

SFB 649 Discussion Paper 2017-026

Dynamic Semiparametric Factor Model with a Common Break

Likai Chen*
Weining Wang*²
Wei Biao Wu*³



* University of Chicago, United States of America

*² Humboldt-Universität zu Berlin, Germany

*³ City, University of London, United Kingdom

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

<http://sfb649.wiwi.hu-berlin.de>
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin



SFB 649 ECONOMIC RISK BERLIN

Dynamic Semiparametric Factor Model with a Common Break

Likai Chen¹, Weining Wang^{2,3}, and Wei Biao Wu¹

¹Department of Statistics, University of Chicago

² C.A.S.E.-Center of Applied Statistics and Economics, Humboldt-Universität zu Berlin

³Department of Economics, City, U of London

Abstract

For change-point analysis of high dimensional time series, we consider a semiparametric model with dynamic structural break factors. The observations are described by a few low dimensional factors with time-invariant loading functions of covariates. The unknown structural break in time models the regime switching effects introduced by exogenous shocks. In particular, the factors are assumed to be nonstationary and follow a Vector Autoregression (VAR) process with a structural break. 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 convergence results for making inference on the common breakpoint in time. The estimation precision is evaluated via a simulation study. Finally we present two empirical illustrations on 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 dependence, vector autoregressive process

*We acknowledge the research support from German Research Foundation (DFG) on "quantile method for complex financial system" and IRTG 1792.

1 Introduction

Emerging situations of big data call for statistical tools to learn the intrinsic complex structure. Modeling the structural break for high dimensional time series is our object to study. A very popular approach is to consider high dimensional time series with factor structures, 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. Emphatically, [Connor et al. \(2012\)](#) and [Fan et al. \(2016\)](#) also consider a similar model majorly 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 it takes full advantage of the information provided by large cross-section and time-series dimensions. For example, DSFM has been applied by [Härdle et al. \(2012\)](#) in modeling and forecasting the limit order book dynamics, [Fengler et al. \(2007\)](#) in describing implied volatility surface dynamics, [van Bömmel et al. \(2014\)](#) in discovering the risk patterns and brain activities, [Härdle and Majer \(2016\)](#) in yield curving modeling, [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 \(1998\)](#) on complex regression models; [Wied et al. \(2012\)](#) and [Preuß et al. \(2015\)](#) on the second-order characteristics of a time series; [Dette and Wied \(2016\)](#) on 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 CUSUM statistics. Notably for factor models, there are also many recent articles on estimating loading changes, such as Cheng et al. (2016) using shrinkage methods and Bai et al. (2016) focusing on establishing the asymptotic distribution of the LS estimator for a structural break.

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 at fitting a structural break VAR process. This brings both new theoretical and empirical perspectives in change-point analysis in semiparametric time series models. Importantly, we establish the asymptotic distribution of our break point by allowing for general temporal and cross-sectional dependence in the error terms. For example, an application is regarding limit order book data modeling, 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 their dependency structure over time.

Our model is thus named as a structural break DSFM (SBDSFM), as we assume that factors follow a structural break vector autoregression model (SBVAR). SBVAR is applied to change-point analysis for 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 does 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. Using SBDSFM allows us therefore to account for the structural breaks in time and cross-section dimension. We show that the model is helpful for detecting and making inference on the latent structural change for high dimensional time series.

We contribute the literature in three aspects. Firstly, we propose a SBDSFM model with an embedded

break structure, and we also study its identification scheme and estimation method. Secondly, we show the consistency of the semiparametric estimates. And for the latent factors, a consistency and distribution theorem allow us to make inference on the breakpoint. Thirdly, we illustrate the good empirical performance of our SBDSFM model via simulations and empirical examples. Our paper is organized as follows, see Section 2 for the model description and estimation, Section 3 for the relevant theorems and discussions. Simulations results are shown in Section 4. Section 5 consists of two applications. The technical details are delegated to the Appendix.

2 Model

In this section, we lay down the general model setup. First of all we list the necessary math notation used throughout the paper. For matrix A , 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 $v = (v_1, \dots, v_d)^\top \in \mathbb{R}^d$ let $|v|_k = (\sum_{i=1}^d |v_i|^k)^{1/k}$ and $|v|_\infty = \max_{i \leq d} |v_i|$. For two positive sequences of numbers (a_n) and (b_n) , 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 $a_n/b_n \leq C$ (resp. $1/C \leq a_n/b_n \leq C$) for all large n , and denote $a_n = o(b_n)$ or $a_n \ll b_n$ (resp. $a_n \sim b_n$), if $a_n/b_n \rightarrow 0$ (resp. $a_n/b_n \rightarrow 1$). For two sequences of random variables (X_n) and (Y_n) , write $X_n = o_{\mathbb{P}}(Y_n)$, if $X_n/Y_n \rightarrow 0$ in probability. Let $\lambda_i(\cdot)$ be the i th largest eigenvalue, $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ be the minimum and maximum eigenvalues respectively.

2.1. Dynamic Semiparametric factor models (DSFMs). Let $\varepsilon_t = (\varepsilon_{t,1}, \varepsilon_{t,2}, \dots, \varepsilon_{t,L})^\top$, $t = 1, 2, \dots, T$, be a sequence of random vectors and the model we consider is $(N, T \rightarrow \infty)$

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

Here $Z_t = (Z_{t,1}, Z_{t,2}, \dots, Z_{t,L})^\top$ can be understood as the common factors, the term $m_0(X_{t,j})$ can be interpreted as the conditional expectation of the individual specific effects and $m_l(X_{t,j}), l \geq 1$, are taken to be generalized nonparametric loading functions. Similar models 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.

To incorporate both temporal and spatial dependencies, we shall consider the commonly used moving average (MA) process for the noise sequence $\varepsilon_t = (\varepsilon_{t,1}, \dots, \varepsilon_{t,N})^\top$,

$$\varepsilon_t = \sum_{k \geq 0} B_k \eta_{t-k}, \quad (2.2)$$

where $\eta_t = (\eta_{t,1}, \eta_{t,2}, \dots, \eta_{t,p})^\top$ with $\eta_{t,j}, t, j \in \mathbb{Z}$, being independent and identically distributed (i.i.d.) random variables with zero mean and unit variance, and $(B_k)_{k \geq 0}$ are matrices in $\mathbb{R}^{p \times p}$ such that ε_t is a proper random vector. If $B_k = 0$ for all $k \geq 1$, then the noise sequences are temporally independent, and 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 . The MA(∞) process is very widely used in practice and it includes many important time series models such as vector autoregressive moving averages (VARMA)

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

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 $|z| \leq 1$.

2.2. Temporal and cross-sectional breaks. To incorporate the temporal break, consider a time break point τ^\diamond . Let $\epsilon_t = (\epsilon_{t,1}, \epsilon_{t,2}, \dots, \epsilon_{t,L})$, $t = 1, \dots, T$, be i.i.d. random vectors. For $A = (A_1, A_2, \dots, A_M)$, denote $\mathcal{B}(A) = A_1 \mathcal{B}^1 + A_2 \mathcal{B}^2 + \dots + A_M \mathcal{B}^M$, where \mathcal{B} is the backward shift operator. Assume the factors Z_t satisfy

$$Z_t = \mathcal{B}(E) Z_t \mathbf{1}_{t \leq \tau^\diamond} + \mathcal{B}(\tilde{E}) Z_t \mathbf{1}_{t > \tau^\diamond} + \epsilon_t, \quad t \geq 1, \quad (2.3)$$

and $Z_0, Z_{-1}, \dots, Z_{-M+1}$ are any vectors in \mathbb{R}^L , where $E = (E_1, E_2, \dots, E_M)$, $\tilde{E} = (\tilde{E}_1, \tilde{E}_2, \dots, \tilde{E}_M)$ and τ^\diamond is an unknown constant. Here the lags for the two regimes, denoted as M_1 and M_2 , can be different or unknown by letting M large enough and $E_i = 0$, $\tilde{E}_j = 0$ for $i > M_1$, $j > M_2$.

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_t s follow a nonstationary SBVAR process, namely the coefficients of VAR changes after an unknown break point, 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 $\varepsilon_{t,j}$. \square

It should be noted that the number of factors stays the same throughout the model. The spatial discontinuity can be handled by taking into account discontinuous bases for $m_l(\cdot)$, and $m_l(\cdot)$ can be approximated by $\sum_{k=1}^J a_{l,k} \phi_k(\cdot)$, where $\phi_k(\cdot)$ is taken to be $\psi_i(\cdot) \mathbf{1}_{\in R_r}$ where R_r s form the space of C in the whole support of $X_{i,t}$. $\psi_i(\cdot)$ are tensor product B-spline basis, and i, r corresponds to the index k . For example, R_r can be used to model the presence of discontinuities of regional minimum wage policies in China. Define the matrix of coefficients $A \stackrel{\text{def}}{=} (a_{l,k})_{l,k}$ (for every factor we have chosen the same number J of basis $\{\phi_k\}$). Denote matrix $\Phi(X_t) = (\phi(X_{t,1}), \phi(X_{t,2}), \dots, \phi(X_{t,N}))^\top \in \mathbb{R}^{N \times J}$, where $\phi(x) = (\phi_1(x), \phi_2(x), \dots, \phi_J(x))^\top$. Let $m(x) = (m_0(x), m_1(x), \dots, m_L(x))^\top$.

2.1 Estimation

With the model [2.1](#) on hand we can estimate A and the dynamics of Z_t according to the following steps.

Step 1. We obtain a group of $\hat{A}, \hat{Z}_t, 1 \leq t \leq T$, by minimizing $h(A, z_1, z_2, \dots, z_T) = \sum_{t=1}^T |Y_t - \Phi(X_t)A(1, z_t^\top)^\top|_2^2$

$$(\hat{A}, \hat{Z}_t, 1 \leq t \leq T) = \operatorname{argmin}_{A, z_t, 1 \leq t \leq T} h(A, z_1, z_2, \dots, z_T). \quad (2.4)$$

It is not hard to see that the minimum point is not unique. More specifically, for any minimum point $(\hat{A}^0, \hat{Z}_t^0, 1 \leq t \leq T)$, 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 $\hat{Z}^0 = (\hat{Z}_1^0, \hat{Z}_2^0, \dots, \hat{Z}_T^0)$, the product $\hat{A}_2^0 \hat{Z}^0$ is unique. Finding $\min h(A, z_1, z_2, \dots, z_T)$ is non-trivial, since it involves a fourth-order problem. In practice, one may follow a Newton-Raphson method proposed in [Park et al. \(2009\)](#).

Step 2. Let $H = (H_1, H_2, \dots, H_M)$, $F = (F_1, F_2, \dots, F_M)$ be the coefficient matrices, τ be the change point.

Consider

$$\hat{S}(\tau, H, F) = \sum_{t=1}^T |\hat{Z}_t - \mathcal{B}(H)\hat{Z}_t\mathbf{1}_{t \leq \tau} - \mathcal{B}(F)\hat{Z}_t\mathbf{1}_{t > \tau}|_2^2. \quad (2.5)$$

Let $(\hat{\tau}, \hat{H}, \hat{F})$ be the minimizer for $\hat{S}(\tau, H, F)$ as the estimates of the parameters. Denote $\hat{V}_\tau = \min_{H, F} \hat{S}(\tau, H, F)$. Then $\hat{\tau} = \operatorname{argmin}_\tau \hat{V}_\tau$, and $(\hat{H}, \hat{F}) = \operatorname{argmin}_{H, F} \hat{S}(\hat{\tau}, H, F)$.

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 on the breakpoints. We consider $\min(N, T) \rightarrow \infty$ asymptotically. The relative rate of N, T is discussed in Remark 5. First we list a few assumptions.

3.1 Assumptions

ASSUMPTION 3.1. (*Properties of ε_t*) Assume that the noise vectors in our SBDSFM ε_t , $1 \leq t \leq T$, satisfy the MA(∞) model in (2.2) with the innovation sequence $(\eta_{i,j})$ and coefficient matrices (B_i) .

Moment. Let $(\eta_{i,j})$ be i.i.d. random variables with zero mean and finite q th moment, $q \geq 4$, i.e. $\|\eta_{1,1}\|_q = \mathbb{E}(|\eta_{1,1}|^q)^{1/q} < \infty$. Denote $\mu_r = \|\eta_{1,1}\|_r$, for any $r \leq q$.

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

ASSUMPTION 3.2. (*Basis function*) For $1 \leq t \leq T$, assume $X_{t,i}$, $i = 1, \dots, N$ are i.i.d random variables, independent of ε_t and ε_t . Assume that basis functions ϕ_j , $j = 1, \dots, J$, are bounded in absolute

value by $c_\phi < \infty$, and

$$l_\phi \leq \lambda_{\min}(\mathbb{E}\phi(X_{t,1})\phi(X_{t,1})^\top) \leq \lambda_{\max}(\mathbb{E}\phi(X_{t,1})\phi(X_{t,1})^\top) \leq u_\phi,$$

where $l_\phi, u_\phi > 0$ are some finite constants, and recall that $\phi(X_{t,1})$ is a $J \times 1$ vector of basis function values.

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 zero mean and $\max_{1 \leq i \leq L} \|\epsilon_{0,i}\|_{q'} < \infty$, for some $q' \geq 4$. And (ϵ_t) are independent of (ε_t) . For the covariance matrix $\Sigma_\epsilon = \mathbb{E}(\epsilon_0\epsilon_0^\top)$, assume $l_\epsilon = \lambda_{\min}(\Sigma_\epsilon) > 0$.

Note when $\epsilon_{t,i}$ are i.i.d for different $1 \leq i \leq L$, with zero mean 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 τ^\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 $E = (E_1, E_2, \dots, E_M)$, $\tilde{E} = (\tilde{E}_1, \tilde{E}_2, \dots, \tilde{E}_M)$,

$$(i) \sum_{i=1}^M |HE_i H^{-1}|_2, \sum_{i=1}^M |H\tilde{E}_i H^{-1}|_2 \leq \gamma_e < 1, \text{ for some constant } \gamma_e > 0.$$

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

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

$$\mathcal{Z}_t^{(l)} = (I - \mathcal{B}(E))^{-1}\epsilon_t, \quad \mathcal{Z}_t^{(r)} = (I - \mathcal{B}(\tilde{E}))^{-1}\epsilon_t \text{ and } \mathcal{Z}_t = \mathcal{Z}_t^{(l)}\mathbf{1}_{t \leq \tau^\diamond} + \mathcal{Z}_t^{(r)}\mathbf{1}_{t > \tau^\diamond}. \quad (3.1)$$

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. We allow δ_e to go to 0, when there is no structural break and the model boils down to the stationary DSFM. Denote covariance matrices

$$W_1 = \mathbb{E}\mathcal{Z}_0^{(l)}\mathcal{Z}_0^{(l)\top}, \quad W_2 = \mathbb{E}\mathcal{Z}_0^{(r)}\mathcal{Z}_0^{(r)\top} \text{ and } W_0 = T^{-1} \sum_{t=1}^T \mathbb{E}\mathcal{Z}_t\mathcal{Z}_t^\top = W_1\tau^\diamond/T + W_2(T - \tau^\diamond)/T. \quad (3.2)$$

For \mathbb{R}^{LM} vectors

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

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

$$\Sigma^{(l)} = \mathbb{E}\xi_0^{(l)}\xi_0^{(l)\top} \text{ and } \Sigma^{(r)} = \mathbb{E}\xi_0^{(r)}\xi_0^{(r)\top}, \quad (3.3)$$

which capture the autocovariance of \mathcal{Z}_t up to M th 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 (3.2),

$$\lambda_{\min}(W_0) \geq l_\epsilon, \quad (3.4)$$

(ii) for matrices $\Sigma^{(l)}$ and $\Sigma^{(r)}$ defined in (3.3),

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

□

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 . Note $(I - \mathcal{B}(E))^{-1} = I + f_1 B^1 + f_2 B^2 + \dots$, where f_i s 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 $\lambda_{\min}(W_1) \geq \lambda_{\min}(\Sigma_\epsilon) \geq l_\epsilon$. Same arguments can be applied for W_2 .

For part (ii), let $x = (x_1^\top, x_2^\top, \dots, x_M^\top)^\top$, 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 $Z_i^{(l)} = \epsilon_i + \sum_{k \geq 1} C_k \epsilon_{k-i}$, some matrices C_k , we have

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

where $H = \sum_{i=1}^M \sum_{k \geq 1} x_i^\top H_k \epsilon_{i^*-k}$, 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. □

ASSUMPTION 3.6. (*Loadings*) Assume $J \leq c_1 N^{\alpha_J}$, some $\alpha_J < 1/2$. For any $j \geq L$, there exists $A^* \in \mathbb{R}^{J \times L}$ and $\beta_J > 0$, such that,

$$(i) \quad \delta_J = \sup_{x \in [0,1]} |m(x)^\top - \phi(x)^\top A^*|_\infty = O(J^{-\beta_J}).$$

(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 $\text{gap}_L(A_2^* W_0 A_2^{*\top}) \geq l'_a > 0$, where 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. Quantity δ_J specifies the approximation speed, which is of some polynomial order and it is also considered as the order of bias for our semiparametric estimation. We also require 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.

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 diagonal entities.

We shall show that the above conditions can be obtained under Assumptions 3.1-3.6. 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 diagonal entities. By Remark 3, W_0 is invertible. Denote $D = Q^\top W_0^{-1/2}$. For $Z'_t = D Z_t$ and $(m'_1(\cdot), m'_2(\cdot), \dots, m'_L(\cdot)) = (m_1(\cdot), m_2(\cdot), \dots, m_L(\cdot)) D^{-1}$, $A_2^{*'} = A_2^* D^{-1}$ and (2.3) 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 $\epsilon'_t = D \epsilon_t$, $E'_i = D E_i D^{-1}$ 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 $H' = H D^{-1}$,

$\sum_{i=1}^M |H' E_i' H'^{-1}|_2 < 1$ and thus we have Assumption 3.5. For $A_2^{*'} = A_2^* D^{-1}$, since $\lambda_{\min}(D) > 0$, Assumption 3.6 holds. By (3.2), $W_0' = DW_0 D^{-1} = I_L$, 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 is concerning the consistency results of the estimated parameters of SBDSFM, and Theorem 2 is on the consistency of the breakpoint estimate. We show that the identified object $A^*(1, Z_t^\top)^\top$ 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.

THEOREM 1. *(Consistency of \hat{A} and \hat{Z}_t) Under Assumptions 3.1-3.7.*

Denote $\rho^2 = (T + J)(TN)^{-(1-2/q)} \log(TN)$ and assume $\rho \rightarrow 0$. Then

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

(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*

$$|A_2^* D_T^{-1} - \hat{A}_2|_F = O_{\mathbb{P}}(\rho + \delta_J) \quad \text{and} \quad T^{-1} \sum_{t=1}^T |D_T Z_t - \hat{Z}_t|_2^2 = O_{\mathbb{P}}(\rho^2 + \delta_J^2).$$

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 slower rate of convergence. \square

For the estimation of the break point, we have the following assumption.

ASSUMPTION 3.8. *Assume $\rho^2 + \delta_J^2 = o(T^{-1})$.*

REMARK 5. Let $N \asymp T^r$. Then under Assumption 3.6 (i), Assumption 3.8 holds if

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

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 . \square

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 δ_e fixed or $\delta_e \rightarrow 0$ and $T^{1/2}\delta_e \rightarrow \infty$, we have $|\hat{\tau} - \tau^\diamond| = O_{\mathbb{P}}(\delta_e^{-2})$.*

The statement implies that $T^{-1}|\hat{\tau} - \tau^\diamond| \rightarrow 0$ in probability. The rate of consistency is determined by the magnitude of the change δ_e .

3.3 Asymptotic Distribution

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

Define the block matrix $\tilde{D}_T = I_M \otimes D_T$ where \otimes is denoted as the Kronecker product between two matrices. Define the true coefficient matrix scaled by D_T and \tilde{D}_T as $H^\diamond = D_T E \tilde{D}_T^{-1}$, $F^\diamond = D_T \tilde{E} \tilde{D}_T^{-1}$. 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 (3.3). The asymptotic normality of the estimated coefficient matrices is shown below.

THEOREM 3 (Central limit theorem for within regime parameters). *Under Assumptions 3.1-3.8. Let $\Theta^{(l)} = (N_{i,j}^{(l)})_{1 \leq i,j \leq L}$ 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 $\Theta^{(r)} = (N_{i,j}^{(r)})_{1 \leq i,j \leq L}$ with $N_{i,j}^{(r)} = \Sigma_{\epsilon,i,j} \Sigma^{(r)-1}$. Then*

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

and

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

REMARK 6. 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 . \square

Then we provide a theorem on the asymptotic distribution of the change point estimate $\hat{\tau}$. Considering 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 3.1-3.8 and additionally $\delta_e \rightarrow 0$. 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

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

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

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

REMARK 7. 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. \square

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

$$[\hat{\tau} - \lfloor \hat{q}_{1-\alpha/2} \rfloor - 1, \hat{\tau} + \lfloor \hat{q}_{\alpha/2} \rfloor + 1], \quad (3.5)$$

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

Denote $q_l = \operatorname{tr}(\mathcal{Q}^{(l)})$, $\sigma_l = 2\operatorname{tr}^{1/2}(\mathcal{Q}^{(l)}\Sigma_\epsilon)$, $q_r = \operatorname{tr}(\mathcal{Q}^{(r)})$ and $\sigma_r = 2\operatorname{tr}^{1/2}(\mathcal{Q}^{(r)}\Sigma_\epsilon)$. Let $\theta_1 = q_l/\sigma_l$ (q_r/σ_r) for $t \leq 0$ ($t > 0$) and $\theta_2 = q_r\sigma_l/\sigma_r^2$ ($q_l\sigma_r/\sigma_l^2$) for $t \leq 0$ ($t > 0$). Denote

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

Then according to [Stryhn \(1996\)](#), the distribution function of $\operatorname{argmin}_s H(s)$ is of the following form.

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

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

4 Simulation

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

Initial Value Selection \hat{Z}^0 and \hat{A}^0 . Denote $\Psi_t \stackrel{\text{def}}{=} (m_l(X_{i,t}))_{1 \leq i \leq N, 1 \leq l \leq L}$. The initial estimation of $Z(T \times L)$ and $A(J \times L)$ can be obtained as follows.

Step 1 We estimate firstly $\Gamma_t \stackrel{\text{def}}{=} A(1, Z_t^\top)^\top$ and let $\hat{\Gamma}_t^0 = \operatorname{argmin}_\Gamma S(\Gamma) = |Y_t - \Phi(X_t)\Gamma|_2^2 = |Y_t - \Phi(X_t)A(1, Z_t^\top)^\top|_2^2 = \{\Phi^\top(X_t)\Phi(X_t)\}^{-1}\Phi^\top(X_t)Y_t$. Define $\hat{\Gamma}^0 = (\hat{\Gamma}_1^0, \hat{\Gamma}_2^0, \dots, \hat{\Gamma}_T^0)_{J \times T}$.

Step 2 Denote the condensed SVD of $\hat{\Gamma}^0$ as $\Gamma^0 = \hat{U}^0 \hat{\Lambda}^0 \hat{V}^{0\top}$, where $\hat{\Lambda}^0 = \operatorname{diag}(\hat{\lambda}_1^0, \hat{\lambda}_2^0, \dots, \hat{\lambda}_L^0)$ (The first L largest singular values.) and $\hat{\lambda}_1^0 \geq \hat{\lambda}_2^0 \geq \dots \geq \hat{\lambda}_L^0$. Set $\hat{Z}^0 = \hat{\Lambda}^0 \hat{V}^{0\top}$, and $A^0 = \hat{U}^0$.

Step 3 Select the number of factors according to a BIC or AIC criteria.

The initial step is a projection based approach. The number of factors is prefixed by the initial selection stage. We focus on the cases of having the minimum fixed number of factors following the parsimonious principle (taking the minimum of BIC and AIC).

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 $Z \stackrel{\text{def}}{=} (Z_1, Z_2, \dots, Z_T)$ following the loss:

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

Step 2 \hat{A} gives us estimates of factors loadings $\hat{m}_l(\cdot)$.

Step 3 Assuming Z_t s follows a SBVAR process in (2.3), plug in (4.1).

Step 4 Apply a binary segmentation algorithm as in Scott and Knott (1974) for estimating the single break τ , namely by minimizing

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

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

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

N, T have the following cases $T = 50, 100, 200, 250$, $N = 50, 100, 200, 300$. Each element of X_{it} is taken to be uniformly distributed over $[-3, 3]$. Let $m_0(x_1, x_2) = 0$. The $m_l(\cdot, \cdot)$ s are taken to be

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

where a_1, a_2, b_1, b_2 are taken to be either 0.5 or 0.7.

The time series $\{Z_t\}$ is taken to be a SBVAR process as in (2.3), with $E = [0.5, -0.2, 0; 0, 0.8, 0.1; 0.1, 0, 0.6]$, and $\tilde{E} = [0.5, -0.2, 0; 0, 0.8, 0.1; 0.1, 0, 0.6]$. And ε_{it} 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 (5 degree of freedom). The true break point τ is taken to be $[T/2]$ or $[T/4]$. In addition, ε_{it} s are set to be i) independent

standard normal noise processes $N(0, 0.1)$, ii) independent ARMA(1,1) processes, $\varepsilon_{it} = 0.5\varepsilon_{i(t-1)} + \eta_{it} + 0.5\eta_{i(t-1)}$, 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 $\sigma_2 = 0.01$. In addition, we allow for the serial correlations for the error processes ε_{it} s by case ii). Figure 1 presents the $m_l(\cdot, \cdot)$ function under the case of $T = 200$, $N = 200$, $a_1 = 0.5$, $a_2 = 0.5$, $b_1 = 0.5$, $b_2 = 0.5$. 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 variances of the factors.

For the estimation of basis functions we consider the tensored quadratic B splines keeping 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 an invertible matrix D_T . Denote the centered $\hat{Z}_{t,c}$ as $\hat{Z}_{t,c} = \hat{Z}_t - T^{-1} \sum_{t=1}^T \hat{Z}_t$, 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 $\hat{D}_T = (\sum_t Z_{t,c} Z_{t,c}^\top)^{-1} (\sum_t Z_{t,c} \hat{Z}_{t,c})$. 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.

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

For evaluating the accuracy of the estimation, the confidence intervals of the estimated change point as in (3.5) is implemented and Table 1 reports the estimated coverage probabilities over 1000 samples in different simulation scenarios. Also Table 2 presents the explained variances of 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, innovations processes and signal to noise ratios. Moreover, we have also shown good recovery rates of the break point over time. When the sample size increases, one sees a tendency of an overall better performance.

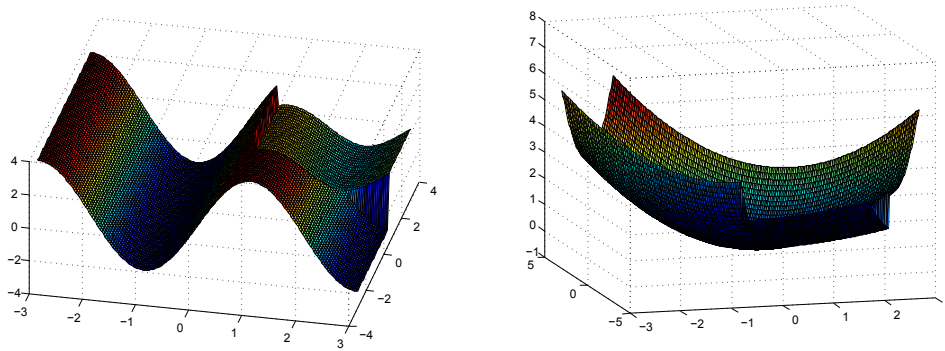


Figure 1: Plot of true loading functions with $T = 200$, $N = 200$, $a_1 = 0.5$, $a_2 = 0.5$, $b_1 = 0.5$, $b_2 = 0.5$.

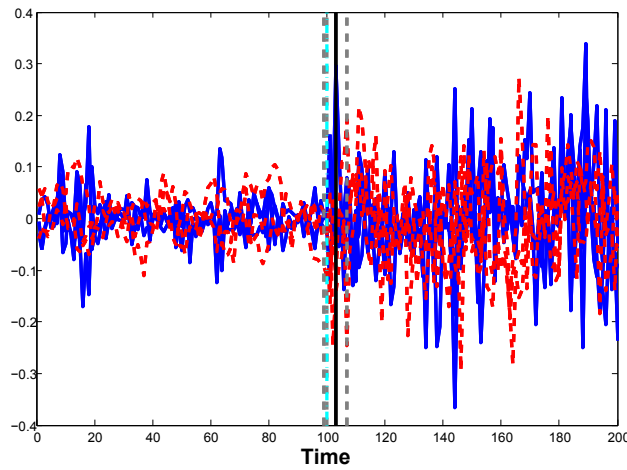


Figure 2: Plot of simulated normalized true factors with $T = 200$, $N = 200$, break point $\tau = 100$ (cyan). Estimated break point $\hat{\tau} = 103$ (black), and estimated confidence region $[99, 107]$ (dotted grey).

Table 1: The coverage probability is $\times 10^2$, G denotes normal innovations and T denotes t location scale distributions (with 5 degree of freedom) innovations, $\tau = [T/2]$ or $[T/4]$. Average over 1000 samples. $\alpha = 0.05$. 50, 100 means $N = 50, T = 100$, and the same for others.

			50, 100	100, 50	200, 200	300, 250
i)	σ_1	$G [T/2]$	60.1	64.2	76.3	77.8
		$[T/4]$	62.1	63.3	73.3	78.2
	T	$[T/2]$	58.7	61.9	74.7	74.8
		$[T/4]$	64.3	65.5	71.4	76.8
	σ_2	$G [T/2]$	72.3	74.6	84.2	87.8
		$[T/4]$	69.2	68.2	83.1	84.6
	T	$[T/2]$	66.1	62.1	75.7	85.5
		$[T/4]$	67.1	69.2	75.1	84.6
ii)	σ_1	$G [T/2]$	59.6	64.0	75.8	75.2
		$[T/4]$	59.3	61.8	74.2	73.8
	T	$[T/2]$	58.3	56.2	74.8	77.8
		$[T/4]$	59.9	59.2	75.5	78.8
	σ_2	$G [T/2]$	69.1	72.5	83.3	85.8
		$[T/4]$	69.3	68.6	79.2	84.2
	T	$[T/2]$	65.1	64.2	73.9	83.1
		$[T/4]$	56.3	63.1	77.3	82.6

Table 2: 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 1000 samples. Exp. var. denotes explained variance.

		$N = 50, T = 100$		$N = 100, T = 50$		$N = 200, T = 200$		$N = 300, T = 250$	
		Exp. var.	e_f	Exp. var.	e_f	Exp. var.	e_f	Exp. var.	e_f
i)	σ_1								
	G	70.03 (0.33)	1.65 (0.03)	74.13 (0.25)	1.60 (0.03)	78.02 (0.24)	1.48 (0.02)	80.17 (0.19)	1.45 (0.01)
	$[T/4]$	71.11 (0.19)	1.73 (0.03)	73.52 (0.23)	1.67 (0.03)	75.65 (0.25)	1.58 (0.02)	76.45 (0.21)	1.51 (0.01)
	T	70.13 (0.45)	1.86 (0.05)	72.22 (0.23)	1.78 (0.05)	72.62 (0.23)	1.76 (0.05)	69.42 (0.23)	1.65 (0.04)
	$[T/4]$	70.15 (0.40)	1.89 (0.05)	71.19 (0.23)	1.60 (0.05)	75.02 (0.21)	1.70 (0.05)	79.01 (0.22)	1.54 (0.04)
	σ_2								
	G	76.14 (0.24)	1.18 (0.03)	81.24 (0.21)	1.16 (0.03)	88.02 (0.19)	1.15 (0.03)	88.14 (0.18)	1.12 (0.02)
	$[T/4]$	75.13 (0.37)	1.20 (0.03)	76.78 (0.34)	1.19 (0.02)	78.12 (0.20)	1.19 (0.02)	89.18 (0.16)	1.24 (0.02)
	T	71.19 (0.26)	1.54 (0.07)	73.13 (0.19)	1.64 (0.05)	74.24 (0.25)	1.48 (0.06)	76.12 (0.20)	1.47 (0.05)
	$[T/4]$	63.68 (0.27)	1.56 (0.10)	64.13 (0.28)	1.84 (0.09)	76.51 (0.21)	1.77 (0.08)	78.18 (0.17)	1.67 (0.07)
ii)	σ_1								
	G	69.17 (0.41)	2.44 (0.11)	74.13 (0.23)	2.18 (0.09)	78.02 (0.19)	1.72 (0.08)	80.02 (0.38)	1.61 (0.07)
	$[T/4]$	72.14 (0.29)	2.56 (0.13)	70.13 (0.23)	2.90 (0.15)	69.02 (0.19)	2.11 (0.19)	76.02 (0.21)	1.77 (0.11)
	T	77.81 (0.28)	2.59 (0.19)	76.43 (0.23)	2.71 (0.19)	75.01 (0.19)	2.55 (0.10)	78.78 (0.20)	1.67 (0.05)
	$[T/4]$	74.28 (0.19)	2.18 (0.16)	73.78 (0.24)	2.32 (0.10)	73.02 (0.19)	2.01 (0.03)	70.62 (0.17)	1.71 (0.02)
	σ_2								
	G	75.13 (0.37)	1.75 (0.11)	73.23 (0.33)	1.57 (0.14)	72.02 (0.19)	1.67 (0.11)	75.12 (0.19)	1.64 (0.10)
	$[T/4]$	76.23 (0.16)	1.59 (0.09)	74.53 (0.38)	1.67 (0.13)	73.18 (0.19)	1.78 (0.11)	85.02 (0.15)	1.63 (0.12)
	T	70.13 (0.17)	1.57 (0.08)	73.68 (0.23)	2.18 (0.19)	71.65 (0.23)	1.99 (0.12)	84.78 (0.19)	1.67 (0.09)
	$[T/4]$	66.19 (0.19)	1.78 (0.09)	78.11 (0.18)	1.65 (0.09)	72.97 (0.16)	1.98 (0.08)	85.11 (0.23)	1.72 (0.09)

5 Application

5.1 Minimum Wage Dataset for China

We consider a Chinese minimum wage dataset. It is collected from 1992 to 2012 for 346 counties over China and the corresponding Chinese regional economic statistics. The data source is the ministry of Human Resources and Social Security, and the China 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 it seems that after 2003, some changes has been made for the minimum wage adjustment policy. The dataset is matched with another one measuring a regional economic situation, namely Chinese Statistical Year Book of China National Knowledge Infrastructure.

The interesting question is to check the geographic heterogeneity of the minimum wages policy. Especially for the economics developed counties around the Pearl River Delta to the Yangtze River Delta, the minimum wage would considerably be 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 discrepancy.

Therefore we apply our estimation procedure as in Section 4. Y_{tj} is taken to be the minimum wage over 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 geographical distribution of minimum wage over 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 on the left panel. And the fitted time varying factors with estimated breakpoints, and confidence intervals. The estimated structural changes for the two factors are closed to the year 2003.

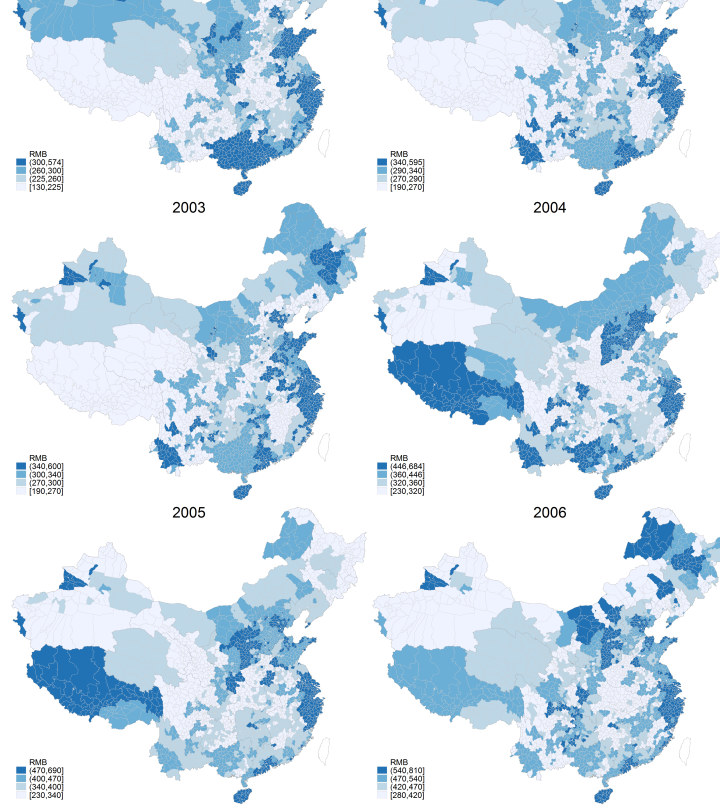


Figure 3: Graphical distribution of the minimum wage in China.

51

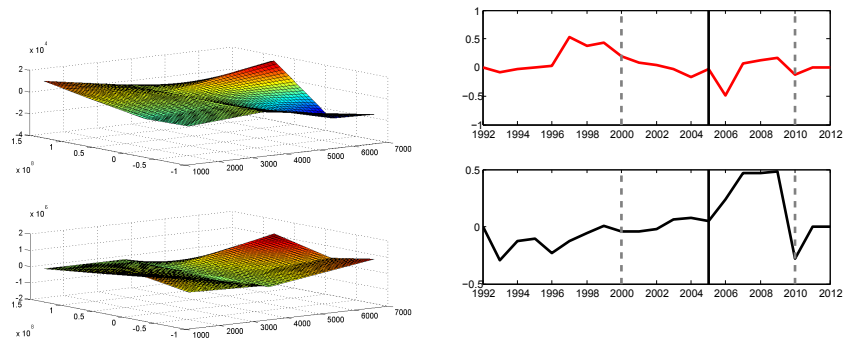


Figure 4: Plot of estimated $m_l(\cdot)$ functions and estimated Z_t , Y : minimum wage on $X1$: first difference of county level gross value added and $X2$: city code. Break point $\hat{\tau} = 2005$, and its confidence interval (dashed grey). Regional cutoff to isolate regions populated with ethnic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

21

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 reflects a snapshot of the stock’s demand and supply curves. The data are collected at the NASDAQ stock market, and are collected at a 60 seconds frequency. 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 day. Normal trading activities take place continuously on all stocks between 9:30 a.m. and 4:00 p.m from Monday to Friday in NASDAQ with totally 390 observations. For illustration, we take one-day trading price as an example for four companies, namely, Amazon, Facebook (on Sep 9th, 2016) AT&T and Tesla on (June 1st, 2016). To show the data structure, the number of shares for the four stocks at time 10:00 am and 10:30 am are plotted in [Figure 5](#). In each minute, five price levels are collected both from the bid and the ask side, with the first one being the lowest sell price and the last one being the highest sell price. As an example, at 10:00am and 216.78 USD(second best ask price), the trading volume is 1798 on June 1st, 2016 for Tesla.

Our Y_{tj} is the trading volume at t th minute and j th 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} is 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 prices 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 [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 break point for \hat{m} functions is always set to 0. $m_1(\cdot)$ represents the average level of trading volume in relationship to the relative to the price level, and $m_2(\cdot)$ corresponds more to the higher order structure of the curves. And we also notice that the estimates with embedded discontinuity in 0 are quite different from the estimation without it. In [Figure 6](#), the estimated two factors are plotted. We also plot the estimated breakpoint and the confidence interval built around it.

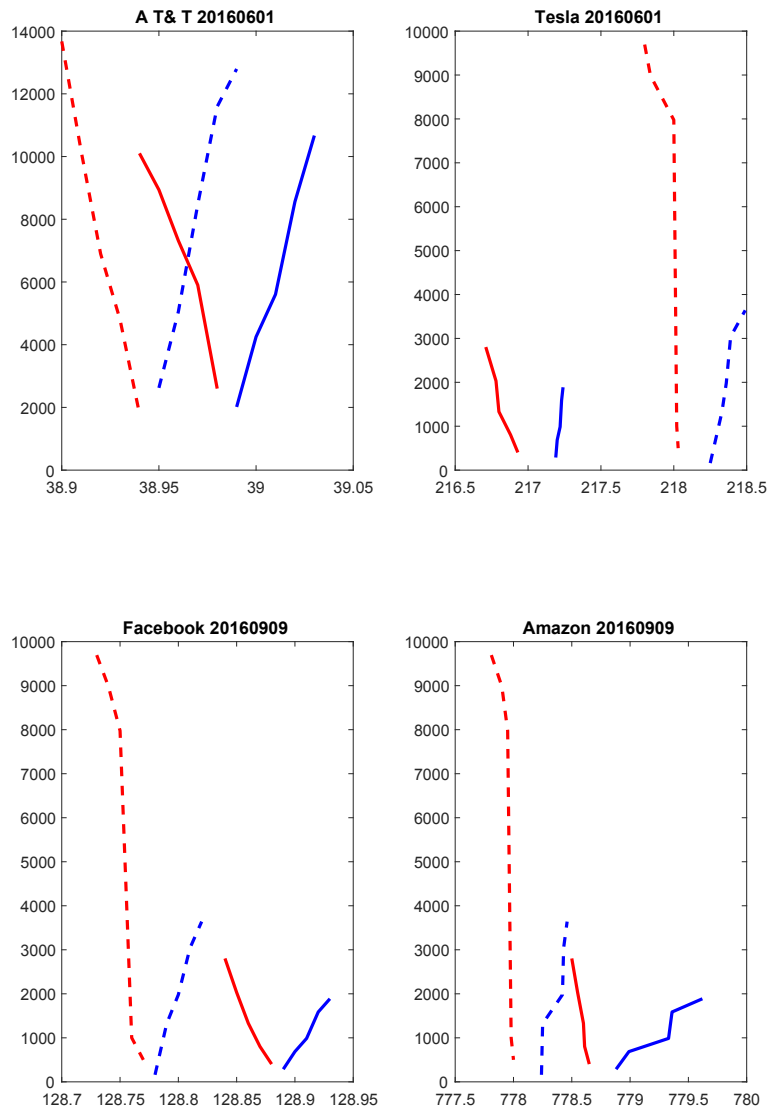
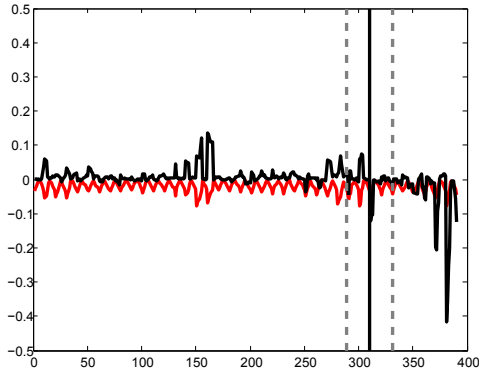
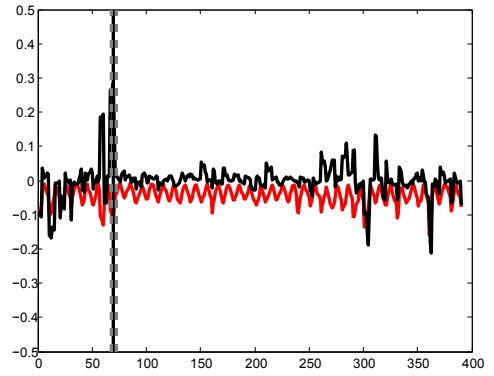


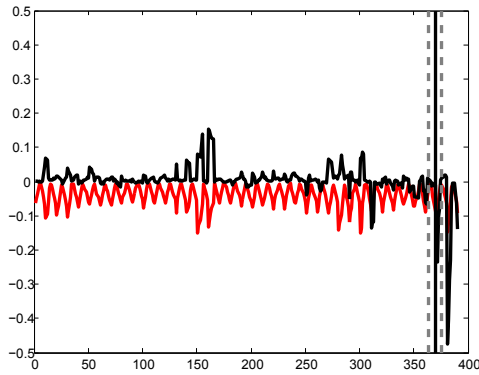
Figure 5: Plot of raw data for different companies at 10:00am (solid line), 10:30 am(dotted line), on the bid side



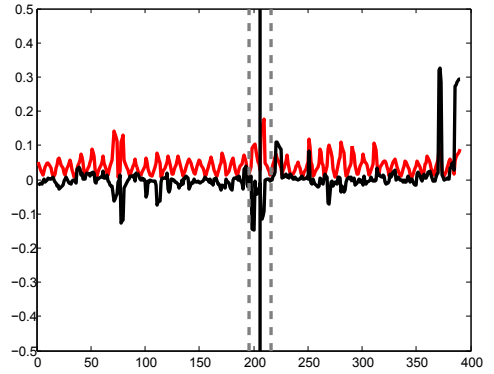
(a) AT& T



(b) Tesla



(c) Facebook



(d) Amazon

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

It is worth noting that the change-point happens at different time point for different stocks, and the width of the confidence interval also varies. This is due to stock specific latent trading dynamics. For AT&T, the change-point is detected at 14 : 40, with a interval of 42 minutes; Tesla switches its latent trading pattern at 10 : 40, with a small 4– minute interval; for facebook, a change happens at late 15 : 35 with a 10– minute length of confidence interval; a change-point is detected for Amazon at 12 : 26 with a 18–minute interval.

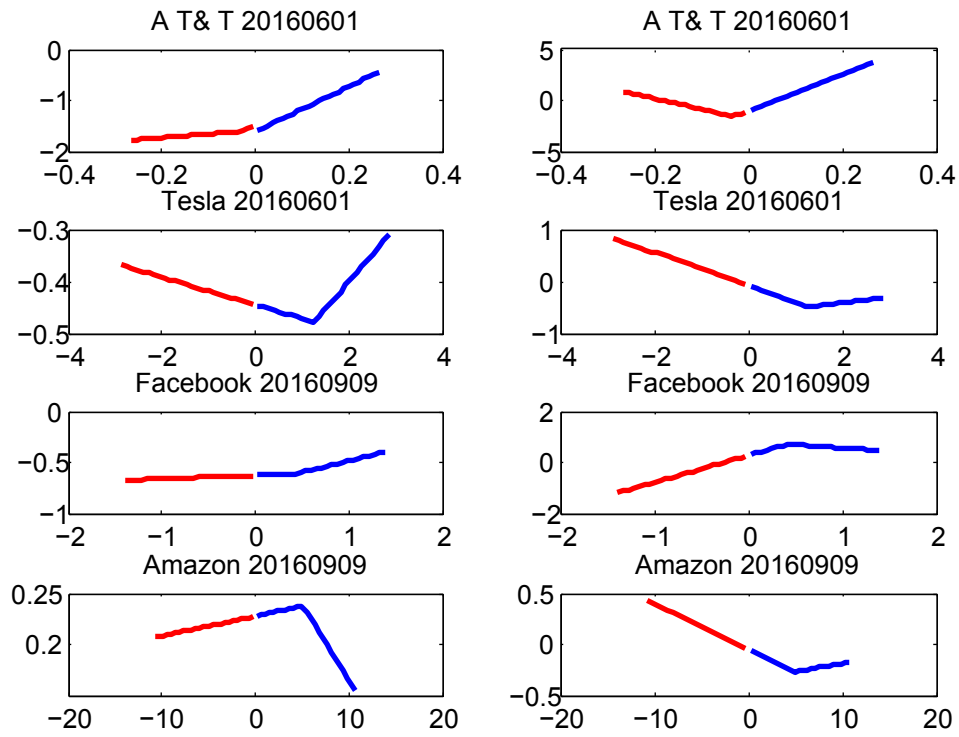


Figure 7: Plot of estimated loading functions $\hat{m}_1(\cdot)$ (left) and $\hat{m}_2(\cdot)$ (right, no break point)

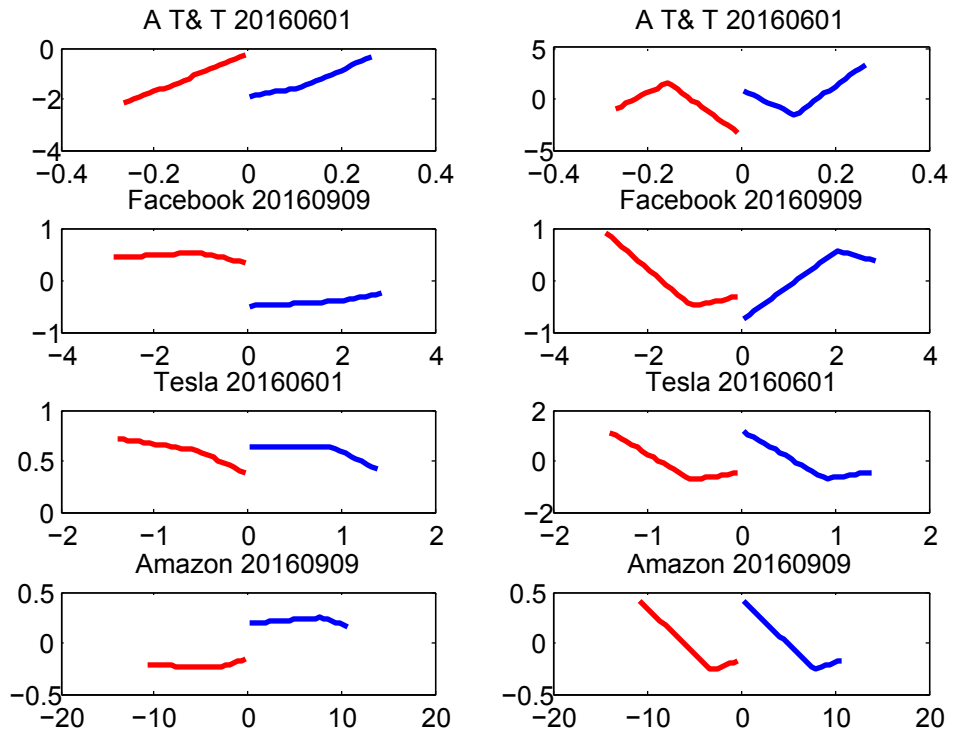


Figure 8: Plot of estimated loading functions $\hat{m}_1(\cdot)$ (left) and (with break point)

6 Conclusion and Further Work

In this paper we propose a dynamic semiparametric factor model with a common structural break. 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, extending the current model to study multiple change points is a very interesting topic. Besides, considering the selection of the number of factors using a $l - 1$ regularization is another interesting direction to pursue.

7 Appendix

Notation. For two matrices $G = (G_{i,t})$, $H = (H_{i,t}) \in \mathbb{R}^{N \times T}$, define

$$\langle G, H \rangle = (TN)^{-1} \sum_{t=1}^T \sum_{i=1}^N G_{i,t} H_{i,t} \quad \text{and} \quad |G|_G^2 = \langle G, G \rangle.$$

Let $\epsilon'_t, t \geq 1$, be an i.i.d. copy of $\epsilon_t, t \geq 1$. Let

$$\mathcal{F}_t = (\epsilon_t, \epsilon_{t-1}, \dots) \quad \text{and} \quad \mathcal{F}_{t, \{t-k\}} = (\epsilon_t, \dots, \epsilon_{t-k+1}, \epsilon'_{t-k}, \epsilon_{t-k-1}, \dots).$$

For any random variable $\xi_t = H(\mathcal{F}_t)$, denote $\xi_{t, \{t-k\}} = H(\mathcal{F}_{t, \{t-k\}})$. Let c_1, c_2, \dots be constants that do not depend on T, N, J , which may change from lemma to lemma. We adopt the functional dependence measure introduced by [Wu \(2005\)](#), i.e. $\|\xi_t - \xi_{t, \{0\}}\|_q$, which measures the effect of ϵ_0 on the observation t .

7.1 Some useful lemmas

LEMMA 1 (Weyl's inequality). *Assume symmetric matrices $\Sigma, \Sigma', \Delta \in \mathbb{R}^{L \times L}$, with $\Sigma = \Sigma' + \Delta$, eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_L$, $\lambda'_1 \geq \lambda'_2 \geq \dots \geq \lambda'_L$ and $d_1 \geq d_2 \geq \dots \geq d_L$ respectively. Then*

$$\lambda'_i + d_L \leq \lambda_i \leq \lambda'_i + d_1.$$

As a result $|\lambda_i - \lambda'_i| \leq |\Delta|_2$.

LEMMA 2 (Davis-Kahan's $\sin \theta$ theorem). Let Σ, Σ' be symmetric matrices in $\mathbb{R}^{L \times L}$, with eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_L$ and $\lambda'_1 \geq \lambda'_2 \geq \dots \geq \lambda'_L$ respectively. Let v_j and v'_j be the corresponding eigenvectors for λ_j and λ'_j respectively. If $v_j^\top v'_j \geq 0$, then

$$|v_j - v'_j|_2 \leq \frac{2^{1/2} |\Sigma - \Sigma'|_F}{\min\{|\lambda_j - \lambda'_{j+1}|, |\lambda_j - \lambda'_{j-1}|\}}.$$

LEMMA 3 (Theorem 1 in [El Machkouri et al. \(2013\)](#)). Denote $Y_i = f(\mathcal{F}_i)$, where f is some measurable function. Let $S_n = \sum_{i=1}^n Y_i$, and $\delta_{i,q} = \|Y_i - Y_{i,\{0\}}\|_q$. If $\mathbb{E}(X_i) = 0$, $\sum_{i \geq 0} \delta_{i,q} < \infty$, some $q \geq 2$, and $\sigma_n^2 := \mathbb{E}(S_n^2) \rightarrow \infty$, then

$$\sigma_n^{-1} S_n \Rightarrow N(0, 1).$$

LEMMA 4. Followings are some useful properties for Z_t and \mathcal{Z}_t . Assume conditions [3.3](#), [3.5](#) and [3.7](#). Then

- (i) For some constant $c_1 > 0$, $\|Z_t - \mathcal{Z}_t\|_2\|_{q'} \leq c_1 \gamma_e^{t/M}$, $t \leq \tau^\diamond$, and $\|Z_t - \mathcal{Z}_t\|_2\|_{q'} \leq c_1 \gamma_e^{(t-\tau^\diamond)/M}$, $t > \tau^\diamond$. Hence $\sum_{t=1}^T \|Z_t - \mathcal{Z}_t\|_2\|_{q'} < \infty$ and $\|Z_t\|_2\|_{q'} \leq c_2$, where $c_2 = c_1 + \|\mathcal{Z}_1\|_2\|_{q'} + \|\mathcal{Z}_{\tau^\diamond+1}\|_2\|_{q'} < \infty$.
- (ii) Let $\delta_t^{(l)} := \|\mathcal{Z}_t^{(l)} - \mathcal{Z}_{t,\{0\}}^{(l)}\|_2\|_{q'}$ and $\delta_t^{(r)} := \|\mathcal{Z}_t^{(r)} - \mathcal{Z}_{t,\{0\}}^{(r)}\|_2\|_{q'}$, then $\delta_t^{(l)} \leq c_3 \gamma_e^{t/M}$ and $\delta_t^{(r)} \leq c_3 \gamma_e^{t/M}$, some constant $c_3 > 0$.

(iii) Recall $W_0 = I_L$. Then

$$|ZZ^\top/T - I_L|_F = O_{\mathbb{P}}(T^{-1/2}). \quad (7.1)$$

Proof. Part (i). First assume $\sum_{i=1}^M |E_i|_2 \leq \gamma_e < 1$. Then for $t \leq \tau^\diamond$

$$\|Z_t - \mathcal{Z}_t\|_2\|_{q'} \leq \gamma_e \max_{t-M \leq s \leq t-1} \|Z_s - \mathcal{Z}_s\|_2\|_{q'}. \quad (7.2)$$

Let $C = \max\{\|Z_s - \mathcal{Z}_s\|_2\|_{q'}, s = 0, -1, \dots, -M+1\}$. We prove the argument by induction, if for any $s < k$, we have $\|Z_s - \mathcal{Z}_s\|_2\|_{q'} \leq C \gamma_e^{s/M}$, then for $s = k$, by (7.2) we have $\|Z_k - \mathcal{Z}_k\|_2\|_{q'} \leq C \gamma_e \gamma_e^{k-M/M} = C \gamma_e^{k/M}$. Hence $\|Z_t - \mathcal{Z}_t\|_2\|_{q'} \leq C \gamma_e^{t/M}$. Similar argument can be applied for $t > \tau^\diamond$. By assumption [3.5](#), there exists a matrix H , such that $\sum_{i=1}^M |HE_i H^{-1}|_2 \leq \gamma_e$ and $HZ_t = \mathcal{B}(HEH^{-1})HZ_t + H\epsilon_t$. Hence by above argument, $\|HZ_t - H\mathcal{Z}_t\|_2\|_{q'} \leq C \gamma_e^{t/M}$. Since $|H^{-1}|_2$ is finite and $|Z_t - \mathcal{Z}_t|_2 \leq |H^{-1}|_2 |HZ_t - H\mathcal{Z}_t|_2$, we complete the proof.

Part (ii). Note $Z_{t,\{0\}}^{(l)} = \mathcal{B}(E)Z_{t,\{0\}}^{(l)} + \epsilon_{t,\{0\}}$. Hence

$$\delta_t^{(l)} \leq \gamma_e \max_{t-M \leq s \leq t-1} \delta_s^{(l)}, \quad t \geq 1, \quad \text{and} \quad \delta_0^{(l)} = \|\epsilon_0 - \epsilon'_0\|_{q'}.$$
 (7.3)

Then same argument as in Part (i), we have $\delta_t^{(l)} \leq c_3 \gamma_e^{t/M}$, where $c_3 = 2\|\epsilon_0\|_{q'} < \infty$. Similarly we have $\delta_t^{(r)} \leq c_3 \gamma_e^{t/M}$.

Part (iii). Define $Z = [Z_1, \dots, Z_T]$. Note $ZZ^\top = \sum_{t=1}^{\tau^\diamond} Z_t Z_t^\top + \sum_{t=\tau^\diamond+1}^T Z_t Z_t^\top =: I_1 + I_2$. Thus by Part (i),

$$|I_1 - \sum_{t=1}^{\tau^\diamond} \mathcal{Z}_t \mathcal{Z}_t^\top|_2 = O_{\mathbb{P}}(1) \quad \text{and} \quad |I_2 - \sum_{t=\tau^\diamond+1}^T \mathcal{Z}_t \mathcal{Z}_t^\top|_2 = O_{\mathbb{P}}(1).$$
 (7.4)

For any $1 \leq i, j \leq L$, we have

$$\|\mathcal{P}_0(\mathcal{Z}_{t,i} \mathcal{Z}_{t,j})\|_{q/2} \leq \|\mathcal{Z}_{t,i}\|_{q'} \|\mathcal{Z}_{t,j} - \mathcal{Z}_{t,j,\{0\}}\|_{q'} + \|\mathcal{Z}_{t,i} - \mathcal{Z}_{t,i,\{0\}}\|_{q'} \|\mathcal{Z}_{t,j}\|_{q'}.$$

Hence Lemma 3 implies $\tau^{\diamond-1/2} \sum_{t=1}^{\tau^\diamond} (\mathcal{Z}_{t,i} \mathcal{Z}_{t,j} - W_{1,i,j}) \Rightarrow N(0, 1)$. Thus $|\sum_{t=1}^{\tau^\diamond} (\mathcal{Z}_t \mathcal{Z}_t^\top - W_1)|_F = O_{\mathbb{P}}(\tau^{\diamond 1/2})$. Similarly we have $|\sum_{t=\tau^\diamond+1}^T (\mathcal{Z}_t \mathcal{Z}_t^\top - W_2)|_F = O_{\mathbb{P}}((T - \tau^\diamond)^{1/2})$. Let \mathcal{Z} be Z with Z_t replaced by \mathcal{Z}_t . Since $\tau^\diamond \asymp T$ and $T - \tau^\diamond \asymp T$, for $W_0 = \tau^\diamond/T W_1 + (T - \tau^\diamond)/T W_2$,

$$|\mathcal{Z} \mathcal{Z}^\top / T - W_0|_F = O_{\mathbb{P}}(T^{-1/2}).$$

Together with (7.4), we complete the proof. \square

LEMMA 5. For some constants $\alpha_J, \alpha_T > 0$, assume $J = O(N^{\alpha_J})$, $T = O(N^{\alpha_T})$ and Assumption 3.2, then with probability $1 - \exp(-cN^{1-2\alpha_J})$, some $c > 0$, we have

$$l_\phi/2 \leq \frac{1}{N} \min_{t=1, \dots, T} \lambda_{\min}(\Phi(X_t)^\top \Phi(X_t)) \leq \frac{1}{N} \max_{t=1, \dots, T} \lambda_{\max}(\Phi(X_t)^\top \Phi(X_t)) \leq 2u_\phi.$$
 (7.5)

Proof. We shall only show the first inequality, since the third one can be similarly derived. By Weyl's inequality (Lemma 1), we have

$$\begin{aligned} & \frac{1}{N} \min_{t=1, \dots, T} \lambda_{\min}(\Phi(X_t)^\top \Phi(X_t)) \\ &= \min_{t=1, \dots, T} \lambda_{\min}\left(\frac{1}{N} \sum_{i=1}^N \phi(X_{t,i}) \phi(X_{t,i})^\top\right) \\ &\geq \min_{t=1, \dots, T} \lambda_{\min}(\mathbb{E} \phi(X_{t,1}) \phi(X_{t,1})^\top) - \max_{t=1, \dots, T} \left| \frac{1}{N} \sum_{i=1}^N \phi(X_{t,i}) \phi(X_{t,i})^\top - \mathbb{E} \phi(X_{t,1}) \phi(X_{t,1})^\top \right|_F. \end{aligned}$$
 (7.6)

Since $|\phi_j|_\infty$ is bounded by c_ϕ , by Hoeffding's inequality

$$\begin{aligned}
& \mathbb{P}\left(\max_{t=1,\dots,T} \left| \frac{1}{N} \sum_{i=1}^N [\phi(X_{t,i})\phi(X_{t,i})^\top - \mathbb{E}\phi(X_{t,1})\phi(X_{t,1})^\top] \right|_F \geq x\right) \\
& \leq \sum_{t=1}^T \sum_{j_1, j_2=1}^J \mathbb{P}\left(\left| \sum_{i=1}^N [\phi_{j_1}(X_{t,i})\phi_{j_2}(X_{t,i}) - \mathbb{E}\phi_{j_1}(X_{t,i})\phi_{j_2}(X_{t,i})] \right| \geq Nx/J\right) \\
& \leq 2TJ^2 \exp\left(-\frac{Nx^2}{2J^2c_\phi^4}\right). \tag{7.7}
\end{aligned}$$

Assumption 3.2 assumes $\min_{t=1,2,\dots,T} \mathbb{E}\phi(X_{t,1})\phi(X_{t,1})^\top \geq L_\phi$. Hence by (7.6) and (7.7), the first inequality of (7.5) holds with probability greater than $\exp\{-L_\phi^2(8c_\phi^4)^{-1}N^{1-2\alpha_J} + \log(2TJ^2)\}$, \square

7.2 Proof of Theorem 1

Following notation will be used throughout this subsection. For $1 \leq t \leq T$, let

$$g(t, x) = \phi(x)^\top A(1, z_t^\top)^\top,$$

and $g^*(t, x)$ (resp. $\hat{g}(t, x)$) be $g(t, x)$ with A, z_t replaced by A^*, Z_t (resp. \hat{A}, \hat{Z}_t). Let

$$g_0(t, x) = m(x)^\top (1, Z_t^\top)^\top,$$

and thus $Y_{t,i} = g_0(t, X_{t,i}) + \varepsilon_{t,i}$. Denote $N \times T$ matrices $M_g = (g(t, X_{t,i}))_{1 \leq i \leq N, 1 \leq t \leq T}$, and let \hat{g} (resp. g^* , g_0) be the same as g with $g(t, X_{t,i})$ replaced by $\hat{g}(t, X_{t,i})$ ((resp. $g^*(t, X_{t,i})$, $g_0(t, X_{t,i})$). For any $R > 0$, denote the function class,

$$\mathcal{G}(R) = \{g(t, x) = \phi(x)^\top A(1, z_t^\top)^\top : A \in \mathbb{R}^{J \times L}, z_i \in \mathbb{R}^d, 1 \leq i \leq T, |M_g - M_{g^*}|_{\mathcal{G}} \leq R\}.$$

Let the δ -entropy of function class $\mathcal{G}(R)$ with respect to the norm $|\cdot|_{\mathcal{G}}$ be $\mathcal{H}(\delta, \mathcal{G}(R))$.

Proof of Part (i). Note that

$$\begin{aligned}
|M_{\hat{g}} - M_{g^*}|_{\mathcal{G}}^2 &= \frac{1}{NT} \sum_{t=1}^T \left| \Phi(X_t) [\hat{A}(1, \hat{Z}_t^\top)^\top - A^*(1, Z_t^\top)^\top] \right|_2^2 \\
&\geq \frac{1}{N} \min_{t=1,\dots,T} \lambda_{\min}(\Phi(X_t)^\top \Phi(X_t)) \frac{1}{T} \sum_{t=1}^T |\hat{A}(1, \hat{Z}_t^\top)^\top - A^*(1, Z_t^\top)^\top|_2^2.
\end{aligned}$$

We shall show in Lemma 5 and Lemma 4 that there exists Ω with $\mathbb{P}(\Omega) \rightarrow 1$, on which

$$|T^{-1}ZZ^\top - W_0|_F = |T^{-1}ZZ^\top - I_L|_F \leq cT^{-1/2}, \quad (7.8)$$

where $c > 0$ some constant, and (7.5) hold.

From now on we shall only work on Ω . Then by (7.5) it suffices to show

$$|M_{\hat{g}} - M_{g^*}|_{\mathcal{G}}^2 = O_{\mathbb{P}}(\rho^2 + \delta_J^2). \quad (7.9)$$

Let $\varepsilon = (\varepsilon_{i,t})_{1 \leq i \leq N, 1 \leq t \leq T}$. Recall h in (2.4). Note $(\hat{A}, \hat{Z}_t, 1 \leq t \leq T)$ is the minimum point for function h , thus $h(\hat{A}, \hat{Z}_t, 1 \leq t \leq T) \leq h(A^*, Z_t, 1 \leq t \leq T)$, which can be rewritten as $|M_{\hat{g}} - M_{g_0}|_{\mathcal{G}}^2 \leq 2\langle \varepsilon, M_{\hat{g}} - M_{g^*} \rangle + |M_{g^*} - M_{g_0}|_{\mathcal{G}}^2$. Therefore

$$|M_{\hat{g}} - M_{g^*}|_{\mathcal{G}}^2 \leq 2|M_{\hat{g}} - M_{g_0}|_{\mathcal{G}}^2 + 2|M_{g^*} - M_{g_0}|_{\mathcal{G}}^2 \leq 4\langle \varepsilon, M_{\hat{g}} - M_{g^*} \rangle + 4|M_{g^*} - M_{g_0}|_{\mathcal{G}}^2. \quad (7.10)$$

Since $\sup_x |\phi(x)^\top A^* - m(x)^\top|_\infty \leq \delta_J$,

$$|M_{g^*} - M_{g_0}|_{\mathcal{G}}^2 \leq \delta_J^2 \sum_{t=1}^T |(1, Z_t^\top)^\top|_2^2 / T = \delta_J^2 (1 + |Z|_F^2 / T) = O_{\mathbb{P}}(\delta_J^2),$$

where the last equality is due to (7.8). Hence, it suffices to consider the event $\langle \varepsilon, M_{\hat{g}} - M_{g^*} \rangle \geq |M_{g^*} - M_{g_0}|_{\mathcal{G}}^2$, since otherwise $|M_{\hat{g}} - M_{g^*}|_{\mathcal{G}} = O_{\mathbb{P}}(\delta_J^2)$. Therefore by (7.10),

$$|M_{\hat{g}} - M_{g^*}|_{\mathcal{G}}^2 \leq 8\langle \varepsilon, M_{\hat{g}} - M_{g^*} \rangle. \quad (7.11)$$

We shall then use truncation technique to further deal with $|M_{\hat{g}} - M_{g^*}|_{\mathcal{G}}$, specifically: recall β_B in Assumption 3.1, for some $m > 0$, $1/q < \beta < \beta_B - 1$, let

$$f(t) = |t|^\beta, \text{ if } t < 0; \quad f(t) = 1, \text{ if } t \geq 0. \quad (7.12)$$

Denote the truncated error $\tilde{\eta}_t = (\tilde{\eta}_{t,1}, \tilde{\eta}_{t,2}, \dots, \tilde{\eta}_{t,N})$ with

$$\tilde{\eta}_{t,i} = (\eta_{t,i} \wedge mf(t)) \vee (-mf(t)). \quad (7.13)$$

Then for constant $C_0 = 2^{1/2}c_B\mu_2$,

$$\begin{aligned} \mathbb{P}(|M_{\hat{g}} - M_{g^*}|_{\mathcal{G}} > x) &\leq \mathbb{P}(|M_{\hat{g}} - M_{g^*}|_{\mathcal{G}} > x, |\varepsilon|_{\mathcal{G}} \leq C_0, \max_{t \leq T, 1 \leq i \leq N} |\eta_{t,i}| \leq mf(t)) \\ &\quad + \mathbb{P}(|\varepsilon|_{\mathcal{G}} > C_0) + \sum_{t \leq T} \sum_{i=1}^N \mathbb{P}(|\eta_{t,i}| > mf(t)) =: I_1 + I_2 + I_3. \end{aligned}$$

By Lemma 6, $I_1 = o_{\mathbb{P}}(1)$ when $x \geq \rho_m$ where

$$\rho_m = 2(C_1^{1/2} \vee 1)(TN)^{-1/2}(T+J)^{1/2}m \log^{1/2}(TN). \quad (7.14)$$

By Lemma 7, $I_2 = O((TN)^{-1})$. For I_3 , by Markov's inequality,

$$I_3 \leq \sum_{t=0}^T \sum_{i=1}^N \mathbb{P}(|\eta_{t,i}| \geq m) + \sum_{t < 0} \sum_{i=1}^N \mathbb{P}(|\eta_{t,i}| \geq m | t|^\beta) \leq (1 + q\beta(q\beta - 1)^{-1}T^{-1})TNm^{-q}\mu_q^q.$$

Our results follow by choosing $m = c(TN)^{1/q}$, some constant c large enough. \square

LEMMA 6. Recall definitions of ρ_m in (7.14). Under assumptions in Theorem 1, we have $I_1 = o_{\mathbb{P}}(1)$ when $x \geq \rho_m$.

Proof. We shall work on the event where (7.5) and (7.8) hold. When $|\varepsilon|_{\mathcal{G}} \leq C_0$, by Cauchy's inequality and (7.11), $|M_{\hat{g}} - M_{g^*}|_{\mathcal{G}} \leq 8C_0$. Let $S = \min\{s : 2^s x \geq 8C_0\}$ and $\tilde{\varepsilon}_t$ be ε_t in (2.2) with η_t therein replaced by $\tilde{\eta}_t$ in (7.13). Then

$$\begin{aligned} I_1 &\leq \sum_{s=0}^{S-1} \mathbb{P}(2^s x < |M_{\hat{g}} - M_{g^*}|_{\mathcal{G}} \leq 2^{s+1}x, |\tilde{\varepsilon}|_{\mathcal{G}} \leq C_0, \max_{t \leq T, 1 \leq i \leq N} |\eta_{t,i}| \leq mf(t)) \\ &\leq \sum_{s=0}^{S-1} \mathbb{P}\left(\sup_{g \in \mathcal{G}(2^{s+1}x)} \langle \tilde{\varepsilon}, M_g - M_{g^*} \rangle \geq 2^{2s-3}x^2, |\tilde{\varepsilon}|_{\mathcal{G}} \leq C_0 \right), \end{aligned} \quad (7.15)$$

where the last inequality is due to (7.11). In the following we shall show:

Step 1 : If there exists some constant $c_0 > 0$ independent of T, N, J, w , such that

$$\sqrt{TN}w \geq c_0 m \max \left\{ \int_{w/(8C_0)}^R H^{1/2}(u, \mathcal{G}(R))du, R \right\},$$

then we have

$$\mathbb{P}\left(\sup_{g \in \mathcal{G}(R)} |\langle \tilde{\varepsilon}, M_g - M_{g^*} \rangle| \geq w, |\tilde{\varepsilon}|_{\mathcal{G}} \leq C_0\right) \leq c'_0 \exp\left(-\frac{TNw^2}{4c'_0{}^2 m^2 R^2}\right), \quad (7.16)$$

where c'_0 is independent of T, N, J, m, w .

Step 2 : For some constant $C_1 > 0$ independent of y, N, T, J , we have

$$H(\delta, \mathcal{G}(y)) \leq C_1 \{(T + J)\log(y/\delta)\}. \quad (7.17)$$

Step 3 : Applying (7.16) and (7.17) to show that $I_1 = o_{\mathbb{P}}(1)$ when $x \geq \rho_m$.

Step 1: Since $(\tilde{\eta}_{t,i})$ are independent and $\sum_{t=1}^T \gamma_t^\top \tilde{\varepsilon}_t = \sum_{k \leq T} \sum_{t=1 \vee k}^T \gamma_t^\top B_{t-k} \tilde{\eta}_k$, by Hoeffding's inequality,

$$\mathbb{P}\left(\left|\sum_{t=1}^T \gamma_t^\top \tilde{\varepsilon}_t\right| \geq w\right) \leq 2 \exp\left\{-\frac{w^2}{2m^2 \sum_{k \leq T} \left|\sum_{t=1 \vee k}^T B_{t-k}^\top \gamma_t\right|_2^2 f(k)^2}\right\}. \quad (7.18)$$

Notice

$$\sum_{k \leq T} \left|\sum_{t=1 \vee k}^T B_{t-k}^\top \gamma_t\right|_2^2 f(k)^2 = \sum_{t_1, t_2=1}^T \gamma_{t_1}^\top \left(\sum_{k \leq t_1 \wedge t_2} B_{t_1-k} B_{t_2-k}^\top f(k)^2\right) \gamma_{t_2} \leq \sum_{t_1, t_2=1}^T \sigma_{t_1, t_2} |\gamma_{t_1}|_2 |\gamma_{t_2}|_2,$$

where $\sigma_{t_1, t_2} = \sum_{k \leq t_1 \wedge t_2} |B_{t_1-k}|_2 |B_{t_2-k}|_2 f(k)^2$. For matrix $\varpi = (\sigma_{t_1, t_2})_{t_1, t_2=1}^T$, ϖ is symmetric with

$$|\varpi|_1 = |\varpi|_\infty = \max_{1 \leq t_2 \leq T} \sum_{t_1=1}^T \sum_{k \leq t_1 \wedge t_2} |B_{t_1-k}|_2 |B_{t_2-k}|_2 f(k)^2 \leq c_0 c_B^2,$$

where constant c_0 only depends on β, β_B . Thus $|\varpi|_2 \leq (|\varpi|_1 |\varpi|_\infty)^{1/2} \leq c_0 c_B^2$. Hence for any vector $\gamma_t \in \mathbb{R}^N$, by (7.18)

$$\mathbb{P}\left(\left|\sum_{t=1}^T \gamma_t^\top \tilde{\varepsilon}_t\right| \geq w\right) \leq 2 \exp\left\{-\frac{c_\beta w^2}{m^2 c_B^2 \sum_{t=1}^T |\gamma_t|_2^2}\right\}, \quad (7.19)$$

where c_β only depends on β and β_B . Recall $X_{t,j}$ and Z_t are independent of ε_t , hence entities in M_{g^*} is independent of ε_t . Therefore result follows by (7.19) and Lemma 3.2 in [van de Geer \(2000\)](#) with W, δ and ϵ therein equal $\tilde{\varepsilon}/m, w/m$ and $w/(2m)$.

Step 2: By (7.5), $\mathcal{G}(y)$ is included in $T^{-1} \sum_{t=1}^T |A(1, z_t^\top)^\top - A^*(1, Z_t^\top)^\top|_2^2 \leq 2y^2/l_\phi$, and is further covered by,

$$\left\{ (A, Z) : |A_1 - A_1^*|_2^2 \leq 2y^2/l_\phi, T^{-1}|A_2 z - A_2^* Z|_F^2 \leq 2y^2/l_\phi \right\}. \quad (7.20)$$

By (7.8), for all large T , recall that u_a is defined in (ii) of Assumption 3.6,

$$l_a/2 \leq T^{-1}|A_2^* Z|_2^2 \leq 2u_a. \quad (7.21)$$

Denote $c_1 = \min\{2l_a'^2/(24^2 u_a), \sqrt{2u_a}\}$.

If $2y^2/l_\phi \geq c_1$, then by (7.20) and (7.21),

$$T^{-1/2}|A_2 z|_F \leq T^{-1/2}|A_2^* Z|_F + (2y^2/l_\phi)^{1/2} \leq c_2 y,$$

where $c_2 = (2/l_\phi)^{1/2}[\sqrt{2Lu_a^2/c_1} + 1]$. Denote the condensed SVD $T^{-1/2}A_2 z = U\Lambda V^\top$. Then above indicates $|\Lambda|_F \leq c_2 y$. Hence $\mathcal{G}(y)$ belongs to

$$\{g(t, x) = \phi(x)^\top A(1, z_t^\top)^\top : |A_1 - A_1^*|_2^2 \leq 2y^2/l_\phi, |\Lambda|_F \leq c_2 y\} \quad (7.22)$$

By (7.5), for $M_{g'}$ equals M_g with A_1, A_2, z therein replaced by A'_1, A'_2, z' , and the condensed SVD $A'_2 z' = U'\Lambda'V'^\top$,

$$\begin{aligned} |M_g - M_{g'}|_{\mathcal{G}} &\leq 2u_\phi(|A_1 - A'_1|_2^2 + |U\Lambda V^\top - U'\Lambda'V'^\top|_F^2) \\ &\leq 2u_\phi[|A_1 - A'_1|_2^2 + 2(|\Lambda(V - V')^\top|_F^2 + |\Lambda - \Lambda'|_F^2 + |(U - U')\Lambda'|_F^2)]. \end{aligned} \quad (7.23)$$

Therefore by (7.22) and (7.23), the δ entropy for $\mathcal{G}(y)$ is of order $O((J + T)\log(y/\delta))$.

If $2y^2/l_\phi < c_1$, then by (7.20) and (7.21)

$$T^{-1}|A_2 z z^\top A_2^\top - A_2^* Z Z^\top A_2^{*\top}|_F \leq T^{-1}|A_2 z - A_2^* Z|_F(|A_2 z|_2 + |A_2^* Z|_2) \leq 6(u_a/l_\phi)^{1/2} y, \quad (7.24)$$

which is less than $l'_a/4$ due to $2y^2/l_\phi < c_1$. By (7.8), $|T^{-1}A_2^* Z Z^\top A_2^{*\top} - A_2^* W_0 A_2^{*\top}|_F \leq O(T^{-1/2})$. Hence for T large, by (iii) of Assumption 3.6

$$\text{gap}_L(T^{-1}A_2^* Z Z^\top A_2^{*\top}) \geq l'_a/2 > 0. \quad (7.25)$$

By Weyl's inequality, (7.24) and (7.25),

$$\max_{1 \leq i \leq L} |\lambda_i(T^{-1}A_2 z z^\top A_2^\top) - \lambda_i(T^{-1}A_2^* Z Z^\top A_2^{*\top})| \leq l'_a/4.$$

Thus by Davis-Kahan $\sin \theta$ theorem, there exists a condensed SVD decomposition $T^{-1/2}A_2 z = U\Lambda V^\top$ and $T^{-1/2}A_2^* Z = U^*\Lambda^*V^{*\top}$, such that $|U - U^*|_F = O(y)$. Similarly we have $|V - V^*|_F = O(y)$ and by Weyl's inequality $|\Lambda - \Lambda^*|_F = O(y)$. Thus the δ entropy for $\mathcal{G}(y)$ is of order $O((J+T)\log(y/\delta))$.

Step 3: Let $\Psi(y) = my \max\{C_1^{1/2}(T+J)^{1/2}\log^{1/2}(2^8 C_0/y), 1\}$. Then

$$\Psi(y) \geq m \max \left\{ \int_{y^2/(2^8 C_0)}^y H^{1/2}(u, \mathcal{G}(y)) du, y \right\},$$

and $\Psi(y)/y^2$ is a non-increasing on $(0, 2^8 C_0)$. For ρ_m in (7.14), $\sqrt{TN}\rho_m^2 \geq \Psi(\rho_m)$, when $TN \geq (2^8 C_0)^{2/7}$. Hence by (7.16), for any $y \geq \rho_m$ and $TN \geq (2^8 C_0)^{2/7}$,

$$\mathbb{P}\left(\sup_{g \in \mathcal{G}(y)} |\langle \tilde{\varepsilon}, M_g - M_{g^*} \rangle| \geq 2^{-5}y^2, |\tilde{\varepsilon}|_{\mathcal{G}} \leq C_0\right) \leq c_3 \exp\left(-\frac{TNy^2}{4c_3m^2}\right),$$

some constant $c_3 = C_1^{1/2}c'_0 > 0$. Inserting above into (7.15) with $x = \rho_m$ implies

$$I_1 \lesssim \exp\{-TN\rho_m^2/(4c_3m^2)\} = o(1).$$

□

LEMMA 7. *Under assumptions of Theorem 1, we have*

$$\mathbb{P}\left(|\varepsilon|_{\mathcal{G}}^2 \geq 2c_B^2\mu_2^2\right) \leq c_1\mu_2^{-4}(TN)^{-1},$$

where $c_1 = \max\{\mu_2^4, \text{Var}(\eta_{1,1}^2)\}$.

Proof. Notice

$$|\varepsilon|_{\mathcal{G}}^2 = \frac{1}{NT} \sum_{k_1, k_2 \leq T} \eta_{k_1}^\top \left(\sum_{t=1 \vee k_1 \vee k_2}^T B_{t-k_1}^\top B_{t-k_2} \right) \eta_{k_2}.$$

Since $\eta_{t,j}, t, j \in \mathbb{Z}$ are i.i.d. with zero mean and $|B|_F^2 \leq N|B|_2^2$ for any matrix $B \in \mathbb{R}^{N \times N}$,

$$\mathbb{E}|\varepsilon|_{\mathcal{G}}^2 = (TN)^{-1}\mathbb{E}(\eta_{1,1}^2) \sum_{k \leq T} \sum_{t=1 \vee k}^T |B_{t-k}|_F^2 \leq c_B^2 \mu_2^2,$$

and

$$\text{Var}(|\varepsilon|_{\mathcal{G}}^2) \leq c_1(TN)^{-2} \sum_{k_1 \leq T} \sum_{k_2 \leq T} \left| \sum_{t=1 \vee k_1 \vee k_2}^T B_{t-k_1}^\top B_{t-k_2} \right|_F^2 \leq c_1 c_B^4 / (TN).$$

The desired result follows from Chebyshev's inequality. \square

Proof of Part (ii). From Part (i) and (7.8) we have

$$T^{-1}|A^*ZZ^\top A^{*\top} - \hat{A}\hat{Z}\hat{Z}^\top \hat{A}^\top|_F \leq T^{-1}(|A^*Z|_2 + |\hat{A}\hat{Z}|_2)|A^*Z - \hat{A}\hat{Z}|_F = O_{\mathbb{P}}(\rho + \delta_J). \quad (7.26)$$

By Weyl's inequality with probability tending to 1, $\min_{1 \leq i \neq j \leq L+1} |\lambda_i(T^{-1}A^*ZZ^\top A^{*\top}) - \lambda_j(T^{-1}\hat{A}\hat{Z}\hat{Z}^\top \hat{A}^\top)| \geq l'_a/4$. Recall $T^{-1/2}\hat{A}_2\hat{Z} = \hat{U}\hat{\Lambda}\hat{V}^\top$. Denote the condensed SVD of $T^{-1/2}A_2^*Z = U^*\Lambda^*V^{*\top}$. Thus Davis-Kahan $\sin \theta$ theorem implies if $U_i^{*\top}\hat{U}_i > 0$, then

$$|U_i^* - \hat{U}_i|_2 \leq \sqrt{2}T^{-1}|A^*ZZ^\top A^{*\top} - \hat{A}\hat{Z}\hat{Z}^\top \hat{A}^\top|_F / (l'_a/4) = O_{\mathbb{P}}(\rho + \delta_J),$$

where U_i^* and \hat{U}_i represent the i th columns of U^* and \hat{U} . That is by choosing the sign for U_i^* , we have $|\hat{U} - U^*|_F^2 = O_{\mathbb{P}}(\rho^2 + \delta_J^2)$. By Weyl's inequality and (7.26), we have $|\hat{\Lambda} - \Lambda^*|_F^2 = O_{\mathbb{P}}(\rho^2 + \delta_J^2)$. Thus $|\hat{A}_2 - U^*\Lambda^*|_F = O_{\mathbb{P}}(\rho + \delta_J)$ in view of $\hat{A}_2 = \hat{U}\hat{\Lambda}$. Hence

$$\begin{aligned} |T^{-1/2}\hat{Z} - V^{*\top}|_F &= |(\hat{A}_2^\top \hat{A}_2)^{-1} \hat{A}_2 (T^{-1/2}\hat{A}_2\hat{Z} - \hat{A}_2V^{*\top})|_F \\ &\leq |(\hat{A}_2^\top \hat{A}_2)^{-1} \hat{A}_2|_2 (|T^{-1/2}\hat{A}_2\hat{Z} - U^*\Lambda^*V^{*\top}|_F + |U^*\Lambda^* - \hat{A}_2|_F) = O_{\mathbb{P}}(\rho + \delta_J). \end{aligned}$$

Since $T^{-1/2}A_2^*Z = U^*\Lambda^*V^{*\top}$, there exists an invertible matrix D_T , such that $T^{-1/2}D_TZ = V^{*\top}$. Since $|T^{-1}ZZ^\top - I_L| = O_{\mathbb{P}}(T^{-1/2})$, we have $\mathbb{P}(\lambda_{\min}(D_T) < 1/2) = O_{\mathbb{P}}(T^{-1/2})$ and thus both $|D_T^\top D_T - I_L|_F, |D_T D_T^\top - I_L|_F$ equal $O_{\mathbb{P}}(T^{-1/2})$. Since $T^{-1}A_2^*ZZ^\top A_2^{*\top} = U^*\Lambda^*{}^2U^{*\top}$, $|A_2^*A_2^{*\top} - U^*\Lambda^*{}^2U^{*\top}|_F = O_{\mathbb{P}}(T^{-1/2})$. Recall $A_2^{*\top}A_2^*$ is diagonal with non-increasing diagonal entities. Hence by Weyl's inequality, we have $|A_2^{*\top}A_2^* - \Lambda^*{}^2|_F = O_{\mathbb{P}}(T^{-1/2})$. Note we also have $A_2^*D_T^{-1} = U^*\Lambda^*$, then $D_TA_2^{*\top}A_2^* = D_TD_T^\top\Lambda^*{}^2D_T$, and therefore $|D_TA_2^{*\top}A_2^* - A_2^{*\top}A_2^*D_T|_F = O_{\mathbb{P}}(T^{-1/2})$. By assumption $A_2^{*\top}A_2^*$ has distinct diagonal values. Therefore $|D_T - D^*|_F = O_{\mathbb{P}}(T^{-1/2})$ where D^* is a diagonal matrix with entities either -1 or 1 . \square

7.3 Proof of Theorem 2

We shall only show the case $\tau < \tau^\diamond$ since the other direction can be similarly dealt with. First let us introduce some notation. For any $k_1 < k_2$, let

$$f(Z, k_1, k_2) = \begin{bmatrix} Z_{k_1-1} & Z_{k_1} & \cdots & Z_{k_2-1} \\ Z_{k_1-2} & Z_{k_1-1} & \cdots & Z_{k_2-2} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{k_1-M} & Z_{k_1-M+1} & \cdots & Z_{k_2-M} \end{bmatrix} \in \mathbb{R}^{LM \times (k_2 - k_1 + 1)}$$

and $h(Z, k_1, k_2) = \begin{bmatrix} Z_{k_1} & Z_{k_1+1} & \cdots & Z_{k_2} \end{bmatrix} \in \mathbb{R}^{L \times (k_2 - k_1 + 1)}.$

We will need to handle three segments to account for the deviation of our estimated break point and the true break point τ^\diamond , namely $1 : \tau$, $\tau + 1 : \tau^\diamond$, $\tau^\diamond + 1 : T$. Denote $\Gamma_1 = f(Z, 1, \tau)$, $\Gamma_2 = f(Z, \tau + 1, \tau^\diamond)$ and $\Gamma_3 = f(Z, \tau^\diamond + 1, T)$. Let $\Gamma_{i,t}$ be the t th column of Γ_i , $i = 1, 2, 3$, and $\Gamma_{i,t,l}$ be the l th coordinate of $\Gamma_{i,t}$. Denote $\zeta_1 = h(Z, 1, \tau)$, $\zeta_2 = h(Z, \tau + 1, \tau^\diamond)$ and $\zeta_3 = h(Z, \tau^\diamond + 1, T)$. Also let $U_{\epsilon_1} = h(\epsilon, 1, \tau)$, $U_{\epsilon_2} = h(\epsilon, \tau + 1, \tau^\diamond)$ and $U_{\epsilon_3} = h(\epsilon, \tau^\diamond + 1, T)$. It can be seen that $\zeta_1 = E\Gamma_1 + U_{\epsilon_1}$, $\zeta_2 = E\Gamma_2 + U_{\epsilon_2}$, and $\zeta_3 = \tilde{E}\Gamma_3 + U_{\epsilon_3}$.

Let

$$S(\tau, H, F) = \sum_{t=1}^{\tau} |D_T Z_t - \mathcal{B}(H)D_T Z_t|_2^2 + \sum_{t=\tau+1}^T |D_T Z_t - \mathcal{B}(F)D_T Z_t|_2^2.$$

Denote $V_\tau = \inf_{H, F} S(\tau, H, F)$ and $\tau^* = \operatorname{argmin}_{1 \leq \tau \leq T} V_\tau$. Let $(H_\tau, F_\tau) = \operatorname{argmin}_{H, F} S(\tau, H, F)$ and $(H^*, F^*) = \operatorname{argmin}_{H, F} S(\tau^*, H, F)$. Then τ^* is the ideal estimation of the break point with known factors Z_t , and H_τ and F_τ are the parameters associated with τ .

Proof. Recall that the block matrix $\tilde{D}_T = I_M \otimes D_T$. Then for any fixed τ , we have the following solutions

$$H_\tau = D_T \zeta_1 \Gamma_1^\top (\Gamma_1 \Gamma_1^\top)^{-1} \tilde{D}_T^{-1} \quad \text{and} \quad F_\tau = D_T (\zeta_2 \Gamma_2^\top + \zeta_3 \Gamma_3^\top) (\Gamma_2 \Gamma_2^\top + \Gamma_3 \Gamma_3^\top)^{-1} \tilde{D}_T^{-1}. \quad (7.27)$$

Note

$$\begin{aligned}
V_\tau &= \sum_{t=1}^{\tau} |(\mathcal{B}(H^\diamond) - \mathcal{B}(H_\tau))D_T Z_t + D_T \epsilon_t|_2^2 + \sum_{t=\tau+1}^{\tau^\diamond} |(\mathcal{B}(H^\diamond) - \mathcal{B}(F_\tau))D_T Z_t + D_T \epsilon_t|_2^2 \\
&+ \sum_{t=\tau^\diamond+1}^T |(\mathcal{B}(F^\diamond) - \mathcal{B}(F_\tau))D_T Z_t + D_T \epsilon_t|_2^2. \tag{7.28}
\end{aligned}$$

Then for $d_t = (\mathcal{B}(H^\diamond) - \mathcal{B}(H_\tau))D_T Z_t$, $t \leq \tau$; $d_t = (\mathcal{B}(H^\diamond) - \mathcal{B}(F_\tau))D_T Z_t$, $\tau + 1 \leq t \leq \tau^\diamond$; $d_t = (\mathcal{B}(F^\diamond) - \mathcal{B}(F_\tau))D_T Z_t$, $t \geq \tau^\diamond + 1$, we have

$$V_\tau - S(\tau^\diamond, H^\diamond, F^\diamond) \geq 2 \sum_{t=1}^T (D_T \epsilon_t)^\top d_t + \sum_{t=\tau+1}^{\tau^\diamond} |d_t|_2^2 =: 2I_1 + I_2. \tag{7.29}$$

By Lemma 8, $|I_1| = O_{\mathbb{P}}(1 + \delta_e(\tau^\diamond - \tau)^{1/2}) + o_{\mathbb{P}}(\delta_e^2(\tau^\diamond - \tau))$, and by Lemma 9, $I_2 \geq c_1(1 + o_{\mathbb{P}}(1))\delta_e^2(\tau^\diamond - \tau)$, some constant $c_1 > 0$. Since $V_{\tau^*} \leq S(\tau^\diamond, H^\diamond, F^\diamond)$, the left hand side of (7.29) should not be positive, therefore $-2I_1$ should be larger than I_2 and thus we have $|\tau^* - \tau^\diamond| = O_{\mathbb{P}}(\delta_e^{-2})$.

Now we prove that plugging in estimated Z_t would not affect our estimation precision. Recall $\hat{S}(\tau, H, F)$ is $S(\tau, H, F)$ with $D_T Z_t$ replaced by \hat{Z}_t . Let $\hat{V}_\tau = \min_{H, F} \hat{S}(\tau, H, F)$ and $(\hat{H}_\tau, \hat{F}_\tau) = \operatorname{argmin}_{H, F} \hat{S}(\tau, H, F)$. Denote

$$\hat{\epsilon}_t = \begin{cases} (I - \mathcal{B}(H^\diamond))\hat{Z}_t = (I - \mathcal{B}(H^\diamond))(\hat{Z}_t - D_T Z_t) + D_T \epsilon_t, & \text{if } t \leq \tau^\diamond \\ (I - \mathcal{B}(F^\diamond))(\hat{Z}_t - D_T Z_t) + D_T \epsilon_t, & \text{if } t > \tau^\diamond. \end{cases} \tag{7.30}$$

Let \hat{d}_t be d_t with H_τ (resp. $F_\tau, D_T Z_t$) replaced by \hat{H}_τ (resp. \hat{F}_τ, \hat{Z}_t). Then similar to (7.29) we have

$$\hat{V}_\tau - \hat{S}(\tau^\diamond, H^\diamond, F^\diamond) \geq 2 \sum_{t=1}^T \hat{\epsilon}_t^\top \hat{d}_t + \sum_{t=\tau+1}^{\tau^\diamond} |\hat{d}_t|_2^2 = 2\hat{I}_1 + \hat{I}_2.$$

Let $\hat{\Gamma}_i$ be Γ_i with Z_t therein replaced by $D_T^{-1} \hat{Z}_t$ and \hat{U}_{ϵ_i} be U_{ϵ_i} with ϵ_i replaced by $D_T^{-1} \hat{\epsilon}_i$. By Theorem 1 and Lemma 4,

$$\sum_{i=1}^3 |\hat{\Gamma}_i - \Gamma_i|_F^2 = \sum_{t=1}^T |D_T^{-1} \hat{Z}_t - Z_t|_2^2 = o_{\mathbb{P}}(1) \quad \text{and thus} \quad \sum_{i=1}^3 |\hat{U}_{\epsilon_i} - U_{\epsilon_i}|_F^2 = o_{\mathbb{P}}(1). \tag{7.31}$$

By (7.37), $|\Gamma_i|_F = O_{\mathbb{P}}(T_i^{1/2})$, where

$$T_i = \tau \text{ (resp. } \tau^\diamond - \tau, T - \tau^\diamond) \text{ for } i = 1 \text{ (resp. } i = 2, 3). \quad (7.32)$$

Then we have

$$|\Gamma_i \Gamma_i^\top - \hat{\Gamma}_i \hat{\Gamma}_i^\top|_F \leq (|\Gamma_i|_F + |\hat{\Gamma}_i|_F) |\Gamma_i - \hat{\Gamma}_i|_F = o_{\mathbb{P}}(T_i^{1/2}) \text{ and } |U_{\epsilon_i} \Gamma_i^\top - \hat{U}_{\epsilon_i} \hat{\Gamma}_i^\top|_F = O_{\mathbb{P}}(T_i^{1/2}). \quad (7.33)$$

With above bounds, same order of bounds as in Lemmas 8 and 9 can be obtained for \hat{I}_1 and \hat{I}_2 . Hence by the same argument as for τ^* case with Z_t replaced by $D_T^{-1} \hat{Z}_t$ and ϵ_t replaced by $D_T^{-1} \hat{\epsilon}_t$, we have $|\hat{\tau} - \tau^\diamond| = O_{\mathbb{P}}(\delta_e^{-2})$ in view of $\hat{V}_{\hat{\tau}} \leq \hat{S}(\tau^\diamond, H^\diamond, F^\diamond)$. \square

LEMMA 8. *Under assumptions of Theorem 2, we have $I_1 = O_{\mathbb{P}}(1 + \delta_e(\tau^\diamond - \tau)^{1/2}) + o_{\mathbb{P}}(\delta_e^2(\tau^\diamond - \tau))$.*

Proof. We shall first show the part $\tau^\diamond + 1 \leq t \leq T$. Recall $H^\diamond = D_T E \tilde{D}_T^{-1}$, $F^\diamond = D_T \tilde{E} \tilde{D}_T^{-1}$ and $\Gamma_{3,t}$ is the t th column of Γ_3 . Note $\zeta_2 = E \Gamma_2 + U_{\epsilon_2}$ and $\zeta_3 = \tilde{E} \Gamma_3 + U_{\epsilon_3}$. Then for $t \in [\tau^\diamond + 1, T]$, by (7.27),

$$\begin{aligned} d_t &= D_T [\tilde{E} - (\zeta_2 \Gamma_2^\top + \zeta_3 \Gamma_3^\top)(\Gamma_2 \Gamma_2^\top + \Gamma_3 \Gamma_3^\top)^{-1}] \Gamma_{3,t-\tau^\diamond} \\ &= -D_T (U_{\epsilon_2} \Gamma_2^\top + U_{\epsilon_3} \Gamma_3^\top)(\Gamma_2 \Gamma_2^\top + \Gamma_3 \Gamma_3^\top)^{-1} \Gamma_{3,t-\tau^\diamond} + D_T (\tilde{E} - E) \Gamma_2 \Gamma_2^\top (\Gamma_2 \Gamma_2^\top + \Gamma_3 \Gamma_3^\top)^{-1} \Gamma_{3,t-\tau^\diamond}. \end{aligned} \quad (7.34)$$

Consequently

$$\begin{aligned} \sum_{t=\tau^\diamond+1}^T (D_T \epsilon_t)^\top d_t &= -\text{tr} \{ D_T (U_{\epsilon_2} \Gamma_2^\top + U_{\epsilon_3} \Gamma_3^\top)(\Gamma_2 \Gamma_2^\top + \Gamma_3 \Gamma_3^\top)^{-1} \Gamma_3 U_{\epsilon_3}^\top D_T^\top \} \\ &\quad + \text{tr} \{ D_T (\tilde{E} - E) \Gamma_2 \Gamma_2^\top (\Gamma_2 \Gamma_2^\top + \Gamma_3 \Gamma_3^\top)^{-1} \Gamma_3 U_{\epsilon_3}^\top D_T^\top \} =: -I_{11} + I_{12}. \end{aligned} \quad (7.35)$$

let $\tilde{\Gamma}_i$ be Γ_i with Z_t replaced by \mathcal{Z}_t in (3.1). Note $U_{\epsilon_3} \tilde{\Gamma}_3^\top = \sum_{t=\tau^\diamond+1}^T \epsilon_t \tilde{\Gamma}_{3,t-\tau^\diamond}^\top$. Hence for any $1 \leq i, j \leq L$, $1 \leq l \leq M$, by Lemma 4, $\delta_k := \|\epsilon_{k,i} \mathcal{Z}_{k-l,j} - \epsilon_{k,i} \mathcal{Z}_{k-l,j,\{0\}}\|_{q'} = \|\epsilon_{k,i}\|_{q'} \|\mathcal{Z}_{k-l,j} - \mathcal{Z}_{k-l,j,\{0\}}\|_{q'} \lesssim \gamma_e^k$, $k \geq 1$, and $\delta_0 := \|\epsilon_{0,i} \mathcal{Z}_{-l,j} - \epsilon'_{0,i} \mathcal{Z}_{-l,j}\|_{q'} \leq 2 \|\epsilon_{0,i}\|_{q'} \|\mathcal{Z}_{-l,j}\|_{q'} \lesssim 1$, where the constant in \lesssim is independent of T . Hence $\sum_{k \geq 0} \delta_k < \infty$ and

$$\sigma^2 := \mathbb{E} \left(\left(\sum_{t=\tau^\diamond+1}^T \epsilon_{t,i} \mathcal{Z}_{t-l,j} \right)^2 \right) = (T - \tau^\diamond) \mathbb{E}(\epsilon_{T,i}^2) \mathbb{E}(\mathcal{Z}_{T-l,j}^2).$$

By the CLT in Lemma 3, $\sigma^{-2} \sum_{t=\tau^\diamond+1}^T \epsilon_{t,i} \mathcal{Z}_{t-l,j} \Rightarrow N(0, 1)$. By Lemma 4, $\sum_{t=1}^T |Z_t - \mathcal{Z}_t|_2 = O_{\mathbb{P}}(\sum_{t \geq 0} \gamma_e^t)$, hence $\|U_{\epsilon_3}(\Gamma_3 - \tilde{\Gamma}_3)^\top\|_{F|q'} \leq \sum_{t=1}^T \|\epsilon_t\|_{q'} \|Z_t - \mathcal{Z}_t\|_{q'} = O(1)$. Therefore

$$\|U_{\epsilon_3} \Gamma_3^\top\|_{F|q'} = O_{\mathbb{P}}((T - \tau^\diamond)^{1/2}). \quad (7.36)$$

Since $\Gamma_3 \Gamma_3^\top = \sum_{t=\tau^\diamond+1}^T \Gamma_{3,t-\tau^\diamond} \Gamma_{3,t-\tau^\diamond}^\top$, using the same argument as above leads to

$$|\tilde{\Gamma}_3 \tilde{\Gamma}_3^\top - \mathbb{E}(\tilde{\Gamma}_3 \tilde{\Gamma}_3^\top)|_F \leq \sum_{i,j=1}^{ML} \left| \sum_{t=\tau^\diamond+1}^T (\tilde{\Gamma}_{3,t-\tau^\diamond,i} \tilde{\Gamma}_{3,t-\tau^\diamond,j} - \mathbb{E} \tilde{\Gamma}_{3,t-\tau^\diamond,i} \tilde{\Gamma}_{3,t-\tau^\diamond,j}) \right|_F = O_{\mathbb{P}}((T - \tau^\diamond)^{1/2}).$$

Since $\mathbb{E}(\tilde{\Gamma}_{3,t} \tilde{\Gamma}_{3,t}^\top) = \Sigma^{(r)}$, we have $\mathbb{E} \tilde{\Gamma}_3 \tilde{\Gamma}_3^\top = (T - \tau^\diamond) \Sigma^{(r)}$. Again by Lemma 4, $|\Gamma_3 \Gamma_3^\top - \tilde{\Gamma}_3 \tilde{\Gamma}_3^\top|_F^2 \leq \sum_{t=1}^T (|Z_t|_2^2 + |\mathcal{Z}_t|_2^2) |Z_t - \mathcal{Z}_t|_2^2 = O_{\mathbb{P}}(1)$. Therefore

$$|\Gamma_3 \Gamma_3^\top - \mathbb{E}(\Gamma_3 \Gamma_3^\top)|_F = O_{\mathbb{P}}((T - \tau^\diamond)^{1/2}) \quad \text{and} \quad |\mathbb{E}(\Gamma_3 \Gamma_3^\top) - (T - \tau^\diamond) \Sigma^{(r)}|_F = O_{\mathbb{P}}(1). \quad (7.37)$$

Since $\lambda_{\min}(\Sigma^{(r)}) \geq c > 0$, $|(\Gamma_3 \Gamma_3^\top)^{-1}|_2 = O_{\mathbb{P}}((T - \tau^\diamond)^{-1})$. Hence by a similar argument for $\Gamma_2 U_{\epsilon_2}^\top$ and $\Gamma_2 \Gamma_2^\top$, we have $I_{11} = O_{\mathbb{P}}(1)$. Recall $|\tilde{E} - E|_2 = \delta_e$. By the same argument as in I_{11} , we have $I_{12} = O_{\mathbb{P}}(\delta_e(\tau^\diamond - \tau) T^{-1/2}) = o_{\mathbb{P}}(\delta_e^2(\tau^\diamond - \tau))$.

Then we comment on the cases of $1 \leq t \leq \tau$ and $\tau + 1 \leq t \leq \tau^\diamond$. Note

$$\sum_{t=1}^{\tau} (D_T \epsilon_t)^\top d_t = -\text{tr}\{D_T U_{\epsilon_1} \Gamma_1^\top (\Gamma_1 \Gamma_1^\top)^{-1} \Gamma_1 U_{\epsilon_1}^\top D_T^\top\},$$

and

$$\begin{aligned} \sum_{t=\tau+1}^{\tau^\diamond} (D_T \epsilon_t)^\top d_t &= -\text{tr}\{D_T (U_{\epsilon_2} \Gamma_2^\top + U_{\epsilon_3} \Gamma_3) (\Gamma_2 \Gamma_2^\top + \Gamma_3 \Gamma_3^\top)^{-1} \Gamma_2 U_{\epsilon_2}^\top D_T^\top\} \\ &\quad + \text{tr}\{D_T (E - \tilde{E}) \Gamma_3 \Gamma_3^\top (\Gamma_2 \Gamma_2^\top + \Gamma_3 \Gamma_3^\top)^{-1} \Gamma_2 U_{\epsilon_2}^\top D_T^\top\}. \end{aligned}$$

Hence similar argument as for $\sum_{t=\tau^\diamond+1}^T (D_T \epsilon_t)^\top d_t$ part leads to $\sum_{t=\tau+1}^{\tau^\diamond} (D_T \epsilon_t)^\top d_t = O_{\mathbb{P}}(1 + \delta_e(\tau^\diamond - \tau)^{1/2})$ and $\sum_{t=1}^{\tau} (D_T \epsilon_t)^\top d_t = O_{\mathbb{P}}(1)$. \square

LEMMA 9. Under assumptions of Theorem 2, we have $I_2 \geq c_1(1 + o_{\mathbb{P}}(1))\delta_e^2(\tau^\diamond - \tau)$, where $c_1 > 0$ only depending on the largest and smallest eigenvalues of $\Sigma^{(l)}$, $\Sigma^{(r)}$ and D_T .

Proof. Similar to d_t in (7.34), for $t \in [\tau + 1, \tau^\diamond]$ we have

$$d_t = D_T(E - \tilde{E})\Gamma_3\Gamma_3^\top(\Gamma_2\Gamma_2^\top + \Gamma_3\Gamma_3^\top)^{-1}\Gamma_{2,t-\tau} + D_T(U_{\epsilon_2}\Gamma_2^\top + U_{\epsilon_3}\Gamma_3^\top)(\Gamma_2\Gamma_2^\top + \Gamma_3\Gamma_3^\top)^{-1}\Gamma_{2,t-\tau}.$$

Hence

$$\begin{aligned} I_2 &\geq \text{tr}\{D_T(E - \tilde{E})\Gamma_3\Gamma_3^\top(\Gamma_2\Gamma_2^\top + \Gamma_3\Gamma_3^\top)^{-1}\Gamma_2\Gamma_2^\top(\Gamma_2\Gamma_2^\top + \Gamma_3\Gamma_3^\top)^{-1}\Gamma_3\Gamma_3^\top(E - \tilde{E})^\top D_T^\top\} \\ &+ \text{tr}\{D_T(E - \tilde{E})\Gamma_3\Gamma_3^\top(\Gamma_2\Gamma_2^\top + \Gamma_3\Gamma_3^\top)^{-1}\Gamma_2\Gamma_2^\top(\Gamma_2\Gamma_2^\top + \Gamma_3\Gamma_3^\top)^{-1}(U_{\epsilon_2}\Gamma_2^\top + U_{\epsilon_3}\Gamma_3^\top)D_T^\top\} =: I_{21} + I_{22}. \end{aligned}$$

Note for some constant $c, C > 0$, $\lambda_{\min}(\Sigma^{(l)}), \lambda_{\min}(\Sigma^{(r)}) \geq c$ and $\lambda_{\max}(\Sigma^{(l)}), \lambda_{\max}(\Sigma^{(r)}) \leq C$. Hence by (7.37) and a similar argument for $\Gamma_2\Gamma_2^\top$, $\lambda_{\min}(\Gamma_2\Gamma_2^\top) \geq (c + o_{\mathbb{P}}(1))(\tau^\diamond - \tau)$ and $\lambda_{\min}(\Gamma_3\Gamma_3^\top) \geq (c + o_{\mathbb{P}}(1))(T - \tau^\diamond)$. Since $T - \tau^\diamond \asymp T$,

$$I_{21} \geq c_1(1 + o_{\mathbb{P}}(1))\delta_e^2(\tau^\diamond - \tau),$$

where $c_1 > 0$ only depends on c, C and $\lambda_{\min}(D_T)$. Same as I_{12} in the proof of Lemma 8, we have $I_{22} = o_{\mathbb{P}}(\delta_e^2(\tau - \tau^\diamond))$. \square

7.4 Proof of Theorem 3

Proof of Theorem 3. Recall definitions of H_τ and F_τ in Subsection 7.3. We shall show the case $\hat{\tau} < \tau^\diamond$ and the other direction can be similarly dealt with.

For \hat{H} : we will work on $H_{\hat{\tau}}$ first, and then show the difference of $H_{\hat{\tau}}$ and \hat{H} is negligible. Let $\tau = \hat{\tau}$ in the construction of Γ_i and U_{ϵ_i} . Note

$$D_T^{-1}(H_{\hat{\tau}} - H^\diamond)\tilde{D}_T = U_{\epsilon_1}\Gamma_1^\top(\Gamma_1\Gamma_1^\top)^{-1}. \quad (7.38)$$

Same argument as (7.37), $\tau^{-1}\Gamma_1\Gamma_1^\top \rightarrow \Sigma^{(l)}$ in probability. For any matrix $A \in \mathbb{R}^{L \times LM}$ with $|A|_F = 1$, we have

$$\text{vect}(A)^\top \text{vect}(U_{\epsilon_1}\Gamma_1^\top \Sigma^{(l)-1}) = \sum_{t=1}^{\hat{\tau}} \text{vect}(A)^\top \text{vect}(\epsilon_t \Gamma_{1,t}^\top \Sigma^{(l)-1}) =: \sum_{t=1}^{\hat{\tau}} \xi_t.$$

Let $\tilde{\xi}_t$ (resp. $\tilde{\Gamma}_1$) be ξ_t (resp. Γ_1) and with Z_t replaced by \mathcal{Z}_t and $\delta_t = \|\mathcal{P}_0 \tilde{\xi}_t\|_{q'}$. Then

$$\delta_t \leq |A|_F \|\epsilon_t(\tilde{\Gamma}_{1,t}^\top - \tilde{\Gamma}_{1,t,\{0\}}) \Sigma^{(l)-1}\|_F \leq \|\epsilon_t\|_2 \|\tilde{\Gamma}_{1,t}^\top - \tilde{\Gamma}_{1,t,\{0\}}\|_2 \|\Sigma^{(l)-1}\|_F, \quad t \geq 1,$$

and $\delta_0 \leq \|\epsilon_0 - \epsilon'_0\|_2 \|\tilde{\Gamma}_{1,1}^\top\|_2 \|\Sigma^{(l)-1}\|_F$. By Lemma 4, $\sum_{t=1}^T \|\mathcal{Z}_t - \mathcal{Z}_{t,\{0\}}\|_2 = O(\sum_{t \geq 0} \gamma_e^t) = O_{\mathbb{P}}(1)$, hence $\sum_{t=1}^{\hat{\tau}} \|\tilde{\Gamma}_{1,t} - \tilde{\Gamma}_{1,t,\{0\}}\|_2 = O_{\mathbb{P}}(1)$ and $\sum_{t=0}^{\infty} \delta_t < \infty$. By Lemma 3,

$$\hat{\tau}^{-1/2} \sum_{t=1}^{\hat{\tau}} \tilde{\xi}_t \Rightarrow N(0, \sigma^2),$$

where $\sigma^2 = \sum_{i \in \mathbb{Z}} \mathbb{E}(\tilde{\xi}_1 \tilde{\xi}_i) = \mathbb{E}(\tilde{\xi}_1^2)$. Let $\Theta := \mathbb{E}(\text{vect}(\epsilon_1 \tilde{\Gamma}_{1,1}^\top \Sigma^{(l)-1}) \text{vect}(\epsilon_1 \tilde{\Gamma}_{1,1}^\top \Sigma^{(l)-1})^\top)$, then $\sigma^2 = \text{vect}(A)^\top \Theta \text{vect}(A)$

Note $\Theta = (N_{i,j})_{1 \leq i,j \leq L}$ where $N_{i,j} \in \mathbb{R}^{LM \times LM}$ with

$$N_{i,j} = \mathbb{E} \left[(\epsilon_{1,i} \tilde{\Gamma}_{1,1}^\top \Sigma^{(l)-1})^\top \epsilon_{1,j} \tilde{\Gamma}_{1,1}^\top \Sigma^{(l)-1} \right] = \mathbb{E}(\epsilon_{1,i} \epsilon_{1,j}) \Sigma^{(l)-1} = \Sigma_{\epsilon,i,j} \Sigma^{(l)-1}.$$

Since $\sum_{t=1}^{\hat{\tau}} \|\tilde{\xi}_t - \xi_t\|_{q'} \leq \sum_{t=1}^{\hat{\tau}} |A|_F \|\epsilon_t\|_2 \|\tilde{\Gamma}_{1,t} - \tilde{\Gamma}_{1,t,\{0\}}\|_2 \|\Sigma^{(l)-1}\|_F = O_{\mathbb{P}}(1)$ and $|\tau^\diamond - \hat{\tau}| = o_{\mathbb{P}}(\tau^\diamond)$, we have

$$\tau^{\diamond-1/2} \text{vect}(U_{\epsilon_1} \Gamma_1^\top \Sigma^{(l)-1}) \Rightarrow N(0, \Theta).$$

By (7.38), $|D_T^{-1}(H_{\hat{\tau}} - \hat{H}) \tilde{D}_T^\top|_F \leq |U_{\epsilon_1} \Gamma_1^\top (\Gamma_1 \Gamma_1^\top)^{-1} - \hat{U}_{\epsilon_1} \hat{\Gamma}_1^\top (\hat{\Gamma}_1 \hat{\Gamma}_1^\top)^{-1}|_F$, where $\hat{\Gamma}_i$ (resp. \hat{U}_{ϵ_i}) is Γ_i (resp. U_{ϵ_i}) with Z_t (resp. ϵ_t) replaced by $D_T^{-1} \hat{Z}_t$ (resp. $D_T^{-1} \hat{\epsilon}_t$ defined in (7.30).) By (7.31), we have $|D_T^{-1}(H_{\hat{\tau}} - \hat{H}) \tilde{D}_T^\top|_F = o_{\mathbb{P}}(T^{-1/2})$ and thus the desired result follows.

For \hat{F} : note

$$\begin{aligned} D_T^{-1}(F_{\hat{\tau}} - F^\diamond) \tilde{D}_T &= (\zeta_2 \Gamma_2^\top + \zeta_3 \Gamma_3^\top) (\Gamma_2 \Gamma_2^\top + \Gamma_3 \Gamma_3^\top)^{-1} - \tilde{E} \\ &= (E - \tilde{E}) \Gamma_2 \Gamma_2^\top (\Gamma_2 \Gamma_2^\top + \Gamma_3 \Gamma_3^\top)^{-1} + (U_{\epsilon_2} \Gamma_2^\top + U_{\epsilon_3} \Gamma_3^\top) (\Gamma_2 \Gamma_2^\top + \Gamma_3 \Gamma_3^\top)^{-1} =: I_1 + I_2. \end{aligned}$$

By (7.37), $|I_1|_F = O_{\mathbb{P}}(\delta_e(\tau^\diamond - \hat{\tau})/T)$. Then Theorem 2 and $\delta_e T^{1/2} \rightarrow \infty$ imply $|I_1|_F = O_{\mathbb{P}}(\delta_e^{-1} T^{-1}) = o_{\mathbb{P}}(T^{-1/2})$.

Decompose I_2 into

$$\begin{aligned} (T - \tau^\diamond)^{1/2} I_2 &= (T - \tau^\diamond)^{-1/2} U_{\epsilon_3} \Gamma_3^\top \Sigma^{(r)-1} + (T - \tau^\diamond)^{1/2} U_{\epsilon_2} \Gamma_2^\top (\Gamma_2 \Gamma_2^\top + \Gamma_3 \Gamma_3^\top)^{-1} \\ &\quad + (T - \tau^\diamond)^{-1/2} U_{\epsilon_3} \Gamma_3^\top (\Gamma_2 \Gamma_2^\top + \Gamma_3 \Gamma_3^\top)^{-1} (\Gamma_2 \Gamma_2^\top + \Gamma_3 \Gamma_3^\top - (T - \tau^\diamond) \Sigma^{(r)}) \Sigma^{(r)-1}. \end{aligned}$$

By (7.36), $|U_{\epsilon_2}\Gamma_2^\top|_F = O_{\mathbb{P}}((\tau^\diamond - \tau)^{1/2})$. By (7.37), $|(T - \tau^\diamond)^{-1}(\Gamma_2\Gamma_2^\top + \Gamma_3\Gamma_3^\top) - \Sigma^{(r)}|_F = o_{\mathbb{P}}(1)$. By Weyl's inequality, with probability tending to 1, the smallest eigenvalue of $(T - \tau^\diamond)^{-1}(\Gamma_2\Gamma_2^\top + \Gamma_3\Gamma_3^\top)$ is lower bounded by some $c > 0$. Then we have

$$|(T - \tau^\diamond)^{1/2}\mathbf{I}_2 - (T - \tau^\diamond)^{-1/2}U_{\epsilon_3}\Gamma_3^\top\Sigma^{(r)-1}|_F = O_{\mathbb{P}}((\tau^\diamond - \hat{\tau})T^{-1}) + o_{\mathbb{P}}(1).$$

By Theorem 2, $\tau^\diamond - \hat{\tau} = O(\delta_e^{-2})$ and $\delta_e^2 T \rightarrow \infty$, hence right hand side of above equality is $o_{\mathbb{P}}(1)$. Same argument as in \hat{H} part, we have

$$(T - \tau^\diamond)^{-1/2}\text{vect}(U_{\epsilon_3}\Gamma_3^\top\Sigma^{(r)-1}) \Rightarrow N(0, \Theta'),$$

where Θ' equals Θ with $\Sigma^{(l)}$ replaced by $\Sigma^{(r)}$. Similar as the \hat{H} part, we have $|D_T^{-1}(F_{\hat{\tau}} - \hat{F})\tilde{D}_T|_F = o_{\mathbb{P}}(T^{-1/2})$ and we complete the proof. \square

7.5 Proof of Theorem 4

Proof of Theorem 4. By Theorem 2, $|\hat{\tau} - \tau^\diamond| = O_{\mathbb{P}}(\delta_e^{-2})$. Hence we shall work on set $\Omega(m) := \{\tau : 0 \leq \tau^\diamond - \tau \leq m\delta_e^{-2}\}$, some constant $m > 0$. The other direction $\tau > \tau^\diamond$ can be similarly dealt with. Recall that $\mathcal{Q}^{(l)} = \delta_e^{-2}(E - \tilde{E})\Sigma^{(l)}(E - \tilde{E})^\top$. Define a standard Wiener process on $[0, \infty)$ as $W(s)$ for $1 \leq s \leq m$. The proof involves the following steps:

Step 1 Show $\sup_{\tau \in \Omega(m)} |V_\tau - V_{\tau^\diamond} - \text{tr}\{2(E - \tilde{E})\Gamma_2 U_{\epsilon_2}^\top + (\tau^\diamond - \tau)\delta_e^2 \mathcal{Q}^{(l)}\}| = o_{\mathbb{P}}(1)$.

Step 2 For $\tau = \tau^\diamond - \lfloor \delta_e^{-2}s \rfloor$, $0 \leq s < m$, show

$$\text{tr}\{(E - \tilde{E})\Gamma_2 U_{\epsilon_2}^\top\} \Rightarrow \text{tr}^{1/2}\{\mathcal{Q}^{(l)}\Sigma_\epsilon\}W(s).$$

Step 3 Show $\sup_{\tau \in \Omega(m)} |V_\tau - \hat{V}_\tau| = o_{\mathbb{P}}(1)$.

Then combining above steps, the desired result follows.

Step 1. Suppose that we know the true break point, and we denote $(H^{*\diamond}, F^{*\diamond}) = \operatorname{argmin}_{H, F} S(\tau^\diamond, H, F)$. Let $r_t^\diamond = (\mathcal{B}(H^\diamond) - \mathcal{B}(H^{*\diamond}))D_T Z_t + D_T \epsilon_t$, if $t \leq \tau^\diamond$; $r_t^\diamond = (\mathcal{B}(F^\diamond) - \mathcal{B}(F^{*\diamond}))D_T Z_t + D_T \epsilon_t$, if $t > \tau^\diamond$. Then

$$V_{\tau^\diamond} = \sum_{t=1}^T |r_t^\diamond|_2^2.$$

Recall definitions of F_τ and H_τ in Subsection 7.3. Let $d_t^\diamond = (\mathcal{B}(H^{*\diamond}) - \mathcal{B}(H_\tau))D_T Z_t$, if $1 \leq t \leq \tau$; $d_t^\diamond = (\mathcal{B}(H^{*\diamond}) - \mathcal{B}(F_\tau))D_T Z_t$, if $\tau + 1 \leq t \leq \tau^\diamond$; $d_t^\diamond = (\mathcal{B}(F^{*\diamond}) - \mathcal{B}(F_\tau))D_T Z_t$, if $\tau^\diamond + 1 \leq t \leq T$. Then

$$V_\tau - V_{\tau^\diamond} = \sum_{t=1}^T |d_t^\diamond + r_t^\diamond|_2^2 - \sum_{t=1}^T |r_t^\diamond|_2^2 = \sum_{t=1}^T |d_t^\diamond|_2^2 + 2 \sum_{t=1}^T r_t^{\diamond\top} d_t^\diamond = \mathbf{I}_1 + 2\mathbf{I}_2. \quad (7.39)$$

Part \mathbf{I}_1 . We shall first deal with $\tau + 1 \leq t \leq \tau^\diamond$ part. Note

$$H^{*\diamond} = D_T(\zeta_1 \Gamma_1^\top + \zeta_2 \Gamma_2^\top)(\Gamma_1 \Gamma_1^\top + \Gamma_2 \Gamma_2^\top)^{-1} \tilde{D}_T^{-1} = D_T \left(E + (U_{\epsilon_1} \Gamma_1^\top + U_{\epsilon_2} \Gamma_2^\top)(\Gamma_1 \Gamma_1^\top + \Gamma_2 \Gamma_2^\top)^{-1} \right) \tilde{D}_T^{-1}$$

and

$$\begin{aligned} F_\tau &= D_T(\zeta_2 \Gamma_2^\top + \zeta_3 \Gamma_3^\top)(\Gamma_2 \Gamma_2^\top + \Gamma_3 \Gamma_3^\top)^{-1} \tilde{D}_T^{-1} \\ &= D_T \left(\tilde{E} + (U_{\epsilon_2} \Gamma_2^\top + U_{\epsilon_3} \Gamma_3^\top + (E - \tilde{E}) \Gamma_2 \Gamma_2^\top)(\Gamma_2 \Gamma_2^\top + \Gamma_3 \Gamma_3^\top)^{-1} \right) \tilde{D}_T^{-1}. \end{aligned}$$

Hence by similar argument as (7.36) and (7.37), we obtain bounds $|\Gamma_i \Gamma_i^\top|_2 = O_{\mathbb{P}}(T_i)$, $|U_{\epsilon_i} \Gamma_i^\top|_2 = O_{\mathbb{P}}(T_i^{1/2})$ and $|(\Gamma_i \Gamma_i^\top)^{-1}|_2 = O_{\mathbb{P}}(T_i^{-1})$, $i \in \{1, 2, 3\}$ and T_i is defined in (7.32). Then

$$|H^{*\diamond} - F_\tau - D_T(E - \tilde{E})\tilde{D}_T^{-1}|_2 = O_{\mathbb{P}}(T^{-1/2} + \delta_e(\tau^\diamond - \tau)/T). \quad (7.40)$$

Thus we have

$$\sum_{t=\tau+1}^{\tau^\diamond} |d_t^\diamond|_2^2 = \operatorname{tr}\{(H^{*\diamond} - F_\tau)\tilde{D}_T \Gamma_2 \Gamma_2^\top \tilde{D}_T^\top (H^{*\diamond} - F_\tau)^\top\} = \operatorname{tr}\{D_T(E - \tilde{E})\Gamma_2 \Gamma_2^\top (E - \tilde{E})^\top D_T^\top\} + o_{\mathbb{P}}(1).$$

Same argument leads to $\sum_{\tau^\diamond+1 \leq t \leq T} |d_t^\diamond|_2^2 = o_{\mathbb{P}}(1)$ and $\sum_{1 \leq t \leq \tau} |d_t^\diamond|_2^2 = o_{\mathbb{P}}(1)$. Since $\mathbb{E}\Gamma_2 \Gamma_2^\top = (\tau^\diamond - \tau)\Sigma^{(l)}$. Similar to (7.37), we have $|\Gamma_2 \Gamma_2^\top - (\tau^\diamond - \tau)\Sigma^{(l)}|_F = O_{\mathbb{P}}((\tau^\diamond - \tau)^{1/2})$. Hence

$$\sum_{t=1}^T |d_t^\diamond|_2^2 = (\tau^\diamond - \tau) \operatorname{tr}\{(E - \tilde{E})\Sigma^{(l)}(E - \tilde{E})^\top D_T^\top D_T\} + o_{\mathbb{P}}(1).$$

Part I₂. Note $H^{*\diamond} - H^\diamond = D_T(U_{\epsilon_1}\Gamma_1^\top + U_{\epsilon_2}\Gamma_2^\top)(\Gamma_1\Gamma_1^\top + \Gamma_2\Gamma_2^\top)^{-1}\tilde{D}_T^{-1}$. Hence $|H^{*\diamond} - H^\diamond|_2 = O_{\mathbb{P}}(T^{-1/2})$. By (7.40),

$$\begin{aligned} \sum_{t=\tau+1}^{\tau^\diamond} r_t^{\diamond\top} d_t^\diamond &= \text{tr}\left\{(H^{*\diamond} - F_\tau)\tilde{D}_T\Gamma_2\Gamma_2^\top\tilde{D}_T^\top(H^\diamond - H^{*\diamond})^\top + (H^{*\diamond} - F_\tau)\tilde{D}_T\Gamma_2U_{\epsilon_2}^\top D_T^\top\right\} \\ &= \text{tr}\{(E - \tilde{E})\Gamma_2U_{\epsilon_2}^\top D_T^\top D_T\} + o_{\mathbb{P}}(1). \end{aligned}$$

Same argument leads to $\sum_{\tau^\diamond+1 \leq t \leq T} r_t^{\diamond\top} d_t^\diamond = o_{\mathbb{P}}(1)$ and $\sum_{1 \leq t \leq \tau} r_t^{\diamond\top} d_t^\diamond = o_{\mathbb{P}}(1)$. Then desired result follows by noting $|D_T^\top D_T - I_L|_F = O_{\mathbb{P}}(T^{-1/2})$.

Step 2. Note

$$\text{tr}\{(E - \tilde{E})\Gamma_2U_{\epsilon_2}^\top\} = \sum_{t=\tau+1}^{\tau^\diamond} \xi_t, \quad \text{where } \xi_t = \epsilon_t^\top (E - \tilde{E})\Gamma_{2,t-\tau}.$$

Let $\tilde{\xi}_k$ be ξ_k with Z_t replaced by $\mathcal{Z}_t^{(l)}$. For $\delta_k := \|\mathcal{P}_0\tilde{\xi}_k\|_{q'}$, since $\|\mathcal{P}_0|\tilde{\Gamma}_{2,t-\tau}|_2\|_{q'} \leq \| |\tilde{\Gamma}_{2,t-\tau} - \tilde{\Gamma}_{2,t-\tau,\{0\}}|_2 \|_{q'}$, we have

$$\delta_k \leq \delta_e \| |\tilde{\Gamma}_{2,k-\tau} - \tilde{\Gamma}_{2,k-\tau,\{0\}}|_2 \|_{q'} \|\epsilon_k\|_2 \|_{q'}, \quad k \geq 1,$$

and $\delta_0 \leq \delta_e \| |\tilde{\Gamma}_{2,-\tau}|_2 \|_{q'} \|\epsilon_0 - \epsilon'_0\|_2 \|_{q'}$. By Lemma 4, $\sum_{t \geq 0} \| |\tilde{Z}_t - \tilde{Z}_{t,\{0\}}|_2 \|_{q'} = O(1)$, hence $\delta_e^{-1} \sum_{k \geq 0} \delta_k < \infty$. By Theorem 3 in Wu (2011), the invariance principle holds as $\delta_e \rightarrow 0$,

$$\frac{\delta_e}{\sigma} \sum_{t=\tau^\diamond - \lfloor \delta_e^{-2}s \rfloor}^{\tau^\diamond} \tilde{\xi}_t \Rightarrow W(s), \quad (7.41)$$

where $W(\cdot)$ is a standard Wiener process on $[0, \infty)$ and

$$\sigma^2 = \mathbb{E}(\tilde{\xi}_{\tau^\diamond}^2) = \text{tr}\left\{\mathbb{E}(\epsilon_{\tau^\diamond}\epsilon_{\tau^\diamond}^\top)(E - \tilde{E})\mathbb{E}(\Gamma_{2,\tau^\diamond-\tau}\Gamma_{2,\tau^\diamond-\tau}^\top)(E - \tilde{E})^\top\right\} = \text{tr}\{\Sigma_\epsilon(E - \tilde{E})\Sigma^{(l)}(E - \tilde{E})^\top\}.$$

Then $\sigma/\delta_e \asymp 1$ in view of the largest and smallest eigenvalues of $\Sigma^{(l)}$ and Σ_ϵ are bounded above and below by some positive constants. Since $\| |(\Gamma_2 - \tilde{\Gamma}_2)U_{\epsilon_2}^\top|_F \|_{q'} \leq \sum_{t=\tau+1}^{\tau^\diamond} \| |\Gamma_{2,t-\tau} - \tilde{\Gamma}_{2,t-\tau}|_2 \|_{q'} \|\epsilon_t\|_2 \|_{q'} = O(1)$, and $\delta_e \rightarrow 0$, (7.41) holds for ξ_t replaced by $\tilde{\xi}_t$.

Step 3. Let $(\hat{H}^\diamond, \hat{F}^\diamond) = \text{argmin}_{H,F} \hat{S}(\tau^\diamond, H, F)$, then $(\hat{H}^\diamond, \hat{F}^\diamond)$ is $(H^{*\diamond}, F^{*\diamond})$ with Z_t and ϵ_t replaced by $D_T^{-1}\hat{Z}_t$ and $D_T^{-1}\hat{\epsilon}_t$. Let \hat{d}_t^\diamond and \hat{r}_t^\diamond be d_t^\diamond and r_t^\diamond with $H^{*\diamond}$, $F^{*\diamond}$, $D_T Z_t$ replaced by \hat{H}^\diamond , \hat{F}^\diamond , \hat{Z}_t respectively.

Then by (7.39),

$$|V_\tau - \hat{V}_\tau| \leq \sum_{t=1}^T \|d_t^\diamond\|_2^2 - \|\hat{d}_t^\diamond\|_2^2 + 2 \sum_{t=1}^T (|r_t^{\diamond\top}(d_t^\diamond - \hat{d}_t^\diamond)| + |(r_t^\diamond - \hat{r}_t^\diamond)^\top \hat{d}_t^\diamond|).$$

Since by Cauchy-Schwartz inequality, $\sum_{t=1}^T \|d_t^\diamond\|_2^2 - \|\hat{d}_t^\diamond\|_2^2 \leq (\sum_{t=1}^T \|d_t^\diamond - \hat{d}_t^\diamond\|_2^2)^{1/2} (\sum_{t=1}^T (\|d_t^\diamond\|_2 + \|\hat{d}_t^\diamond\|_2)^2)^{1/2}$. By (7.33),

$$\sum_{t=1}^T \|d_t^\diamond - \hat{d}_t^\diamond\|_2^2 = o_{\mathbb{P}}(1) \quad \text{and} \quad \sum_{t=1}^T \|r_t^\diamond - \hat{r}_t^\diamond\|_2^2 = o_{\mathbb{P}}(1).$$

Hence $\sum_{t=1}^T \|d_t^\diamond\|_2^2 - \|\hat{d}_t^\diamond\|_2^2 = o_{\mathbb{P}}(1)$. Similarly we have $\sum_{t=1}^T |r_t^{\diamond\top}(d_t^\diamond - \hat{d}_t^\diamond)| \leq (\sum_{t=1}^T \|r_t^\diamond\|_2^2)^{1/2} (\sum_{t=1}^T \|d_t^\diamond - \hat{d}_t^\diamond\|_2^2)^{1/2} = o_{\mathbb{P}}(1)$ and $\sum_{t=1}^T |(r_t^\diamond - \hat{r}_t^\diamond)^\top \hat{d}_t^\diamond| \leq (\sum_{t=1}^T \|r_t^\diamond - \hat{r}_t^\diamond\|_2^2)^{1/2} (\sum_{t=1}^T \|\hat{d}_t^\diamond\|_2^2)^{1/2} = o_{\mathbb{P}}(1)$. \square

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. (1997). Estimation of a change point in multiple regression models. *The Review of Economics and Statistics* 79(4), 551–563.
- Bai, J., X. Han, and Y. Shi (2016). Estimation and inference of change points in high dimensional factor models. *Manuscript*.
- 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.
- 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.
- El Machkouri, M., D. Volný, and W. B. Wu (2013). A central limit theorem for stationary random fields. *Stochastic Process. Appl.* 123(1), 1–14.
- 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.
- Mihoci, A. (2017). Modelling limit order book volume covariance structures. In *Advances in Statistical Methodologies and Their Application to Real Problems*. InTech.
- 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.
- van de Geer, S. (2000). *Empirical Processes in M-estimation*, Volume 6. Cambridge university press.
- 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. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences of the United States of America* 102(40), 14150–14154.
- Wu, W. B. (2011). Asymptotic theory for stationary processes. *Stat. Interface* 4(2), 207–226.
- 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.

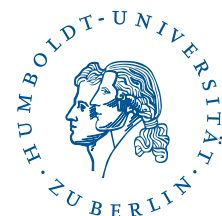
SFB 649 Discussion Paper Series 2017

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 001 "Fake Alpha" by Marcel Müller, Tobias Rosenberger and Marliese Uhrig-Homburg, January 2017.
- 002 "Estimating location values of agricultural land" by Georg Helbing, Zhiwei Shen, Martin Odening and Matthias Ritter, January 2017.
- 003 "FRM: a Financial Risk Meter based on penalizing tail events occurrence" by Lining Yu, Wolfgang Karl Härdle, Lukas Borke and Thijs Benschop, January 2017.
- 004 "Tail event driven networks of SIFIs" by Cathy Yi-Hsuan Chen, Wolfgang Karl Härdle and Yarema Okhrin, January 2017.
- 005 "Dynamic Valuation of Weather Derivatives under Default Risk" by Wolfgang Karl Härdle and Maria Osipenko, February 2017.
- 006 "RiskAnalytics: an R package for real time processing of Nasdaq and Yahoo finance data and parallelized quantile lasso regression methods" by Lukas Borke, February 2017.
- 007 "Testing Missing at Random using Instrumental Variables" by Christoph Breunig, February 2017.
- 008 "GitHub API based QuantNet Mining infrastructure in R" by Lukas Borke and Wolfgang K. Härdle, February 2017.
- 009 "The Economics of German Unification after Twenty-five Years: Lessons for Korea" by Michael C. Burda and Mark Weder, April 2017.
- 010 "Data Science & Digital Society" by Cathy Yi-Hsuan Chen and Wolfgang Karl Härdle, May 2017.
- 011 "The impact of news on US household inflation expectations" by Shih-Kang Chao, Wolfgang Karl Härdle, Jeffrey Sheen, Stefan Trück and Ben Zhe Wang, May 2017.
- 012 "Industry Interdependency Dynamics in a Network Context" by Ya Qian, Wolfgang Karl Härdle and Cathy Yi-Hsuan Chen, May 2017.
- 013 "Adaptive weights clustering of research papers" by Larisa Adamyan, Kirill Efimov, Cathy Yi-Hsuan Chen, Wolfgang K. Härdle, July 2017.
- 014 "Investing with cryptocurrencies - A liquidity constrained investment approach" by Simon Trimborn, Mingyang Li and Wolfgang Karl Härdle, July 2017.
- 015 "(Un)expected Monetary Policy Shocks and Term Premia" by Martin Kliem and Alexander Meyer-Gohde, July 2017.
- 016 "Conditional moment restrictions and the role of density information in estimated structural models" by Andreas Tryphonides, July 2017.
- 017 "Generalized Entropy and Model Uncertainty" by Alexander Meyer-Gohde, August 2017.
- 018 "Social Security Contributions and the Business Cycle" by Anna Almosova, Michael C. Burda and Simon Voigts, August 2017.
- 019 "Racial/Ethnic Differences In Non-Work At Work" by Daniel S. Hamermesh, Katie R. Genadek and Michael C. Burda, August 2017.
- 020 "Pricing Green Financial Products" by Awdesch Melzer, Wolfgang K. Härdle and Brenda López Cabrera, August 2017.
- 021 "The systemic risk of central SIFIs" by Cathy Yi-Hsuan Chen and Sergey Nasekin, August 2017.
- 022 "Das deutsche Arbeitsmarktwunder: Eine Bilanz" by Michael C. Burda and Stefanie Seele, August 2017.

SFB 649, Spandauer Straße 1, D-10178 Berlin
<http://sfb649.wiwi.hu-berlin.de>

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".



SFB 649 Discussion Paper Series 2017

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 023 "Penalized Adaptive Method in Forecasting with Large Information Set and Structure Change" by Xinjue Li, Lenka Zbonakova and Wolfgang Karl Härdle, September 2017.
- 024 "Smooth Principal Component Analysis for High Dimensional Data" by Yingxing Li, Wolfgang K. Härdle and Chen Huang, September 2017.
- 025 "Realized volatility of CO2 futures" by Thijs Benschop and Brenda López Cabrera, September 2017.
- 026 "Dynamic Semiparametric Factor Model with a Common Break" by Likai Chen, Weining Wang and Wei Biao Wu, November 2017.

SFB 649, Spandauer Straße 1, D-10178 Berlin
<http://sfb649.wiwi.hu-berlin.de>

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

