = \mathbb{E}\mathcal{Z}_0^{(r)}\mathcal{Z}_0^{(r)\top} \text{ and } W_0 = T^{-1}\sum_{t=1}^T \mathbb{E}\mathcal{Z}_t\mathcal{Z}_t^\top = W_1\tau^\diamond/T + W_2(T-\tau^\diamond)/T. \tag{3.3}$$

For $\mathbb{R}^{LM}$ vectors,

$$\xi_t^{(l)} = (\mathcal{Z}_{t-1}^{(l)\top}, \mathcal{Z}_{t-2}^{(l)\top}, ..., \mathcal{Z}_{t-M}^{(l)\top})^\top, \quad \xi_t^{(r)} = (\mathcal{Z}_{t-1}^{(r)\top}, \mathcal{Z}_{t-2}^{(r)\top}, ..., \mathcal{Z}_{t-M}^{(r)\top})^\top,$$

we define the $\mathbb{R}^{LM\times LM}$ matrices

$$\Sigma^{(l)} = \mathbb{E}\xi_0^{(l)}\xi_0^{(l)\top} \text{ and } \Sigma^{(r)} = \mathbb{E}\xi_0^{(r)}\xi_0^{(r)\top}, \tag{3.4}$$

which capture the autocovariance of $\mathcal{Z}_t$ up to the Mth lag, beyond that the covariance is 0.

REMARK 3. It is worth noting that under Assumption 3.3 and 3.5 (i), we have the positive definiteness of W_0, $\Sigma^{(l)}$ and $\Sigma^{(r)}$.

(i) For matrix W_0 defined in (3.3),

$$\lambda_{\min}(W_0) \geq l_\epsilon, \tag{3.5}$$

(ii) for matrices $\Sigma^{(l)}$ and $\Sigma^{(r)}$ defined in (3.4),

$$\lambda_{\min}(\Sigma^{(l)}), \lambda_{\min}(\Sigma^{(r)}) > 0.$$

□

Proof of Remark 3. Part (i) is due to the fact that both $\lambda_{\min}(W_1)$ and $\lambda_{\min}(W_2)$ are greater than l_ϵ. We note that $(I - \mathcal{B}(E))^{-1} = I + f_1 B^1 + f_2 B^2 + ...$, where f_is are matrices depending on E_i. Since ϵ_t are i.i.d., $\mathbb{E}\mathcal{Z}^{(l)}\mathcal{Z}^{(l)\top} = \Sigma_\epsilon + \sum_{i\geq 1} f_i \Sigma_\epsilon f_i^\top$ and thus $\lambda_{\min}(W_1) \geq \lambda_{\min}(\Sigma_\epsilon) \geq l_\epsilon$. The same arguments can be applied for W_2.

For part (ii), let $x = (x_1^\top, x_2^\top, ..., x_M^\top)^\top$, with $x_i \in \mathbb{R}^L$ and $|x|_2^2 = 1$. Denote i^* as the largest i such that $x_i \neq 0$. Since $Z_i^{(l)} = \epsilon_i + \sum_{k\geq 1} C_k \epsilon_{k-i}$, for some matrices C_k, we have

$$x^\top \Sigma^{(l)} x = \mathbb{E}(\sum_{i=1}^{M} x_i^\top \mathcal{Z}_i^{(l)})^2 = \mathbb{E}(x_{i^*}^\top \epsilon_{i^*} + H)^2,$$

where $H = \sum_{i=1}^{i^*-1} x_i^\top Z_i^{(l)} = \sum_{i=1}^{i^*-1} \sum_{k\geq 0} x_i^\top H_k \epsilon_{i-k}$, for some matrices H_k, which are independent of ϵ_{i^*}. Hence $x^\top \Sigma^{(l)} x \geq l_\epsilon |x_{i^*}|_2^2 > 0$. Same argument can be applied to $\Sigma^{(r)}$ and we complete the proof. □

Hence, M, L are both fixed constants, $\lambda_{\max}(\Sigma^{(l)}), \lambda_{\max}(\Sigma^{(r)}) < C$, where $C < \infty$ is some constant. We now set assumptions on the number of basis functions J, which should be a positive integer which diverges to infinity with N and T. We suppress the dependency of J with respect to N and T in the notation.

ASSUMPTION 3.6 (Loadings). *We assume $J \leq c_1 N^{\alpha_J}$, for some $0 < \alpha_J < 1/2$. There exists $A^* \in \mathbb{R}^{J\times L}$ and $\beta_J > 0$, such that,*

(i) $\delta_J = \sup_{x\in R} |m(x)^\top - \phi(x)^\top A^*|_\infty = O(J^{-\beta_J})$, *where R is the support of $\{X_{t,j}, 1 \leq t \leq T, 1 \leq j \leq p\}$.*

(ii) For A_2^* being the $(2 : (L+1))$ columns of A^*, assume

$$0 < l_a \leq \lambda_{\min}(A_2^{*\top} A_2^*) \leq \lambda_{\max}(A_2^{*\top} A_2^*) \leq u_a,$$

where l_a, u_a are some finite constants.

(iii) Assume that $A_2^* W_0 A_2^{*\top}$ *has* L *non-zero distinct eigenvalues and* $\text{gap}_L(A_2^* W_0 A_2^{*\top}) \geq l'_a$, *where* $l'_a > 0$ *is some constant and function* $\text{gap}_k(A) := \min_{1 \leq i \leq k}(\lambda_i(A) - \lambda_{i+1}(A))$.

This condition states that the factor loadings can be better approximated by basis functions $\phi(x)$ as the number of basis functions J increases. The value α_J restricts the dimensionality growth rate of J with respect to N. The value of β_J describes the approximation speed. For example, it can be taken as s/d for power series and splines, where d is the fixed dimension of $m(.)$ and s is the lower bound of the degree of smoothness of $m(.)$ within each regime R_r. This rate can be found for example in (Newey (1997), p.150). Regarding the selection of J, by minimizing the mean square error of the estimation, a cross-validation method is usually used for selecting J. We also require in (ii) that the decomposition is genuine in the sense that A_2^* always has full column rank by restricting the minimum eigenvalue of $A_2^{*\top} A_2^*$ to be greater than some positive number. This condition restricts the factors to be strong by assuming that the coefficient matrix A_2^* is well conditioned.

ASSUMPTION 3.7 (Identification condition). *Without loss of generality, let* $W_0 = I_L$ *and* $A_2^{*\top} A_2^*$ *be a diagonal matrix with distinct decreasing diagonal entities.*

We shall show that the feasibility of Assumption 3.1-3.7. By assumption 3.6 (iii), there exists an orthogonal matrix Q such that $Q^\top W_0^{1/2} A_2^{*\top} A_2^* W_0^{1/2} Q$ is diagonal with distinct decreasing diagonal entities. By Remark 3, W_0 is invertible. We denote $D = Q^\top W_0^{-1/2}$. For $Z'_t = DZ_t$ and $(m'_1(\cdot), m'_2(\cdot), \ldots, m'_L(\cdot)) = (m_1(\cdot), m_2(\cdot), \ldots, m_L(\cdot))D^{-1}$, $A_2^{*'} = A_2^* D^{-1}$ and (2.2) becomes

$$Z'_t = \mathcal{B}(E')Z'_t \mathbf{1}_{t \leq \tau^\diamond} + \mathcal{B}(\tilde{E}')Z'_t \mathbf{1}_{t > \tau^\diamond} + \epsilon'_t, \quad t \geq 1,$$

where $\epsilon'_t = D\epsilon_t$, $E'_i = DE_i D^{-1}$ and $\tilde{E}' = D\tilde{E}_i D^{-1}$. Then Assumptions 3.1 and 3.2 are unchanged, Assumption 3.3 holds in view of $\Sigma'_\epsilon = \mathbb{E}(\epsilon'_t \epsilon_t^{'\top}) = D\Sigma_\epsilon D^{-1}$ and $\lambda_{\min}(D\Sigma_\epsilon D^{-1}) > 0$. Note for $H' = HD^{-1}$, $\sum_{i=1}^M |H' E'_i H'^{-1}|_2 < 1$ and thus we have Assumption 3.5. For $A_2^{*'} = A_2^* D^{-1}$, since $\lambda_{\min}(D) > 0$, Assumption 3.6 holds. By (3.3), $W'_0 = DW_0 D^{-1} = I_L$, and we have $A_2^{*'\top} A_2^{*'} = D^{-\top} A_2^{*\top} A_2^* D^{-1} = Q^\top W_0^{1/2} A_2^{*\top} A_2^* W_0^{1/2} Q$, which is diagonal, hence Assumption 3.7 holds.

3.2 Estimation Consistency

Next, we show theorems regarding parameter consistency. Theorem 1 concerns the consistency results of the estimated parameters of SBDSFM, and Theorem 2 is on the consistency of the breakpoint estimate. We show that the identifiable part $A^*(1, Z_t^\top)^\top$ can be consistently estimated. Moreover both the coefficient matrix A_2^* and the factors Z_t can be estimated consistently up to an invertible matrix.

Regarding the semiparametric estimation of our loading functions and factors as in (2.5), we adopt the commonly used chaining technique in the empirical process theory. The proof strategy for the following theorem is utilizing empirical processes for the time series, in particular, the central task lies in examining the tail probability of the partial sum of products of the covariates and error terms. However, the object we study is nonstandard as we incorporate dynamic factors and dependent errors.

THEOREM 1 (Consistency of $\hat{Z}_t$). *Under Assumptions 3.1-3.7, we denote* $\rho^2 = (T+J)(TN)^{-(1-2/q)}\log(TN)$ *and assume* $\rho \to 0$. *Then*

(i) $T^{-1}\sum_{t=1}^{T} |\hat{A}(1, \hat{Z}_t^\top)^\top - A^*(1, Z_t^\top)^\top|_2^2 = O_{\mathbb{P}}(\rho^2 + \delta_J^2)$.

(ii) There exists a matrix D_T *such that* $|D_T - D^*|_F = O_{\mathbb{P}}(T^{-1/2})$, *where* D^* *is some diagonal matrix with diagonal entities either* -1 *or* 1, *and*

$$T^{-1}\sum_{t=1}^{T} |D_T Z_t - \hat{Z}_t|_2^2 = O_{\mathbb{P}}(\rho^2 + \delta_J^2).$$

REMARK 4. The rate of Theorem 1 is similar to Theorem 2 in Park et al. (2009). It can be seen that moment (Assumption 3.1) on the innovations $\eta_{i,j}$ plays a role in the rate of convergence in Theorem 1. In particular, a larger value of q means a stronger moment assumption and thus a faster rate of convergence. The introducation of D^* is unavoidable due to the identification issue. The appearance of D_T is because of the structure on Z_t so that we cannot impose the condition that $T^{-1}ZZ^\top = I_L$ but only $|T^{-1}ZZ^\top - \mathrm{I}_L|_F = O_{\mathbb{P}}(T^{-1/2})$. □

REMARK 5. We do not restrict the temporal dependence of X_t. In the special case that X_t does not change over time t, then our model (2.1) is the same as the model considered in Fan et al. (2016) and our least

square estimator in (2.5) would have a principal component interpretation. However, as we have numerous different theoretical assumptions (e.g., the moment and dependence assumptions on the noise sequence), we can only compare the rates of the estimators in some special cases. If we assume the innovations have sub-exponential tails, then we would have $\rho^2 = (T+J)(TN)^{-1}\log^3(TN)$. Ignoring the sieve approximation bias δ_J^2 and assuming the dependence assumptions on noises are satisfied in both papers, our convergence rate of factor would be

$$T^{-1}\sum_{t=1}^{T}|D_T Z_t - \hat{Z}_t|_2^2 = O_{\mathbb{P}}((T+J)(TN)^{-1}\log^3(TN)). \tag{3.6}$$

The corresponding rate in their paper is $1/N$. In comparison, our convergence rate is slightly slower up to a $\log^3(NT)$ factor if $T \geq J$. □

For the estimation of the breakpoint, we need the following assumption.

ASSUMPTION 3.8. *We assume that* $\rho^2 + \delta_J^2 = O(T^{-1})$.

REMARK 6. Let $N \asymp T^r$. Then under Assumption 3.6 (i), Assumption 3.8 holds if

$$r \geq \max\{1/(2\alpha_J\beta_J), (q+2)/(q-2), 2/(q-2-\alpha_J q)\}.$$

Recall that $J = O(N^{\alpha_J})$ and $\delta_J = N^{-\alpha_J\beta_J}$. This condition assumes a larger rate of N than T, as $r > 1$. Also the rate of N interplays with q regarding the moment assumption, the number of basis functions and the bias δ_J. □

Next we provide a theorem on the consistency of the change-point estimate.

THEOREM 2. *(Consistency of* $\hat{\tau}$*) Under Assumptions 3.1-3.8, for* $T^{1/2}\delta_e \to \infty$, *we have* $|\hat{\tau} - \tau^\diamond| = O_{\mathbb{P}}(\delta_e^{-2})$.

The rate of consistency is determined by the magnitude of the change δ_e. This means that in terms of the fractions of the sample size, the estimated break points converge rapidly to the true fraction. In the special case with the fixed break size δ_e, we have

$$|\hat{\tau} - \tau^\diamond| = O_{\mathbb{P}}(1),$$

which means that the difference between the estimated breakpoint and the true one is stochastically bounded. A similar rate can be found in Bai (1995). Note that the break estimate consistency rate is determined by the break size δ_e and as δ_e is defined to be aggregated from low-dimensional factors, our rate is in line with the univariate cases. This is consistent with the setting and results in Baltagi et al. (2016) for breaks estimation in the classical factor model. Different rates of consistentency of $\hat{\tau}$ can be achieved for breaks in a panel data model where the cross-sectional dimension contributes to the order of the break sizes, see Bai (2010) and Bai et al. (2018) for breaks in the factor model.

3.3 Asymptotic Distribution

In this subsection, we show the distribution theory of the estimated coefficient matrix within the regime and the change point estimate. For matrix $A \in \mathbb{R}^{m\times n}$, denote $vect(A) = (A_1^\top, A_2^\top, ..., A_m^\top)^\top$, where $A_i = (A_{i,1}, A_{i,2}, ..., A_{i,n})^\top$ is the ith row of matrix A.

We define the block matrix $\tilde{D}_T = I_M \otimes D_T$ where $\otimes$ is denoted as the Kronecker product between two matrices. We define the true coefficient matrix scaled by D_T and $\tilde{D}_T$ as

$$H^\diamond = D_T E \tilde{D}_T^{-1}, \quad F^\diamond = D_T \tilde{E} \tilde{D}_T^{-1}. \tag{3.7}$$

We define that $\Sigma_{\epsilon,i,j}$ is the i,j th entry of Σ_ϵ in Assumption 3.3, and $\Sigma^{(l)}$ is defined in (3.4). The asymptotic normality of the estimated coefficient matrices is shown below.

ASSUMPTION 3.9. *We assume that* $\rho^2 + \delta_j^2 = o(T^{-1})$.

THEOREM 3 (Central limit theorem for within regime parameters). *Under Assumptions 3.1-3.7 and 3.9. Let* $\Theta^{(l)} = (N_{i,j}^{(l)})_{1\le i,j\le L}$ *where* $N_{i,j}^{(l)} \in \mathbb{R}^{LM\times LM}$ *with* $N_{i,j}^{(l)} = \Sigma_{\epsilon,i,j}\Sigma^{(l)-1}$, *and let* $\Theta^{(r)} = (N_{i,j}^{(r)})_{1\le i,j\le L}$ *with* $N_{i,j}^{(r)} = \Sigma_{\epsilon,i,j}\Sigma^{(r)-1}$. *Then*

$$\tau^{\diamond 1/2} vect(D_T^{-1}(\hat{H} - H^\diamond)\tilde{D}_T) \Rightarrow N(0, \Theta^{(l)})$$

and

$$(T - \tau^\diamond)^{1/2} vect(D_T^{-1}(\hat{F} - F^\diamond)\tilde{D}_T) \Rightarrow N(0, \Theta^{(r)}).$$

REMARK 7. The rate of convergence depends on the number of observations available within each regime, and the asymptotic efficiency of the estimation is related to the auto-covariance structure of the process Z_t. □

Then we provide a theorem on the asymptotic distribution of the change point estimate $\hat{\tau}$. We consider the type of contiguous asymptotics, where δ_e tends to zero in the limit. We show that the loss function for estimating the change point can be approximated by a two-sided Brownian motion with a triangular drift.

THEOREM 4. *(Asymptotic distribution of $\hat{\tau}$) Under Assumptions 3.1-3.7, 3.9 and additionally $\delta_e \to 0$. Let $\mathcal{Q}^{(l)} = \delta_e^{-2}(E - \tilde{E})\Sigma^{(l)}(E - \tilde{E})^\top$ and $\mathcal{Q}^{(r)} = \delta_e^{-2}(E - \tilde{E})\Sigma^{(r)}(E - \tilde{E})^\top$. Then*

$$\delta_e^2(\hat{\tau} - \tau^\diamond) \Rightarrow \text{argmin}_s H(s),$$

$$\text{where} \quad H(s) = \begin{cases} -\mathsf{tr}(\mathcal{Q}^{(l)})s + 2\mathsf{tr}^{1/2}(\mathcal{Q}^{(l)}\Sigma_\epsilon)W_1(-s), & \text{if } s \le 0, \\ \mathsf{tr}(\mathcal{Q}^{(r)})s + 2\mathsf{tr}^{1/2}(\mathcal{Q}^{(r)}\Sigma_\epsilon)W_2(s), & \text{if } s > 0, \end{cases}$$

where $W_1(\cdot)$ and $W_2(\cdot)$ are independent standard Wiener processes.

REMARK 8. For the change-point detection in a univariate regression model, Bai (1997) provides a similar type of consistency and asymptotic normality results. In our setup, we consider a VAR model with a structural break, and $\hat{Z}_t$ is with generated error from our semiparametric estimation. □

REMARK 9. Importantly we see that the change-point estimation is not related to the matrix D_T and is not subject to an identification issue. □

Regarding the limit distribution of Theorem 4, we can obtain the analytical form of the cumulative distribution function (cdf). Then we can obtain the quantile by inverting the cdf for the confidence interval. Alternatively we can simulate the path of the limit distribution directly. Also one can create confidence intervals for $\hat{\tau}$ via a multiplier bootstrap method on the estimated residuals which does not require an asymptotic distribution.

From the above theorem, with estimates of $\mathsf{tr}(\mathcal{Q}^{(l)})$, $\mathsf{tr}^{1/2}(\mathcal{Q}^{(l)}\Sigma_\epsilon)$, $\mathsf{tr}(\mathcal{Q}^{(r)})$ and $\mathsf{tr}^{1/2}(\mathcal{Q}^{(r)}\Sigma_\epsilon)$, we can construct a $100(1-\alpha)\%$ confidence interval for $\hat{\tau}$:

$$[\hat{\tau} - \lfloor \hat{q}_{1-\alpha/2} \rfloor - 1, \hat{\tau} + \lfloor \hat{q}_{\alpha/2} \rfloor + 1], \tag{3.8}$$

where $q_{1-\alpha/2}$ ($q_{\alpha/2}$) is $1-\alpha/2$ ($\alpha/2$)th quantile of $\mathrm{argmin} H(s)$, and $\hat{q}_{\alpha/2}(\hat{q}_{1-\alpha/2})$ is a estimate.

Denote $q_l = \mathsf{tr}(\mathcal{Q}^{(l)})$, $\sigma_l = 2\mathsf{tr}^{1/2}(\mathcal{Q}^{(l)}\Sigma_\epsilon)$, $q_r = \mathsf{tr}(\mathcal{Q}^{(r)})$ and $\sigma_r = 2\mathsf{tr}^{1/2}(\mathcal{Q}^{(r)}\Sigma_\epsilon)$. We let $\theta_1 = q_l/\sigma_l$ (q_r/σ_r) for $t \le 0$ $(t > 0)$ and $\theta_2 = q_r\sigma_l/\sigma_r^2$ $(q_l\sigma_r/\sigma_l^2)$ for $t \le 0$ $(t > 0)$. We denote the cdf

$$\begin{aligned} F(s, \theta_1, \theta_2) &= (2\pi)^{-1/2} 2\theta_1 s^{1/2} \exp(-\theta_1^2 s/2) - (2\theta_1^2 x + [\theta_1^2 + 2\theta_2^2 + 2\theta_1\theta_2]/[\theta_2(\theta_1+\theta_2)])\Phi(-\theta_1 s^{1/2}) \\ &\quad + (\theta_1(\theta_1+2\theta_2))/(\theta_2(\theta_1+\theta_2))\exp\{2\theta_2(\theta_1+\theta_2)s\}\Phi(-(\theta_1+2\theta_2)s^{1/2}). \end{aligned}$$

Then according to Stryhn (1996), the distribution function of $\mathrm{argmin}_s H(s)$ is of the following form,

$$F(s) = -F(|s|, \theta_1, \theta_2), \quad s \le 0,$$

$$F(s) = 1 + F(|s|, \theta_1, \theta_2), \quad s > 0.$$

4 Simulation

In this section, we run simulations under different settings to evaluate our model performance. First, we suggest an algorithm for our estimation:

Initial Value for $\hat{Z}^0$ and $\hat{A}^0$. We denote the matrix of the loading functions as $\Psi_t \overset{\text{def}}{=} (m_l(X_{i,t}))_{1\le i\le N, 1\le l\le L+1}$. The initial estimation of $Z(T\times L)$ and $A(J\times L)$ can be obtained as follows.

Step 1 We estimate firstly $\Gamma_t \overset{\text{def}}{=} A(1, Z_t^\top)^\top$ $(\Gamma = (\Gamma_1, \Gamma_2, \cdots, \Gamma_T)_{J\times T})$ and we let $\hat{\Gamma}^0 = \mathrm{argmin}_\Gamma S(\Gamma) = \mathrm{argmin}_\Gamma \sum_t |Y_t - \Phi(X_t)\Gamma|_2^2 = \mathrm{argmin}_\Gamma \sum_t |Y_t - \Phi(X_t)A(1, Z_t^\top)^\top|_2^2$. It is not hard to see that $\hat{\Gamma}_t^0 = \{\Phi^\top(X_t)\Phi(X_t)\}^{-1}\Phi^\top(X_t)Y_t$. We can obtain that $\hat{\Gamma}^0 = (\hat{\Gamma}_1^0, \hat{\Gamma}_2^0, \cdots, \hat{\Gamma}_T^0)_{J\times T}$.

Step 2 We denote the condensed SVD of $\hat{\Gamma}^0$ as $\hat{U}^0\hat{\Lambda}^0\hat{V}^{\top 0}$, where $\hat{\Lambda}^0 = \text{diag}(\hat{\lambda}_1^0, \hat{\lambda}_2^0, ..., \hat{\lambda}_L^0)$ (The first L largest singular values), $\hat{\lambda}_1^0 \geq \hat{\lambda}_2^0 \geq ... \geq \hat{\lambda}_L^0$. We set the initial factors to be $\hat{Z}^0 = \hat{\Lambda}^0\hat{V}^{0\top}$, and the loadings to be $A^0 = \hat{U}^0$.

The initial step based on a SVD of the estimated coefficient matrix $\hat{\Gamma}^0$. The number of factors is prefixed by the initial selection stage. We focus on the cases of having the minimum fixed number of factors i.e. $\hat{L}$ following the principle of parsimony; ($\hat{L}$ is taken from the BIC criteria $\log(NT^{-1}h(L)) + L(N+T)/(NT)\log(NT/(N+T))$, where $h(L)$ is the sum of square residuals with respect to different factors). In our applications senarios, we can always achieve more than 99% precision for 5000 simulation samples with the factor number selection.

Iteration and Change-Point Next, we show that given $\hat{Z}^0$ and $\hat{A}^0$, we can further obtain an estimate as follows.

Step 1 Given the estimates $\hat{A}^0$ and $\hat{Z}^0$, one can iterate between the estimation of A and $Z \overset{\text{def}}{=} (Z_1, Z_2, \cdots, Z_T)$ following the loss:

$$\text{argmin}_{A,Z} \sum_t |Y_t - \Phi(X_t)A(1, Z_t^\top)^\top|_2^2. \tag{4.1}$$

Step 2 We follow a Newton-Raphson method proposed in Park et al. (2009) to obtain $\hat{A}$ and $\hat{Z}_t$. $\hat{A}$ gives us estimates of factors loadings $\hat{m}_l(\cdot)$, and $\hat{Z}_t$s are the estimated factors.

Step 3 Assuming Z_ts follows a SBVAR process in (2.2), we plug in the least square loss as in (2.7).

Step 4 We apply a binary segmentation algorithm as in Scott and Knott (1974) for estimating the single break τ over the interval $[Td_0\%, T(1-d_0)\%]$ ($d_0 > 0$), namely by minimizing

$$\text{argmin}_{\tau,H,F} S_{1:\tau}(H) + S_{(\tau+1):n}(F). \tag{4.2}$$

To set up the simulation, the following data generating processes are taken,

$$Y_{i,t} = m_0(X_{i,t}) + \sum_l Z_{l,t} m_l(X_{i,t}) + \sigma\varepsilon_{it}, \tag{4.3}$$

where N, T have the cases $T = 50, 100, 200, 300$ and $N = 50, 100, 200, 300$. Each element of X_{it} is taken to be uniformly distributed over $[-3, 3]$. We let $m_0(x_1, x_2) = 0$. The $m_l(\cdot, \cdot)s$ are taken to be Model I:

$$\begin{aligned} m_1(x_1, x_2) &= \mathbf{1}(x_1 < a_1, x_2 < a_2)\{(9.45((x_1 - 0.5)^2 + (x_2 - 0.5)^2) - 1.6)/30\} \\ &+ \mathbf{1}(x_1 \geq a_1, x_2 \geq a_2)\{(2.45((x_1 - 0.5)^2 + (x_2 - 0.5)^2) - 1.6)/30\}, \\ m_2(x_1, x_2) &= 3\sin(0.5\pi x_2)\mathbf{1}(x_1 < b_1, x_2 < b_2) \\ &\quad + \sin(0.7\pi x_2)\mathbf{1}(x_1 \geq b_1, x_2 \geq b_2), \end{aligned}$$

where $a_1 = a_2, b_1 = b_2$ are selected to be either 0.5 or 0.7.

As a second alternative, we consider a different set of functions with $x_1, x_2 \in [-1, 1]$, which is denoted as Model II:

$$\begin{aligned} m_1(x_1, x_2) &= \mathbf{1}(x_1 < a_1, x_2 < a_2)\{(\exp(-(x_1 - 0.1)^2/0.3) + 0.1\exp(-(x_2 - 0.5)^2/2))/2.39\} \\ &+ \mathbf{1}(x_1 \geq a_1, x_2 \geq a_2)\{\sin(0.7\pi x_2)\}, \\ m_2(x_1, x_2) &= \{2\sin(2\pi x_1) - 6|x_2 - 0.4|^{0.3}\}\mathbf{1}(x_1 < b_1, x_2 < b_2) \\ &\quad + \{2.45((x_1 - 0.5)^2 + (x_2 - 0.5)^2 - 1.6)\}\mathbf{1}(x_1 \geq b_1, x_2 \geq b_2). \end{aligned}$$

The time series $\{Z_t\}$ is generated from a SBVAR process as in (2.2), with $E = [0.5, -0.2, 0; 0, 0.8, 0.1; 0.1, 0, 0.6]$, and $\tilde{E} = [0.6, 0.1, 0; 0, 0.3, -0.1; 0.01, 0, 0.2]$. Also, ϵ_{tl}s are either i.i.d. normal random variables with standard deviation 0.001 before the break and 0.01 after the break or variables following t location scale distributions with mean zero and the same variances (with 5 degrees of freedom). The true breakpoint τ is $[T/2]$ or $[T/4]$. In addition, ε_{it}s are set to be: i) an independent standard normal distributed $N(0, 0.1)$, and ii) an independent ARMA(1,1) process, $\varepsilon_{it} = 0.3\varepsilon_{i(t-1)} + \eta_{it} + 0.5\eta_{i(t-1)}$, where η_{it} and $\eta_{i(t-1)}$ are normal random variables of $N(0, 0.1)$.

The simulation setup is to account for different types of signal-to-noise ratios for different values of σ, in particular $\sigma_1 = 0.1$ or $\sigma_2 = 0.2$. For instance, $\sigma_2 = 0.2$ amount to roughly 20% to 40% in terms of the proportion of the error variance to the variance in the observations. In addition, we allow for the serial correlations of the error processes $\varepsilon_{it}s$ by case ii). Figure 1 presents the $m_l(\cdot, \cdot)$ function under the case

of $a_1 = 0.5$, $a_2 = 0.5$, $b_1 = 0.5$, $b_2 = 0.5$. Figure 2 shows the plots of the simulated two factors and the estimated confidence intervals; with a change-point at 100, one can observe a switching of the dispersion of the factors.

For the estimation of basis functions, we consider keeping the tensored quadratic B-splines to be the same within the regions defined according to a_1, a_2, b_1 and b_2. By Theorem 1, the covariance structure of $\hat{Z}_t$ is identified up to a sign matrix D_T. Denote the centered $\hat{Z}_{t,c}$ as $\hat{Z}_{t,c} = \hat{Z}_t - T^{-1}\sum_{t=1}^{T}\hat{Z}_t$, and the estimated $\hat{D}_T$ can be the solution to minimize $\sum_t |\hat{Z}_{t,c} - D_T Z_{t,c}|_2^2$, which is $\hat{D}_T = (\sum_t Z_{t,c}Z_{t,c}^\top)^{-1}(\sum_t Z_{t,c}\hat{Z}_{t,c}^\top)$. We work with the transformed estimate $\tilde{Z}_t = \hat{D}_T^{-1}\hat{Z}_t$. We define a measure of the scale differences between the estimated covariance matrix and the true one,

$$e_f = |\frac{1}{\sqrt{T}}\{\sum_{t=1}^{T}(\tilde{Z}_t - \bar{\tilde{Z}})(\tilde{Z}_t - \bar{\tilde{Z}})^\top - \sum_{t=1}^{T}(Z_t - \bar{Z})(Z_t - \bar{Z})^\top\}|_2. \tag{4.4}$$

To evaluate the accuracy of the estimation, the 90% confidence intervals of the estimated change point as in (3.8) is implemented, and Table 1 reports the estimated coverage probabilities over 5000 samples in different simulation scenarios. Also, Table 2 and 3 presents the explained variances of the fitted model and e_f.

The estimation errors appear to be moderate across different estimation cases. In particular, they are robust against different error distributions, innovation processes and signal-to-noise ratios. Moreover, we have also shown good recovery rates of the breakpoint over time. When the sample size increases, one sees a tendency of an overall better performance.

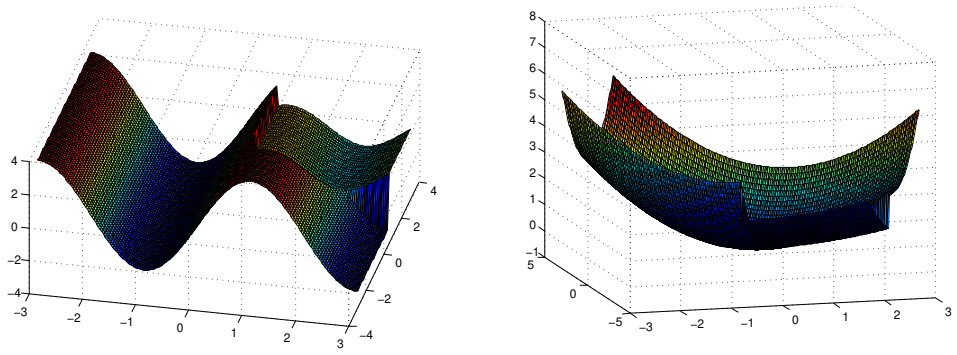


Figure 1: Plot of true loading functions with $a_1 = 0.5$, $a_2 = 0.5$, $b_1 = 0.5$, $b_2 = 0.5$.

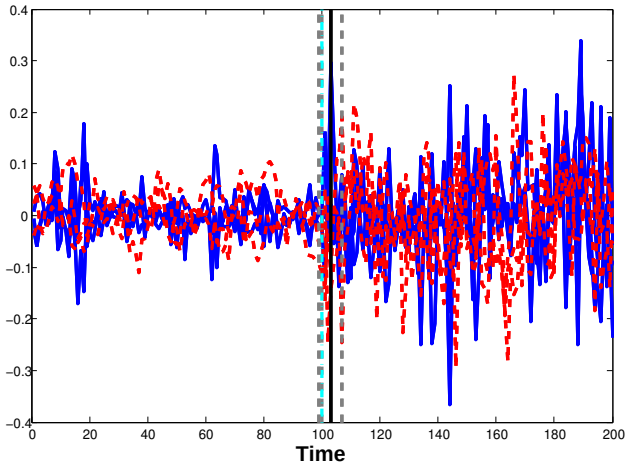


Figure 2: Plot of simulated normalized true factors with $T = 200$, $N = 200$, breakpoint $\tau = 100$(cyan). Estimated breakpoint $\hat{\tau} = 103$(black), and estimated confidence region $[99, 107]$(dotted grey).

Table 1: The coverage probability is $\times 10^2$. G denotes normal innovations and T denotes t location scale distributions (with 5 degrees of freedom) innovations. $\tau = [T/2]$ or $[T/4]$. The average is taken over 5000 samples. $\alpha = 0.05$. $50, 100$ means $N = 50, T = 100$, and the same for others. I and II indicate Model I and II.

			$50, 100$	$100, 50$	$200, 200$	$300, 300$
I,i)	σ_1	G $[T/2]$	75.3	77.2	85.3	89.6
		$[T/4]$	76.9	77.9	83.3	90.1
		T $[T/2]$	72.7	73.8	82.7	88.8
		$[T/4]$	70.3	72.5	78.4	87.8
	σ_2	G $[T/2]$	71.1	73.4	84.2	87.5
		$[T/4]$	72.2	73.2	79.1	87.6
		T $[T/2]$	68.1	72.1	78.9	88.5
		$[T/4]$	70.1	74.1	83.1	89.3
I,ii)	σ_1	G $[T/2]$	66.5	71.0	73.3	87.2
		$[T/4]$	65.3	66.8	74.1	87.1
		T $[T/2]$	64.3	66.2	74.1	86.5
		$[T/4]$	62.9	59.2	75.5	86.9
II,ii)	σ_2	G $[T/2]$	63.3	66.7	71.2	86.8
		$[T/4]$	64.2	63.6	70.6	85.1
		T $[T/2]$	59.8	62.5	70.2	84.3
		$[T/4]$	59.1	61.7	70.1	85.0

Table 2: Model I: The averaged percentage of explained variance of the model, e_f and their variances (in parentheses). All numbers are presented in $\times 10^2$, G denotes normal innovations and $T(5)$ denotes a t location scale distribution (with 5 degrees of freedom). The results are averaged over 5000 samples. Exp. var. denotes explained variance.

			$N = 50, T = 100$		$N = 100, T = 50$		$N = 200, T = 200$		$N = 300, T = 300$	
			Exp. var.	e_f	Exp. var.	e_f	Exp. var.	e_f	Exp. var.	e_f
i)	σ_1	G $[T/2]$	75.12 (0.34)	1.59 (0.02)	79.12 (0.24)	1.55 (0.02)	93.32 (0.25)	1.45 (0.02)	94.11 (0.18)	1.44 (0.01)
		$[T/4]$	73.09 (0.39)	1.62 (0.03)	82.56 (0.27)	1.61 (0.03)	86.57 (0.24)	1.55 (0.03)	93.44 (0.17)	1.50 (0.01)
		$T(5)$ $[T/2]$	70.34 (0.44)	1.74 (0.05)	74.35 (0.32)	1.62 (0.06)	78.35 (0.27)	1.54 (0.05)	92.32 (0.20)	1.53 (0.04)
		$[T/4]$	71.89 (0.41)	1.72 (0.06)	75.85 (0.30)	1.67 (0.05)	81.02 (0.26)	1.57 (0.04)	92.76 (0.23)	1.42 (0.03)
	σ_2	G $[T/2]$	69.43 (0.40)	1.66 (0.05)	70.34 (0.33)	1.67 (0.03)	72.02 (0.28)	1.63 (0.03)	73.14 (0.25)	1.56 (0.02)
		$[T/4]$	70.04 (0.42)	1.65 (0.06)	71.34 (0.32)	1.64 (0.06)	73.65 (0.29)	1.60 (0.06)	74.44 (0.28)	1.58 (0.04)
		$T(5)$ $[T/2]$	69.28 (0.48)	1.83 (0.07)	72.13 (0.35)	1.74 (0.07)	73.24 (0.31)	1.64 (0.05)	74.12 (0.30)	1.56 (0.05)
		$[T/4]$	64.68 (0.49)	1.86 (0.10)	65.13 (0.37)	1.84 (0.09)	69.51 (0.34)	1.70 (0.08)	74.18 (0.32)	1.57 (0.07)
ii)	σ_1	$G[T/2]$	72.23 (0.41)	1.80 (0.11)	74.03 (0.33)	1.76 (0.09)	83.42 (0.32)	1.65 (0.07)	88.12 (0.30)	1.63 (0.06)
		$[T/4]$	72.06 (0.46)	1.99 (0.10)	75.43 (0.34)	1.89 (0.08)	79.01 (0.31)	1.78 (0.08)	86.54 (0.29)	1.65 (0.07)
		$T(5)$ $[T/2]$	71.63 (0.48)	2.01 (0.13)	73.38 (0.43)	1.89 (0.10)	76.56 (0.38)	1.65 (0.09)	87.36 (0.41)	1.59 (0.08)
		$[T/4]$	70.16 (0.54)	2.04 (0.16)	74.16 (0.46)	1.88 (0.11)	77.11 (0.36)	1.69 (0.08)	86.53 (0.36)	1.56 (0.08)
	σ_2	G $[T/2]$	65.03 (0.47)	1.98 (0.13)	67.89 (0.42)	1.85 (0.10)	68.01 (0.36)	1.74 (0.10)	71.89 (0.33)	1.70 (0.09)
		$[T/4]$	66.23 (0.45)	2.09 (0.14)	68.67 (0.47)	1.97 (0.11)	68.73 (0.45)	1.89 (0.09)	70.27 (0.43)	1.89 (0.10)
		$T(5)$ $[T/2]$	63.07 (0.55)	2.13 (0.16)	66.89 (0.49)	1.96 (0.15)	67.59 (0.47)	1.87 (0.12)	72.13 (0.42)	1.75 (0.11)
		$[T/4]$	61.89 (0.56)	2.12 (0.17)	65.96 (0.55)	1.98 (0.17)	66.78 (0.49)	1.89 (0.11)	70.08 (0.45)	1.71 (0.12)

Table 3: Model II: The averaged percentage of explained variance of the model, e_f and their variances (in bracket). All numbers are presented in $\times 10^2$, G denotes normal innovations and $T(5)$ denotes a t location scale distribution (with 5 degree of freedom). The results are averaged over 5000 samples. Exp. var. denotes explained variance.

			$N=50, T=100$		$N=100, T=50$		$N=200, T=200$		$N=300, T=300$	
			Exp. var.	e_f	Exp. var.	e_f	Exp. var.	e_f	Exp. var.	e_f
i)	σ_1	G $[T/2]$	74.55 (0.61)	3.47 (0.13)	76.98 (0.59)	3.11 (0.10)	81.40 (0.48)	2.48 (0.08)	85.98 (0.43)	2.35 (0.07)
		$[T/4]$	73.04 (0.69)	3.72 (0.12)	74.59 (0.63)	3.40 (0.08)	80.75 (0.55)	2.98 (0.06)	85.90 (0.48)	2.51 (0.05)
		$T(5)$ $[T/2]$	72.80 (0.68)	3.84 (0.15)	72.31 (0.60)	3.50 (0.12)	77.24 (0.56)	3.01 (0.10)	83.17 (0.45)	2.65 (0.09)
		$[T/4]$	71.67 (0.70)	3.86 (0.16)	73.19 (0.69)	3.60 (0.14)	74.67 (0.61)	3.10 (0.11)	82.98 (0.40)	2.54 (0.08)
	σ_2	G $[T/2]$	60.14 (0.85)	3.78 (0.19)	63.25 (0.71)	3.46 (0.15)	68.32 (0.78)	3.25 (0.14)	73.43 (0.61)	2.82 (0.14)
		$[T/4]$	61.56 (0.87)	3.90 (0.18)	66.78 (0.74)	3.79 (0.16)	68.43 (0.70)	3.19 (0.12)	72.19 (0.63)	3.04 (0.13)
		$T(5)$ $[T/2]$	59.19 (0.96)	3.84 (0.21)	63.03 (0.83)	3.63 (0.18)	65.33 (0.79)	3.48 (0.17)	71.11 (0.65)	3.32 (0.15)
		$[T/4]$	56.54 (0.94)	3.76 (0.22)	60.03 (0.89)	3.64 (0.19)	63.48 (0.82)	3.37 (0.18)	70.18 (0.66)	3.17 (0.16)
ii)	σ_1	$G[T/2]$	69.17 (0.84)	3.84 (0.20)	74.13 (0.71)	3.58 (0.18)	78.02 (0.67)	3.42 (0.15)	81.78 (0.57)	3.11 (0.14)
		$[T/4]$	72.14 (0.86)	3.86 (0.23)	74.13 (0.73)	3.70 (0.17)	76.02 (0.65)	3.51 (0.18)	83.09 (0.51)	3.07 (0.15)
		$T(5)$ $[T/2]$	77.81 (0.88)	4.09 (0.23)	76.43 (0.76)	4.01 (0.19)	75.01 (0.71)	2.55 (0.16)	80.18 (0.63)	1.67 (0.16)
		$[T/4]$	74.28 (0.87)	4.08 (0.24)	73.78 (0.75)	3.92 (0.18)	73.02 (0.73)	3.51 (0.17)	80.90 (0.65)	3.27 (0.17)
	σ_2	G $[T/2]$	62.13 (1.09)	4.35 (0.23)	63.67 (0.93)	4.07 (0.20)	65.02 (0.90)	3.89 (0.19)	68.45 (0.78)	3.49 (0.15)
		$[T/4]$	61.78 (1.11)	4.59 (0.29)	62.89 (1.08)	4.07 (0.23)	65.56 (0.97)	3.78 (0.21)	69.02 (0.79)	3.20 (0.16)
		$T(5)$ $[T/2]$	60.13 (1.15)	4.57 (0.28)	63.45 (1.01)	4.18 (0.24)	66.46 (0.93)	3.99 (0.22)	68.78 (0.80)	3.67 (0.18)
		$[T/4]$	60.56 (1.19)	4.78 (0.30)	63.09 (1.10)	4.19 (0.26)	65.87 (0.98)	3.98 (0.24)	66.11 (0.87)	3.85 (0.19)

4.1 Multiple Change Points

For the generality of our results, we formulate a numerical extension of our algorithm to test and estimate multiple change points. Also we propose a multiplier bootstrap procedure for the finite sample performance of our tests. The theoretical results of the multiple change points are highly involved and therefore out of the scope of this paper. We look at supreme Wald statistics to test the existence of a breakpoint. First, it can be shown that the statistics for testing the null hypothesis of the true coefficients in both regime are the same as $H_0 : H^\diamond = F^\diamond$ is
$\sup_{\tau\in[Td_0\%,T(1-d_0)\%]}\{S(\tau,\hat{H},\hat{F}) - \overline{S}(\hat{\overline{H}})\}(n-LM)/S(\tau,\hat{H},\hat{F})(LM)$ where $S(\tau,H,F) \overset{\text{def}}{=} \sum_{t=1}^{\tau}|D_TZ_t - \mathcal{B}(H)D_TZ_t|_2^2 + \sum_{t=\tau+1}^{T}|D_TZ_t - \mathcal{B}(\hat{F})D_TZ_t|_2^2$ and $\overline{S}(\hat{\overline{H}}) = \sum_{t=1}^{T}|D_TZ_t - \mathcal{B}(\hat{\overline{H}})D_TZ_t|_2^2$ are the estimated loss with breaks and without breaks respectively. It can be seen that if there is no break then the statistics should be small.

In the following, we show how to calculate the statistics and conduct a bootstrap procedure for calculating the critical value. Define for any $k_1 < k_2$, let

$$f(Z,k_1,k_2) = \begin{bmatrix} Z_{k_1-1} & Z_{k_1} & \dots & Z_{k_2-1} \\ Z_{k_1-2} & Z_{k_1-1} & \dots & Z_{k_2-2} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{k_1-M} & Z_{k_1-M+1} & \dots & Z_{k_2-M} \end{bmatrix} \in \mathbb{R}^{LM\times(k_2-k_1+1)}$$
$$\text{and} \quad h(Z,k_1,k_2) = \begin{bmatrix} Z_{k_1} & Z_{k_1+1} & \dots & Z_{k_2} \end{bmatrix} \in \mathbb{R}^{L\times(k_2-k_1+1)}.$$

Then we define the matrix of regressor in each regime and the whole period as $\hat{\Gamma}_1 = f(Z,1,\hat{\tau})$ and $\hat{\Gamma}_2 = f(Z,\hat{\tau},T)$, $\Gamma_{1,2} = f(Z,1,T)$. As the estimated parameters are perturbed by the sign matrix D_T with $D_T^{-1}\hat{H}_\tau\tilde{D}_T = \zeta_1\hat{\Gamma}_1^\top(\hat{\Gamma}_1\hat{\Gamma}_1^\top)^{-1}$, $D_T^{-1}\hat{F}_\tau\tilde{D}_T = (\zeta_2\hat{\Gamma}_2^\top)(\hat{\Gamma}_2\hat{\Gamma}_2^\top)^{-1}$ and let the projection matrix be $P_{\Gamma_{1,2}} = \Gamma_{1,2}^\top(\Gamma_{1,2}\Gamma_{1,2}^\top)^{-1}\Gamma_{1,2}$, $\mathcal{P}_{\Gamma_1} = \hat{\Gamma}_1^\top(\hat{\Gamma}_1\hat{\Gamma}_1^\top)^{-1}\hat{\Gamma}_1$. $S(\tau,\hat{H},\hat{F}) = |D_T\zeta_1(I_T - P_{\Gamma_1})|_F^2 + |D_T\zeta_2(I_T - P_{\Gamma_2})|_F^2$ and $\overline{S}(\hat{\overline{H}}) = |D_T\zeta_{1,2}(I - P_{\Gamma_{1,2}})|_F^2$.

Therefore we can express the difference of the loss in terms of estimator in each regime $S(\tau,\hat{H},\hat{F}) - \overline{S}(\hat{\overline{H}}) = \text{trace}((\hat{H}_\tau - \hat{F}_\tau)\tilde{D}_T\{(\Gamma_1\Gamma_1^\top)^{-1} + (\Gamma_2\Gamma_2^\top)^{-1}\}\tilde{D}_T^\top(\hat{H}_\tau - \hat{F}_\tau)^\top)$. Define the matrices of the errors as $\hat{U}_{\epsilon_1} = h(\hat{\epsilon},1,\hat{\tau})$ and $\hat{U}_{\epsilon_3} = h(\hat{\epsilon},\hat{\tau},T)$.

The multiplier bootstrap procedure is set on $\hat{U}_{\epsilon_1}, \hat{U}_{\epsilon_2}$ via $\zeta_1^* = D_T^{-1}\hat{E}_{\hat{\tau}}\tilde{D}_T\hat{\Gamma}_1 + U_{\epsilon_1}^*$, $U_{\epsilon_1}^{*b} = [U_{\epsilon_1,j,t}e_{j,t}^b]$, where $e_{j,t}^b$, $b = 1, \cdots B, j = 1, \cdots, N, t = 1, \cdots, T$ (the same for ζ_2^*, $\zeta_2^* = D_T^{-1}\hat{H}_{\hat{\tau}}\tilde{D}_T\hat{\Gamma}_2 + \hat{U}_{\epsilon_2}^*$) are i.i.d. random variables simulated from the standard normal distribution function.

We can then obtain $\tau^{*b}, \hat{F}_{\hat{\tau}^{*b}}^{*b}$, $\hat{E}_{\hat{\tau}^{*b}}^{*b}$ from the generated bootstrap sample ζ_1^*, ζ_2^*. We summarize the detailed implementation as follows, note that we work with $\hat{Z}_t$ replacing Step 3 and Step 4 in the algorithm on Iteration and Change Point in Section 4.

Algorithm on multiple change points

Step 1 Estimating a single break τ for the interval $[Td_0\%, T(1-d_0)\%]$ $(d_0 > 0)$, namely by minimizing over the interval,

$$(\hat{\tau}, \hat{H}, \hat{F}) = \text{argmin}_{\tau,H,F} S_{1:\tau}(H) + S_{(\tau+1):n}(F). \tag{4.5}$$

Step 2 Use the aforementioned multiplier bootstrap method to obtain $(\hat{\tau}^{*b}, \hat{H}^{*b}, \hat{F}^{*b})$, $b = 1, \cdots, B$, with B as the number of bootstrap samples.

Step 3 Calculate the bootstrap value of the test statistics for the breakpoint $W_b^* = \sup_{\tau\in[Td_0\%,T(1-d_0)\%]}\{\overline{S}^*(\hat{\overline{H}}^*) - S^*(\tau, \hat{H}^*, \hat{F}^*)\}(n-LM)/S^*(\tau, \hat{H}^*, \hat{F}^*)(LM)$. Obtain the critical values $q_{W,1-\alpha}^*$ as the $1-\alpha$ quantile of the bootstrap sample W_b^*. Also the $\alpha/2$ and $1-\alpha/2$ quantile of $(\hat{\tau}^{*b} - \hat{\tau})$ are denoted as $q_{\alpha/2,\tau}^*$ and $q_{1-\alpha/2,\tau}^*$.

Step 4 Reject the null hypothesis of $H = F$ if $\hat{W}_b > q_{W,1-\alpha}^*$. This means that we can find a break over the interval of interest. If we fail to reject, stop the algorithm and claim no change point in $[Td_0\%, T(1-d_0)\%]$.

Step 5 Proceed to the test procedure in Steps 1-4 on the half intervals of Step 4 if the test in Step 4 is significant, or else we terminate the procedure.

Similar to the above procedure, we obtain the bootstrap confidence interval for the change-point estimate $[\hat{\tau} - \lfloor q_{0.05,\tau}^* \rfloor - 1, \hat{\tau} + \lfloor q_{0.95,\tau}^* \rfloor + 1]$.

Table 4: Estimated RMSE and coverage probability ($\times 100$) of the three breaks $\hat{\tau}_i/T$, where G stands for the normal distribution and $T(3)$ stands for the t distribution. $N = 300, T = 300$. Nominal level $= 0.9$, averaged over 5000 samples.

	G		$T(3)$	
	RMSE	Cov. Prob.	RMSE	Cov. Prob.
$\hat{\tau}_1/T$	0.001	91.2	0.003	92.5
$\hat{\tau}_2/T$	0.003	89.3	0.005	88.6
$\hat{\tau}_3/T$	0.003	92.6	0.006	89.8

We consider the following data generating process for our multiple change-point algorithm. The Y_{it}s are generated by the same model as in equation (4.3), but instead we consider multiple breaks for Z_t. The time series $\{Z_t\}$ is taken to be a SBVAR process as in (2.2),
with $\tilde{E}_{(1)} = [0.6, 0.1, 0; 0, 0.3, -0.1; 0.01, 0, 0.2]$, $E_{(2)} = [0.6, 0.1, 0; 0, 0.3, -0.1; 0.01, 0, 0.2] - I_3 * T^{-1/3}$, $\tilde{E}_{(3)} = [0.5, -0.2, 0; 0, 0.8, 0.1; 0.1, 0, 0.6]$, and $\tilde{E}_{(4)} = [0.5, -0.2, 0; 0, 0.8, 0.1; 0.1, 0, 0.6] - I_3 * T^{-1/4}$. The ϵ_{tl}s are either i.i.d. normal random variables with standard deviation 0.001 before the break 1 and 0.01 after the break 1. The true breakpoints τ are taken to be $[T/4]$, $[T/2]$ and $[3T/4]$. In addition, we take case i) as in the previous single break case for the distribution of ε_{it} and $\sigma_1 = 0.1$. Table 4 shows the results on the multiple change-points estimation, the estimation accuracy stays the same as the single break case in the previous section.

5 Application

5.1 Minimum Wage Dataset for China

We consider a Chinese minimum wage dataset. It was collected from 1992 to 2012 for $N = 346$ counties over China and the corresponding Chinese regional economic statistics. The data have a yearly frequency with $T = 20$. The data source is the ministry of Human Resources and Social Security, and the Chinese Academy of Labor and Social Security. For more detailed descriptions, please refer to Huang et al. (2014). The minimum wage is set upon by the local government and the levels of the minimum wage may vary within a province. It is also known that after 2003 some changes have been made in the minimum wage adjustment policy. The dataset is matched with another one measuring the regional economic situation, namely the Chinese Statistical Year Book of China National Knowledge Infrastructure.

Why do we study this dataset using factor model? As can be seen, we have panel data situations with many counties, and the minimum wage policies may vary with respect to regional factors as well as common policy changes over time. Applying our SBDSFM model can help to disentangle the changes over the years and the time-invariant policy discrepancies over different regions. The interesting question is to check the geographic heterogeneity of the minimum wage policy, especially for the economically developed counties around the Pearl River Delta to the Yangtze River Delta; the minimum wage would be considerably different from the other regions. Thus, one would also be interested in modeling the location difference for the regions populated with minorities, such as Xinjiang or Tibet. It is in general a difficult task to jointly analyze the time-changing policy effect and the geographical discrepancies.

Therefore, we apply our estimation procedure as in Section 4. Y_{tj} is taken to be the minimum wage over the year. X_{tj1} is taken to be the difference of the countywise gross value added which measures the regional economics indicator, and X_{tj2} is taken to be scaled regional postal code. Figure 3 is from Huang et al. (2014), showing snapshots of the geographical distribution of minimum wage over the years. One sees that there are time changes and location discrepancies for the minimum wage policy in China. Figure 4 presents the estimated location loading functions in the left panel and the fitted time-varying factors

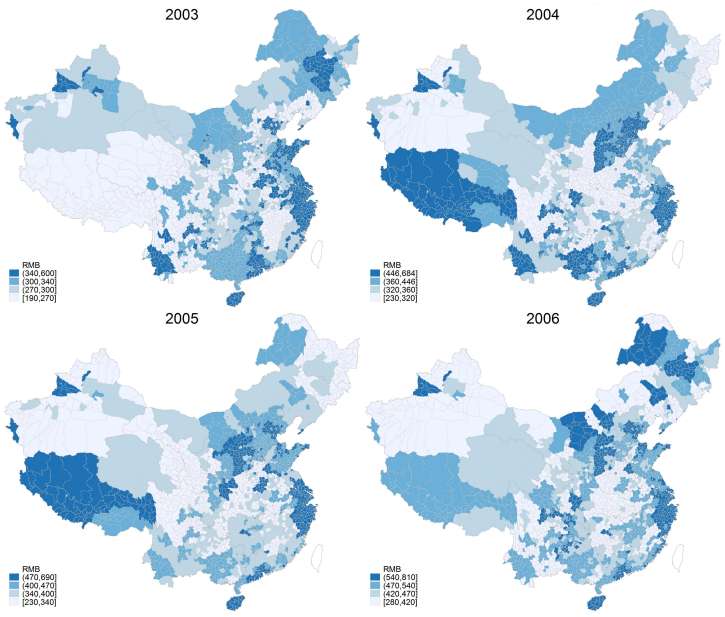


Figure 3: Graphical distribution of the minimum wage in China.

with estimated breakpoints and confidence intervals.

We select two factors with an explained variance (estimated proportion of variance of fitted values divided by estimated variance of the response variables) proportion as 0.5. The estimated structural changes for the two factors are closed to the year 2003. In particular, the two fitted factors follow a VAR(1) processes break and after the year of 2005 with the coefficients as $[0.5826, -0.6559; -0.0253, 0.4676]$ and $[0.7975, 0.2349; -0.7256, -0.0922]$ before and after the break respectively. The break size δ_e is estimated as 1.2821. The estimated confidence interval length is 5 years. It indicates that we can document a clear overall minimum wage policy change at the year 2005. This finding corresponds to the known fact of a common minimum wage policy change. We can see that our methodology is helpful to detect the common breakpoint in time for a dataset with a large cross-sectional dimension.

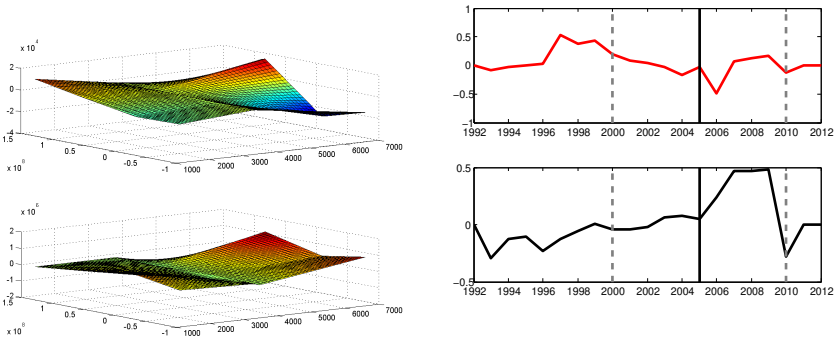


Figure 4: Plot of estimated $m_l(\cdot)$ functions and estimated Z_t, Y: minimum wage on X1: first difference of county-level gross value added and X2: city code. Breakpoint $\hat{\tau} = 2005$, and its confidence interval (dashed grey). Regional cutoff to isolate regions populated with ethnic minorities, Tibet, Xinjiang, Qinghai and Gansu and the economically developed regions.

5.2 Limit Order Book Volume Dataset

In this subsection, we illustrate our methodology using a limit order book dataset. For a specific stock, the limit order book is about the volume of pending buying or selling orders at certain price levels. Prices for the asset under consideration reflect a snapshot of the stock's demand and supply curves. The data are collected every 60 seconds at the NASDAQ stock market. The data source is from LOBSTER (lobsterdata.com), see Härdle et al. (2012) and Mihoci (2017) for more details on the data.

We consider a 60-second frequency over one day. Normal trading activities take place continuously on all stocks between 9:30 AM and 4:00 PM (i.e., 16:00) from Monday to Friday in NASDAQ with a total of $T = 390$ observations. For illustration, we take a one-day trading price as an example for four companies, namely Amazon and Facebook (on Sep 9th, 2016), and AT&T and Tesla (on June 1st, 2016). To show the data structure, the number of shares for the four stocks at times 10:00 AM and 10:30 AM are plotted in Figure 5. At each minute, ten (N = 10) price levels are collected from both the bid and the ask sides, with the first one being the lowest sell price and the last one being the highest sell price. As an example, at 10:00 AM on June 1, 2016, the trading volume is 1798 for Tesla and 216.78 USD is the second best ask price.

Our Y_{tj} is the trading volume at the tth minute and the jth price level, and X_{tj} is the ordered relative price level. As we measure the spread in relative terms, on the bid side, the price levels X_{tj} are divided by the highest bid price while on the ask side, the price levels X_{tj} are divided by the lowest price on the ask side. As we work with relative price levels, we do not consider to model the relative shift of the level of the curve. The connection point of the bid and ask curve are thus at a fixed point.

In Figure 7 and Figure 8, we show the $\hat{m}_l(\cdot)$s $(l = 1, 2)$ estimated with and without discontinuity. Note that the relative price level is considered and therefore, the breakpoint for $\hat{m}$ functions is always set to 0. $m_1(\cdot)$ represents the average level of trading volume in relationship to the relative price level, and $m_2(\cdot)$ corresponds more to the higher order structure of the curves. We also notice that the estimates with embedded discontinuity in 0 are quite different from the estimations without it. In Figure 6, the estimated two factors are plotted. We also plot estimated breakpoints and the confidence intervals built around them. It is worth noting that change-points happen at different time points for different stocks, and the width of the confidence interval also varies. This is due to stock specific latent trading dynamics. For AT&T, a change-point is detected at $14:40$, with an interval of 42 minutes; Tesla have a switch in its latent trading pattern at $10:40$ and $14:30$, with a $5-$ and $4-$ minute interval; for Facebook, changes happen at $12:03$ and $15:35$ with a $18-$ and $10-$ minute length of confidence interval; change-points are detected for Amazon at $10:50$ and $12:26$ with a $5-$ and $18-$minute interval. The fitted explained variance are respectively, 0.63, 0.75, 0.73, 0.56.

6 Conclusion and Further Work

In this paper we propose a dynamic semiparametric factor model with structural breaks. This approach contributes to the literature on change-point analysis in high-dimensional time series. We show good empirical performance in simulations and applications. We provide results on parameter consistency and we establish the asymptotic distribution of the estimated change point. Regarding future work, consideration of the selection of the number of factors using a $L-1$ regularization is another interesting direction to pursue.

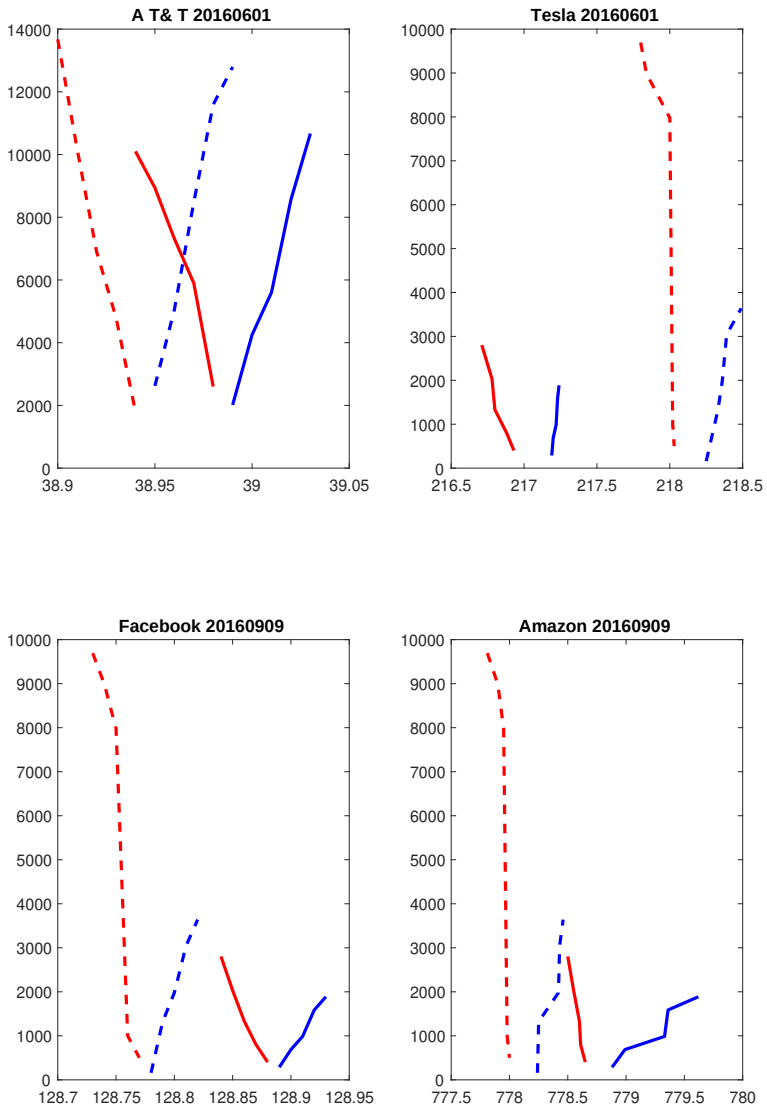
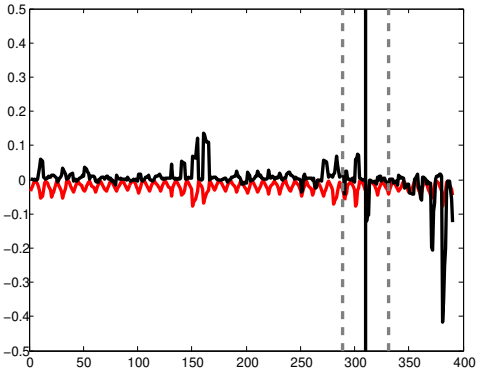
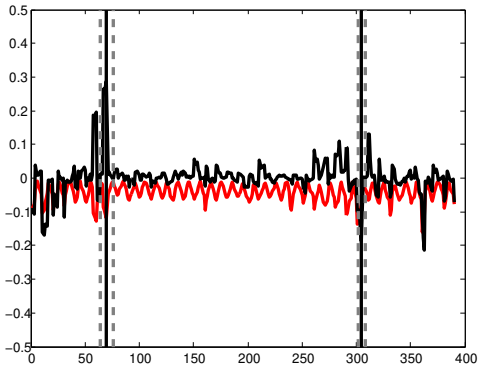


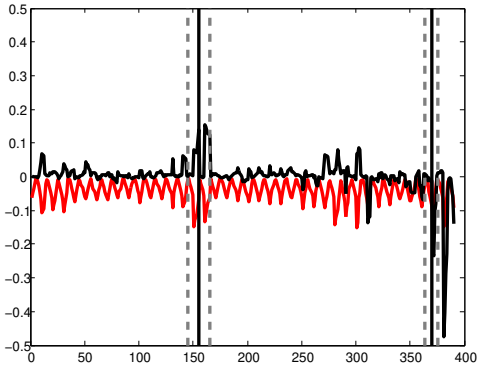
Figure 5: Plot of raw data for different companies at 10:00 AM (solid line), 10:30 AM (dotted line), on the bid side



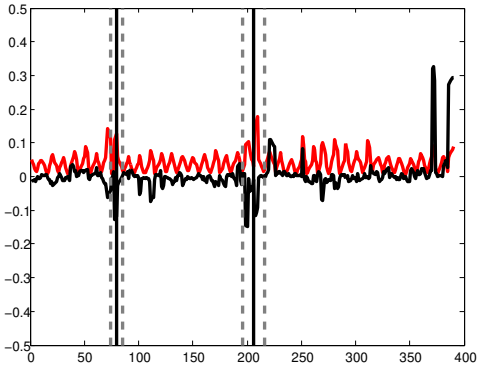
(a) AT& T



(b) Tesla



(c) Facebook



(d) Amazon

Figure 6: Plot of two factors and their estimated breakpoints(black lines), and their confidence intervals.

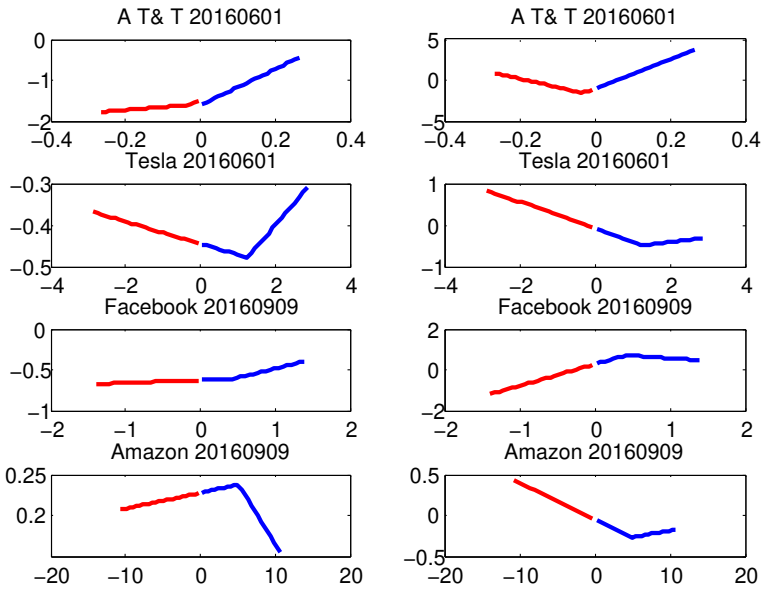


Figure 7: Plot of estimated loading functions $\hat{m}_1(\cdot)$ (left) and $\hat{m}_2(\cdot)$ (right, no breakpoint)

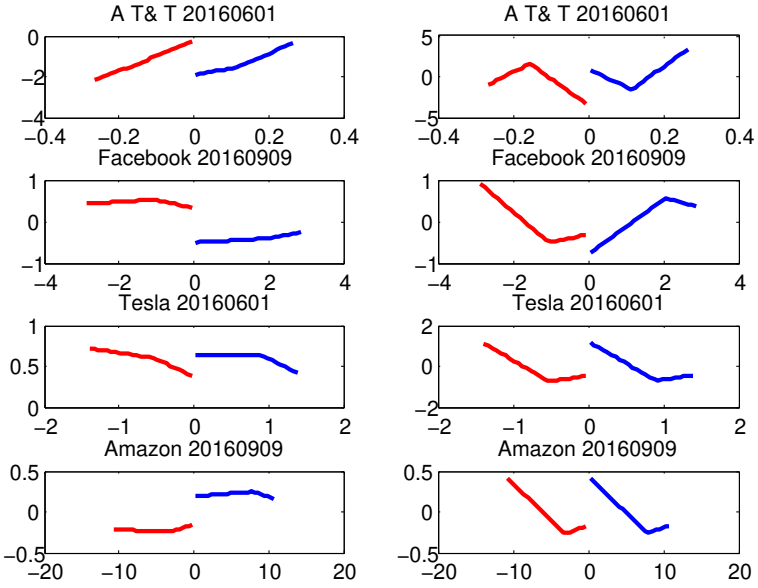


Figure 8: Plot of estimated loading functions $\hat{m}_1(\cdot)$ (left) and $\hat{m}_2(\cdot)$ (with breakpoint)

REFERENCES

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

Bai, J. (1995). Least absolute deviation estimation of a shift. *Econometric Theory 11*(3), 403–436.

Bai, J. (1997). Estimation of a change point in multiple regression models. *The Review of Economics and Statistics 79*(4), 551–563.

Bai, J. (2010). Common breaks in means and variances for panel data. *Journal of Econometrics 157*(1), 78–92.

Bai, J., X. Han, and Y. Shi (2018). Estimation and inference of change points in high dimensional factor models. *Available at SSRN 2875193*.

Bai, J. and S. Ng (2008). Recent developments in large dimensional factor analysis. Technical report, Working Paper, Mimeo.

Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 47–78.

Baltagi, B. H., Q. Feng, and C. Kao (2016). Estimation of heterogeneous panels with structural breaks. *Journal of Econometrics 191*(1), 176–195.

Brüggemann, R., W. Härdle, J. Mungo, and C. Trenkler (2008). VAR modeling for dynamic loadings driving volatility strings. *Journal of Financial Econometrics 6*(3), 361–381.

Cheng, X., Z. Liao, and F. Schorfheide (2016). Shrinkage estimation of high-dimensional factor models with structural instabilities. *The Review of Economic Studies 83*(4), 1511–1543.

Connor, G., M. Hagmann, and O. Linton (2012). Efficient semiparametric estimation of the Fama–French model and extensions. *Econometrica 80*(2), 713–754.

Dette, H. and D. Wied (2016). Detecting relevant changes in time series models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 78*(2), 371–394.

Fan, J., Y. Liao, and W. Wang (2016). Projected principal component analysis in factor models. *Annals of statistics 44*(1), 219.

Fengler, M. R., W. K. Härdle, and E. Mammen (2007). A semiparametric factor model for implied volatility surface dynamics. *Journal of Financial Econometrics 5*(2), 189–218.

Galvão, A. B. C. (2006). Structural break threshold VARs for predicting US recessions using the spread. *Journal of Applied Econometrics 21*(4), 463–487.

Härdle, W. K., N. Hautsch, and A. Mihoci (2012). Modelling and forecasting liquidity supply using semiparametric factor dynamics. *Journal of Empirical Finance 19*(4), 610–625.

Härdle, W. K. and P. Majer (2016). Yield curve modeling and forecasting using semiparametric factor dynamics. *The European Journal of Finance 22*(12), 1109–1129.

Huang, Y., P. Loungani, and G. Wang (2014). Minimum wages and firm employment: Evidence from China. *IMF working paper*.

Jirak, M. (2015). Uniform change point tests in high dimension. *The Annals of Statistics 43*(6), 2451–2483.

Ma, S. and L. Su (2016). Estimation of large dimensional factor models with an unknown number of breaks.

Manner, H., F. Stark, and D. Wied (2019). Testing for structural breaks in factor copula models. *Journal of Econometrics 208*(2), 324–345.

Mihoci, A. (2017). Modelling limit order book volume covariance structures. In *Advances in Statistical Methodologies and Their Application to Real Problems*. InTech.

Newey, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics 79*(1), 147–168.

Park, B. U., E. Mammen, W. Härdle, and S. Borak (2009). Time series modelling with semiparametric factor dynamics. *Journal of the American Statistical Association 104*(485), 284–298.

Preuß, P., R. Puchstein, and H. Dette (2015). Detection of multiple structural breaks in multivariate time series. *Journal of the American Statistical Association 110*(510), 654–668.

Scott, A. J. and M. Knott (1974). A cluster analysis method for grouping means in the analysis of variance. *Biometrics*, 507–512.

Shao, X. and X. Zhang (2010). Testing for change points in time series. *Journal of the American Statistical Association 105*(491), 1228–1240.

Stock, J. H. and M. W. Watson (2011). Dynamic factor models. *Oxford Handbook of Economic Forecasting 1*, 35–59.

Stryhn, H. (1996). The location of the maximum of asymmetric two-sided Brownian Motion with triangular drift. *Statistics & Probability Letters 29*(3), 279 – 284.

Trück, S., W. Hardle, and R. Weron (2014). The relationship between spot and futures CO2 emission allowance prices in the EU-ETS.

van Bömmel, A., S. Song, P. Majer, P. N. Mohr, H. R. Heekeren, and W. K. Härdle (2014). Risk patterns and correlated brain activities. multidimensional statistical analysis of fmri data in economic decision making study. *Psychometrika 79*(3), 489–514.

Wied, D., W. Krämer, and H. Dehling (2012). Testing for a change in correlation at an unknown point in time using an extended functional delta method. *Econometric Theory 28*(3), 570–589.

Wu, W. B. and Z. Zhao (2007). Inference of trends in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69*(3), 391–410.