Nonconvex High-Dimensional Time-Varying Coefficient Estimation for Noisy High-Frequency Observations with a Factor Structure

Minseok Shin[a] and Donggyu Kim[b*]
[a]Pohang University of Science and Technology (POSTECH)
[b]University of California, Riverside

November 8, 2024

Abstract

In this paper, we propose a novel high-dimensional time-varying coefficient estimator for noisy high-frequency observations with a factor structure. In high-frequency finance, we often observe that noises dominate the signal of underlying true processes and that covariates exhibit a factor structure due to their strong dependence. Thus, we cannot apply usual regression procedures to analyze high-frequency observations. To handle the noises, we first employ a smoothing method for the observed dependent and covariate processes. Then, to handle the strong dependence of the covariate processes, we apply Principal Component Analysis (PCA) and transform the highly correlated covariate structure into a weakly correlated structure. However, the variables from PCA still contain non-negligible noises. To manage these non-negligible noises and the high dimensionality, we propose a nonconvex penalized regression method for each local coefficient. This method produces consistent but biased local coefficient estimators. To estimate the integrated coefficients, we propose a debiasing scheme and obtain a debiased integrated coefficient estimator using debiased local coefficient estimators. Then, to further account for the sparsity structure of the coefficients, we apply a thresholding scheme to the debiased integrated coefficient estimator. We call this scheme the Factor Adjusted

*Corresponding author. E-mail address: donggyu.kim@ucr.edu.

Thresholded dEbiased Nonconvex LASSO (FATEN-LASSO) estimator. Furthermore, this paper establishes the concentration properties of the FATEN-LASSO estimator and discusses a nonconvex optimization algorithm.

Keywords: debias, diffusion process, PCA, LASSO, smoothing, sparsity.

1 Introduction

Regression models are widely used in statistical analysis. With the wide availability of high-frequency data, increasing attention has been paid to high-frequency regression. The framework of high-frequency regression enables us to accommodate the time variation in the coefficient process, which is often observed in financial practice (Ferson and Harvey, 1999; Kalnina, 2023; Reiß et al., 2015). Thus, various statistical methods have been developed to analyze high-frequency regression. For example, Barndorff-Nielsen and Shephard (2004); Andersen et al. (2005) proposed a realized coefficient estimator, which was constructed using the ratio of realized covariance to realized variance. Mykland and Zhang (2009) estimated the integrated coefficient by aggregating the spot coefficients obtained from local blocks. See also Aït-Sahalia et al. (2020); Oh et al. (2024); Reiß et al. (2015). Chen (2018) suggested the statistical inference for volatility functionals of general Itô semimartingales. Andersen et al. (2021) proposed the measure for market beta dispersion and studied the intra-day variation in market betas. These models and estimation methods perform well under the assumption that the number of factors is finite. Recently, Chen et al. (2024) proposed high-dimensional market beta estimation procedure with a large number of dependent variables, where the number of common factors diverges slowly as the number of dependent variables goes to infinity.

However, in finance, we often encounter a large number of factor candidates (Bali et al., 2011; Cochrane, 2011; Harvey et al., 2016; Hou et al., 2020; McLean and Pontiff, 2016). This causes the curse of dimensionality, and the estimation methods designed for the finite dimension cannot consistently estimate the coefficients. To overcome the curse of dimensionality, LASSO (Tibshirani, 1996), SCAD (Fan and Li, 2001), and the Dantzig selector (Candes and Tao, 2007) are often employed under the sparsity assumption on the model parameters. Loh and Wainwright (2012) introduced the nonconvex high-dimensional regression to handle noisy and missing variables. However, these estimation methods cannot account for the time-varying property of coefficient processes. Recently, to handle both the curse of dimensionality and the time-varying feature of the coefficient process, Kim et al. (2024) proposed a Thresholded dEbiased Dantzig (TED) estimator under the sparsity assumption on the coefficient process. They first employed a time-localized Dantzig selector (Candes and Tao, 2007) to estimate the instantaneous coefficient and then applied debiasing and truncation schemes to estimate the integrated coefficient. However, the TED estimator cannot handle the microstructure noise of high-frequency data, since the noises and regression variables have an unbalanced order relationship. For example, Figure 1 plots the log max, ℓ_1, and ℓ_2 norm errors of the TED, LASSO, and Zero estimators for estimating integrated coefficients with sample sizes $n = 1170, 7800, 23400$, where the dependent and covariate processes are contaminated by microstructure noises. The Zero estimator estimates the coefficients as zero. The detailed simulation setting is described in Section 4 and Appendix A.2, while the rank of the covariate process is set to 0 in this example. As seen in Figure 1, the TED and LASSO estimators cannot estimate the integrated coefficients consistently because the microstructure noise dominates the signal of the coefficients. As the sample size n increases, the effect of noise increases, and the TED estimator shows even worse performance than the Zero estimator in terms of the ℓ_1 norm. Thus, handling microstructure noise is essential when estimating high-dimensional time-varying coefficient processes. On the other hand, we often observe that the financial returns have strong comovements due to the factor structure. When the covariates are highly correlated, the existing methods, such as LASSO, SCAD, and the Dantzig selector, fail to consistently estimate the coefficients (Barigozzi et al., 2024; Fan et al., 2020; Kneip and Sarda, 2011). That is, the direct application of the usual regression procedures cannot guarantee consistency. These findings lead to the demand for developing an estimation method that can simultaneously handle the high dimensionality and time variation in the coefficient process, the microstructure noise of high-frequency data, and the factor structure in the covariate process.

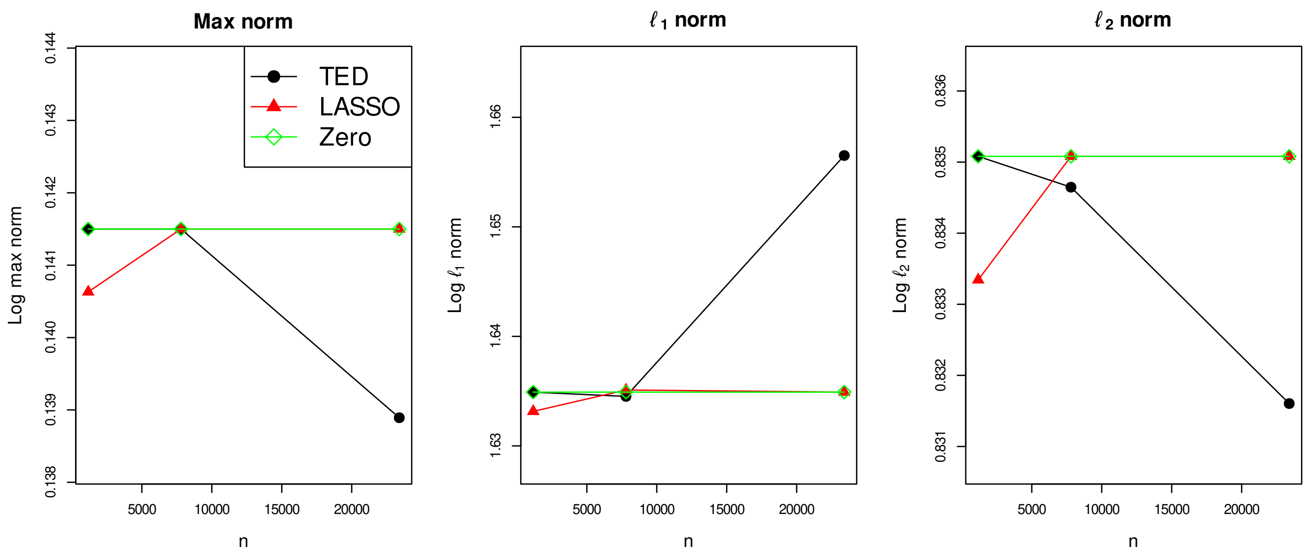

Figure 1: The log max, ℓ_1, and ℓ_2 norm error plots of the TED (black dot), LASSO (red triangle), and Zero (green diamond) estimators for sample sizes $n = 1170, 7800, 23400$.

In this paper, we develop a novel high-dimensional integrated coefficient estimator with factor-based regression jump diffusion processes contaminated by microstructure noises. To handle the high dimensionality and time variation in the coefficient process, we impose a sparse structure on the coefficient process and assume that the coefficient process follows a diffusion process. To accommodate the highly correlated structure of the covariate process, we impose an approximate factor structure (Bai, 2003; Fan et al., 2013), where the factor loading matrix process follows a diffusion process. Due to the time-varying property of the coefficient process, we first estimate the instantaneous coefficients. Specifically, since noises dominate the signals of the regression variables, we smooth the observed dependent and covariate processes using a kernel function. Then, we apply the Principal Component Analysis (PCA) (Aït-Sahalia and Xiu, 2017; Bai, 2003; Dai et al., 2019; Fan et al., 2013, 2020) on the smoothed covariate process to separate the latent factor part from the idiosyncratic part. This procedure transforms the highly dependent covariates into weakly dependent ones. Then, we perform a local regression procedure using the smoothed variables. Due to the noises, the direct application of the LASSO procedure to the smoothed variables cannot guarantee that the derivative of the empirical loss function with the true parameter goes to zero as the sample size goes to infinity, which is called the deviation condition. The deviation condition is essential to obtaining the consistency of the LASSO-type estimator. Thus, we adjust the bias in a loss function using the noise covariance matrix estimator and employ ℓ_1-regularization to accommodate the sparsity of the coefficient process. Due to the bias adjustment, it becomes a nonconvex optimization problem. We demonstrate that the resulting instantaneous coefficient estimator achieves the sharp convergence rate. However, the instantaneous coefficient estimator has another type of bias coming from the ℓ_1-regularization. To handle this bias, we employ a debiasing scheme and estimate the integrated coefficient using debiased instantaneous coefficient estimators. However, the debiasing scheme causes non-sparsity of the integrated coefficient estimates. To accommodate sparsity, the integrated coefficient estimator is further regularized. We call it the Factor Adjusted Thresholded dEbiased Nonconvex LASSO (FATEN-LASSO) estimator. We show that the FATEN-LASSO estimator has a sharp convergence rate. Finally, to implement nonconvex optimization, we adopt the composite gradient descent method (Agarwal et al., 2012) and investigate its properties.

The rest of paper is organized as follows. Section 2 introduces the high-dimensional factor-based regression jump diffusion process. Section 3 proposes the FATEN-LASSO estimator and establishes its concentration properties. In Section 4, we conduct a simulation study to check the finite sample performance of the proposed FATEN-LASSO estimation procedure. In Section 5, we apply the proposed estimation procedure to high-frequency financial data. The conclusion is presented in Section 6, and we provide technical proofs and miscellaneous materials in the Appendix.

2 The model setup

We first fix some notations. For any p_1 by p_2 matrix $\mathbf{G} = (G_{ij})$, define

$$\|\mathbf{G}\|_1 = \max_{1\le j\le p_2} \sum_{i=1}^{p_1} |G_{ij}|, \quad \|\mathbf{G}\|_\infty = \max_{1\le i\le p_1} \sum_{j=1}^{p_2} |G_{ij}|, \quad \text{and} \quad \|\mathbf{G}\|_{\max} = \max_{i,j} |G_{ij}|.$$

We denote the Frobenius norm by $\|\mathbf{G}\|_F = \sqrt{\mathrm{tr}(\mathbf{G}^\top \mathbf{G})}$, and the matrix spectral norm $\|\mathbf{G}\|_2$ is denoted by the square root of the largest eigenvalue of $\mathbf{G}\mathbf{G}^\top$. The vectorization of the matrix $\mathbf{G}$, $\mathrm{vec}(\mathbf{G})$, is the column vector obtained by stacking the columns of $\mathbf{G}$. In addition, $\det(\mathbf{G})$ is the determinant of $\mathbf{G}$. For any vector $\mathbf{x} \in \mathbb{R}^{p_3}$, $\mathrm{Diag}\,(\mathbf{x})$ denotes a p_3 by p_3 diagonal matrix with the elements of $\mathbf{x}$ on the main diagonal. For any process $f(t)$ and $\Delta_n = 1/n$, we define $\Delta_i^n f = f(i\Delta_n) - f((i-1)\Delta_n)$ for $1 \le i \le n$. The sign function is defined as

$$\mathrm{sign}(x) = \begin{cases} -1 & \text{if } x < 0; \\ 0 & \text{if } x = 0; \\ 1 & \text{if } x > 0. \end{cases}$$

We use the subscript 0 to represent the true parameters. We use C's to denote generic positive constants whose values are free of n and p and may vary from appearance to appearance.

Let $Y(t)$ and $\mathbf{X}(t) = (X_1(t), \ldots, X_p(t))^\top$ be the true dependent process and the vector of the true p-dimensional covariate process, respectively. We consider the regression diffusion model as follows:

$$\begin{aligned} &dY(t) = dY^c(t) + dY^J(t), \\ &dY^c(t) = \boldsymbol{\beta}^\top(t) d\mathbf{X}^c(t) + dZ(t), \quad \text{and} \quad dY^J(t) = J^Y(t) d\Lambda^Y(t), \end{aligned} \tag{2.1}$$

where $Y^c(t)$ and $Y^J(t)$ denote the continuous part and jump part of $Y(t)$, respectively, $J^Y(t)$ is a jump size process, $\Lambda^Y(t)$ is a Poisson process with a bounded intensity, $\mathbf{X}^c(t) = \left(X_1^c(t), \ldots, X_p^c(t)\right)^\top$ is the continuous part of $\mathbf{X}(t)$, $\boldsymbol{\beta}(t) = (\beta_1(t), \ldots, \beta_p(t))^\top$ is a coefficient process, and $Z(t)$ is a residual process. The superscripts c and J represent the continuous and jump parts of the process, respectively. The true covariate process $\mathbf{X}(t)$ satisfies the following factor-based jump diffusion model:

$$
\begin{aligned}
& d\mathbf{X}(t) = d\mathbf{X}^c(t) + d\mathbf{X}^J(t), \quad d\mathbf{X}^c(t) = \boldsymbol{\mu}(t)dt + \mathbf{B}(t)d\mathbf{f}(t) + d\mathbf{u}(t), \\
& d\mathbf{f}(t) = \boldsymbol{\nu}_f(t)d\mathbf{W}_f(t), \quad d\mathbf{u}(t) = \boldsymbol{\nu}_u(t)d\mathbf{W}_u(t), \quad \text{and} \quad d\mathbf{X}^J(t) = \mathbf{J}(t)d\boldsymbol{\Lambda}(t),
\end{aligned} \tag{2.2}
$$

where $\mathbf{X}^J(t)$ denotes the jump part of $\mathbf{X}(t)$, $\mathbf{J}(t) = (J_1(t), \ldots, J_p(t))^\top$ is a jump process, and $\boldsymbol{\Lambda}(t) = (\Lambda_1(t), \ldots, \Lambda_p(t))^\top$ denotes a p-dimensional Poisson process with the bounded intensity processes. We note that $\boldsymbol{\Lambda}(t)$ is a vector of p individual Poisson processes and these individual processes are allowed to be dependent. Additionally, $\boldsymbol{\mu}(t)$ is a p-dimensional drift process, $\mathbf{B}(t) = (B_{ij}(t))_{1\le i\le p, 1\le j\le r}$ is a factor loading matrix process, $\mathbf{f}(t) = (f_1(t), \ldots, f_r(t))^\top$ is a latent factor process, $\mathbf{u}(t) = (u_1(t), \ldots, u_p(t))^\top$ is an idiosyncratic process, $\boldsymbol{\nu}_f(t)$ and $\boldsymbol{\nu}_u(t)$ are r by q_1 and p by q_2 instantaneous volatility matrices, respectively, and $\mathbf{W}_f(t)$ and $\mathbf{W}_u(t)$ are q_1-dimensional and q_2-dimensional independent Brownian motions, respectively. The residual process $Z(t)$ satisfies

$$
dZ(t) = \nu_z(t)dW_z(t), \tag{2.3}
$$

where $\nu_z(t)$ is a one-dimensional instantaneous volatility process and $dW_z(t)$ is a one-dimensional independent Brownian motion. The processes $\boldsymbol{\mu}(t)$, $\boldsymbol{\nu}_f(t)$, $\boldsymbol{\nu}_u(t)$, and $\nu_z(t)$ are predictable. The coefficient process $\boldsymbol{\beta}(t)$ and factor loading matrix process $\mathbf{B}(t)$ satisfy the following diffusion models:

$$
d\boldsymbol{\beta}(t) = \boldsymbol{\mu}_\beta(t)dt + \boldsymbol{\nu}_\beta(t)d\mathbf{W}_\beta(t) \quad \text{and} \quad d\mathrm{vec}\left(\mathbf{B}(t)\right) = \boldsymbol{\mu}_B(t)dt + \boldsymbol{\nu}_B(t)d\mathbf{W}_B(t),
$$

where $\boldsymbol{\mu}_\beta(t)$ and $\boldsymbol{\mu}_B(t)$ are p-dimensional and pr-dimensional drift processes, respectively, $\boldsymbol{\nu}_\beta(t)$ and $\boldsymbol{\nu}_B(t)$ are p by q_3 and pr by q_4 instantaneous volatility matrix processes, respectively, $\boldsymbol{\mu}_\beta(t)$, $\boldsymbol{\mu}_B(t)$, $\boldsymbol{\nu}_\beta(t)$, and $\boldsymbol{\nu}_B(t)$ are predictable, and $\mathbf{W}_\beta(t)$ and $\mathbf{W}_B(t)$ are q_3-dimensional and q_4-dimensional independent Brownian motions, respectively. In this paper, the parameter of interest is the following integrated coefficient:

$$I\beta = (I\beta_j)_{j=1,\ldots,p} = \int_0^1 \boldsymbol{\beta}(t)dt.$$

In financial practices, there exist hundreds of potential factor candidates (Bali et al., 2011; Campbell et al., 2008; Cochrane, 2011; Harvey et al., 2016; Hou et al., 2020; McLean and Pontiff, 2016). To accommodate the large set of factor candidates, we assume that the dimension of the covariate process, p, is large, which causes the curse of dimensionality. To handle this issue, we impose the exact sparsity condition for the coefficient process. That is, there exists a set $S_\beta \subset \{1, \ldots, p\}$ with cardinality at most s_p such that $\beta_j(t) = 0$ for $0 \le t \le 1$ and $j \notin S_\beta$.

Unfortunately, we cannot observe the true processes $\mathbf{X}(t)$ and $Y(t)$, since the high-frequency data are contaminated by microstructure noises. These noises result from market inefficiencies, such as the bid–ask spread, the rounding effect, and asymmetric information. To account for this feature, we assume that the observed processes satisfy

$$Y^o(t_i) = Y(t_i) + \epsilon^Y(t_i) \quad \text{and} \quad \mathbf{X}^o(t_i) = \mathbf{X}(t_i) + \boldsymbol{\varepsilon}^X(t_i) \quad \text{for } i = 0, \ldots, n, \tag{2.4}$$

where $t_i \in [0, 1]$ is the ith observation time point, $Y^o(t_i)$ is the observed dependent process for time t_i, $\mathbf{X}^o(t_i) = \left(X_1^o(t_i), \ldots, X_p^o(t_i)\right)^\top$ is the observed covariate process for time t_i, and $\epsilon^Y(t_i)$ and $\boldsymbol{\varepsilon}^X(t_i) = \left(\epsilon_1^X(t_i), \ldots, \epsilon_p^X(t_i)\right)^\top$ are independent one-dimensional and p-dimensional microstructure noises for $Y(t_i)$ and $\mathbf{X}(t_i)$, respectively. The noises are independent over time and have a mean of zero and variances of $\mathrm{E}\{\epsilon^Y(t_i)\}^2 = V^Y$ and $\mathrm{E}\left\{\boldsymbol{\varepsilon}^X(t_i)\left(\boldsymbol{\varepsilon}^X(t_i)\right)^\top\right\} = \mathbf{V}^X$, where $\mathbf{V}^X = \left(V_{jj'}^X\right)_{1 \le j,j' \le p}$. For simplicity, we assume that the observation time points are synchronized and equally spaced:

$t_i - t_{i-1} = 1/n$ for $i = 1, \ldots, n$.

Remark 1. It is more realistic to consider dependent microstructure noises and asynchronous observation time points. In fact, we can relax the conditions for the observation time points by employing the generalized sampling time (Aït-Sahalia et al., 2010), refresh time (Barndorff-Nielsen et al., 2011), and previous tick (Zhang, 2011) schemes. See also Chen et al. (2024). Then, the above condition can be extended to the non-synchronized and unequally spaced condition. On the other hand, various studies, including Chen et al. (2024); Oh et al. (2024), have developed estimation methods that can handle both the time variation of the coefficient process and the dependence structure of microstructure noises. For example, Oh et al. (2024) addresses price-dependent and autocorrelated microstructure noise with a time-varying coefficient process. In our setting, we can handle the dependent microstructure noise by employing the bias adjustment scheme proposed by Oh et al. (2024) when estimating the covariance, $\mathbf{V}^X$, of the microstructure noise. That is, the dependent microstructure noise can be handled by introducing a robust covariance estimator for the dependent structure of the microstructure noise. In this paper, to focus on developing an integrated coefficient estimation method for dependent covariate processes, we assume the synchronized and equally spaced observation time and independent microstructure noise conditions for simplicity.

3 Nonconvex high-dimensional high-frequency regression

3.1 Integrated coefficient estimation procedure

In this section, we propose a high-dimensional integrated coefficient estimation procedure in the presence of microstructure noises, factor structure in covariate processes, and jumps. Recently, Kim et al. (2024) developed an integrated coefficient estimator that can handle the high dimensionality and time variation in the coefficient process without microstructure noises. However, in practice, when employing higher-frequency observations, microstructure noises and strong dependence tend to be observed. To accommodate the noises, we impose the noise structure as in (2.4). To handle the strong dependence, we employ the approximate factor structure as in (2.2). Based on the noisy and strongly correlated high-frequency observation structure, we propose an integrated coefficient estimation procedure as follows. Due to the time variation in the coefficient process, we first need to estimate the instantaneous coefficients. To handle the high dimensionality of instantaneous coefficients, we usually employ a penalized regression method for the observed log-returns, $\Delta_i^n Y^o$ and $\Delta_i^n \mathbf{X}^o$ (Kim et al., 2024; Shin and Kim, 2023). However, in the presence of noises, the noises dominate the signals of true log-returns. This relationship ruins the regression structure in (2.1). To overcome this, we first construct smoothed variables for the observed processes. Specifically, we consider the following:

$$\Delta_i^n \widehat{Y} = \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right) \Delta_{i+l+1}^n Y^o \quad \text{and} \quad \Delta_i^n \widehat{X}_j = \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right) \Delta_{i+l+1}^n X_j^o,$$

where the kernel function $g(x)$ is Lipschitz continuous and satisfies $g(0) = g(1) = 0$ and $\int_0^1 \{g(t)\}^2 dt > 0$, and k_1 is the bandwidth parameter for $g(x)$. We choose $k_1 = c_{k,1} n^{1/2}$ for some constant $c_{k,1}$, which ensures that the signals of the continuous underlying log-return and the noise are of the same magnitude. Thus, this bandwidth choice provides the optimal rate. Then, we employ a local regression with the smoothed variables as follows:

$$\mathcal{Y}_i = \begin{pmatrix} \Delta_i^n \widehat{Y}^{\text{trunc}} \\ \Delta_{i+1}^n \widehat{Y}^{\text{trunc}} \\ \vdots \\ \Delta_{i+k_2-k_1}^n \widehat{Y}^{\text{trunc}} \end{pmatrix} \quad \text{and} \quad \mathcal{X}_i = \begin{pmatrix} \left(\Delta_i^n \widehat{\mathbf{X}}^{\text{trunc}}\right)^\top \\ \left(\Delta_{i+1}^n \widehat{\mathbf{X}}^{\text{trunc}}\right)^\top \\ \vdots \\ \left(\Delta_{i+k_2-k_1}^n \widehat{\mathbf{X}}^{\text{trunc}}\right)^\top \end{pmatrix},$$

where k_2 is the number of observed log-returns used for each local regression,

$$\Delta_i^n \widehat{Y}^{\text{trunc}} = \Delta_i^n \widehat{Y} \, \mathbf{1}\left(|\Delta_i^n \widehat{Y}| \le w_n\right), \quad \Delta_i^n \widehat{\mathbf{X}}^{\text{trunc}} = \left(\Delta_i^n \widehat{X}_j^{\text{trunc}}\right)_{j=1,\ldots,p},$$

$$\Delta_i^n \widehat{X}_j^{\text{trunc}} = \Delta_i^n \widehat{X}_j \mathbf{1}\left(|\Delta_i^n \widehat{X}_j| \leq v_{j,n}\right),$$

$\mathbf{1}(\cdot)$ is an indicator function, and w_n and $v_{j,n}$, $j = 1, \ldots, p$, are the truncation parameters to handle the jumps. We choose $k_2 = c_{k,2} n^{3/4}$ for some constant $c_{k,2}$. In addition, we choose

$$w_n = C_w s_p \sqrt{\log p} n^{-1/4} \quad \text{and} \quad v_{j,n} = C_{j,v} \sqrt{\log p} n^{-1/4} \tag{3.1}$$

for some large constants C_w and $C_{j,v}$, and $j = 1, \ldots, p$.

Remark 2. The truncation parameters w_n and $v_{j,n}$ control the information of the continuous processes by detecting and truncating the jumps after obtaining the smoothed variables. Note that we do not truncate the jumps in each observed log-return. This is because the order of the noise dominates the order of the continuous return, which makes it difficult to control the information of the continuous return when truncating jumps one by one. To detect the jumps, we need the conditions $w_n \geq C s_p \sqrt{\log p} n^{-1/4}$ and $v_{j,n} \geq C \sqrt{\log p} n^{-1/4}$. The $\log p$ term is required to bound the continuous parts of the dependent and covariate processes with high probability. The s_p term in w_n is the cost to handle the continuous parts of s_p significant factors. However, the s_p term is not required when $Y^c(t)$ follows a continuous Itô diffusion model with bounded drift and volatility processes, which is a common assumption for single processes. That is, we theoretically use the s_p term to handle a diverging significant factor summation. However, in practice, we can consider the continuous part of the dependent process as an individual Itô diffusion process and assume that the summation of the factor part is bounded. On the other hand, to obtain the deviation and restricted eigenvalue conditions, we require sharp w_n and $v_{j,n}$. For example, to obtain the restricted eigenvalue condition, we need $\left(\max_j v_{j,n}\right)^2 n^{1/4} \log p \to 0$ as $n, p \to \infty$. Thus, we choose w_n and $v_{j,n}$ as outlined in (3.1).

However, due to the highly correlated structure of the true covariate process $\mathbf{X}(t)$, it is difficult to directly apply commonly used model selection methods, such as LASSO, SCAD, and the Dantzig selector (Barigozzi et al., 2024; Fan et al., 2020, 2024; Kneip and Sarda, 2011). To address this, we employed a decorrelation step based on the factor structure (2.2). Specifically, using PCA (Aït-Sahalia and Xiu, 2017; Dai et al., 2019; Fan et al., 2020, 2013), we first estimate the factor loading matrix $\mathbf{B}(i\Delta_n)$ and smoothed latent factor variable

$$\mathbf{F}_i = \begin{pmatrix} \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right) \Delta_{i+l+1}^n \mathbf{f}^\top \\ \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right) \Delta_{i+l+2}^n \mathbf{f}^\top \\ \vdots \\ \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right) \Delta_{i+l+k_2-k_1+1}^n \mathbf{f}^\top \end{pmatrix}$$

as follows:

$$(\widehat{\mathbf{B}}_{i\Delta_n}, \widehat{\mathbf{F}}_i) = \arg \min_{\mathbf{B}\in\mathbb{R}^{p\times r}, \mathbf{F}\in\mathbb{R}^{(k_2-k_1+1)\times r}} \|\mathcal{X}_i - \mathbf{F}\mathbf{B}^\top\|_F^2, \tag{3.2}$$

subject to

$$p^{-1}\mathbf{B}^\top\mathbf{B} = \mathbf{I}_r \quad \text{and} \quad \mathbf{F}^\top\mathbf{F} \text{ is an } r \times r \text{ diagonal matrix,}$$

where $\mathbf{I}_r$ is the r-dimensional identity matrix. The above constraint is imposed to handle the identification problem for the latent factor and factor loading matrix. The identifiability assumption is described in Assumption 1(c). Then, we estimate the smoothed idiosyncratic variable

$$\mathbf{U}_i = \begin{pmatrix} \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right) \Delta_{i+l+1}^n \mathbf{u}^\top \\ \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right) \Delta_{i+l+2}^n \mathbf{u}^\top \\ \vdots \\ \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right) \Delta_{i+l+k_2-k_1+1}^n \mathbf{u}^\top \end{pmatrix}$$

by

$$\widehat{\mathbf{U}}_i = \mathcal{X}_i - \widehat{\mathbf{F}}_i\widehat{\mathbf{B}}_{i\Delta_n}^\top \tag{3.3}$$

and define

$$\widehat{\mathbf{G}}_i = \left(\widehat{\mathbf{U}}_i, \widehat{\mathbf{F}}_i\right).$$

We note that $\widehat{\mathbf{B}}_{i\Delta_n}$ and $\widehat{\mathbf{F}}_i$ estimate $\mathbf{B}\left(i\Delta_n\right)\mathbf{H}_i$ and $\mathbf{F}_i\left(\mathbf{H}_i^{\top}\right)^{-1}$, respectively, where the nonsingular r by r matrix $\mathbf{H}_i$ comes from handling the identification issue (see Proposition 3 in the Appendix). Let $\boldsymbol{\gamma}(t) = \mathbf{B}^{\top}(t)\boldsymbol{\beta}(t)$. From (2.1) and (2.2), we have

$$dY^c(t) = \boldsymbol{\beta}^{\top}(t)\boldsymbol{\mu}(t)dt + \boldsymbol{\beta}^{\top}(t)d\mathbf{u}(t) + \boldsymbol{\gamma}^{\top}(t)d\mathbf{f}(t) + dZ(t).$$

Thus, based on $\mathcal{Y}_i$ and $\widehat{\mathbf{G}}_i$, we can estimate $\boldsymbol{\beta}(i\Delta_n)$ and $\boldsymbol{\gamma}(i\Delta_n)$ using the weakly correlated structure of $\widehat{\mathbf{G}}_i$. Specifically, the local regression coefficient for $\mathcal{Y}_i$ and $\widehat{\mathbf{G}}_i$ can be approximated by

$$\boldsymbol{\theta}\left(i\Delta_n\right) = \left(\boldsymbol{\beta}^{\top}(i\Delta_n), \boldsymbol{\gamma}^{\top}(i\Delta_n)\mathbf{H}_i\right)^{\top},$$

where $\boldsymbol{\theta}\left(i\Delta_n\right)$ is defined for $i = 0, \ldots, n - k_2$. This implies that we can estimate the instantaneous coefficient by choosing the first p entries after estimating $\boldsymbol{\theta}\left(i\Delta_n\right)$. Now, we estimate $\boldsymbol{\theta}\left(i\Delta_n\right)$ based on $\mathcal{Y}_i$ and $\widehat{\mathbf{G}}_i$ as follows. For each local regression, we need to handle the curse of dimensionality. To do this, we often utilize a penalized regression method, such as LASSO (Tibshirani, 1996) or Dantzig (Candes and Tao, 2007), under the sparsity assumption. However, they cannot consistently estimate instantaneous coefficients due to the bias from microstructure noises in $\widehat{\mathbf{U}}_i$. For example, the usual LASSO leads to the following instantaneous coefficient estimator at time $i\Delta_n$:

$$\widehat{\boldsymbol{\beta}}_{i\Delta_n}^{\text{LASSO}} = \left(\widehat{\theta}_{i\Delta_n,j}^{\text{LASSO}}\right)_{j=1,\ldots,p},$$

where

$$\widehat{\boldsymbol{\theta}}_{i\Delta_n}^{\text{LASSO}} = \left(\widehat{\theta}_{i\Delta_n,j}^{\text{LASSO}}\right)_{j=1,\ldots,p+r} = \underset{\boldsymbol{\theta}\in\mathbb{R}^{p+r}}{\arg\min}\, \frac{n}{k_1 k_2}\left\|\mathcal{Y}_i - \widehat{\mathbf{G}}_i\boldsymbol{\theta}\right\|_2^2 + \eta\left\|\boldsymbol{\theta}\right\|_1$$

and $\eta > 0$ is some regularization parameter. To obtain the consistency of $\widehat{\boldsymbol{\beta}}_{i\Delta_n}^{\text{LASSO}}$, we need the deviation condition $\frac{2n}{k_1k_2}\left\|\widehat{\mathbf{G}}_i^\top\widehat{\mathbf{G}}_i\boldsymbol{\theta}_{0,i\Delta_n}-\widehat{\mathbf{G}}_i^\top\mathcal{Y}_i\right\|_{\max}\xrightarrow{p}0$. However, in the presence of noises, this condition cannot be satisfied, since $\mathrm{E}\left(\widehat{\mathbf{U}}_i^\top\widehat{\mathbf{U}}_i\right)$ contains noise covariance terms for the covariate process $\mathbf{X}(t)$. Thus, we need to estimate the noise covariance matrix and adjust the bias. The noise covariance matrix for the covariate process is estimated by

$$\widehat{\mathbf{V}}^X=\frac{1}{2n}\sum_{i=1}^{n}\Delta_i^n\mathbf{X}^{\mathrm{trunc}}\left(\Delta_i^n\mathbf{X}^{\mathrm{trunc}}\right)^\top, \tag{3.4}$$

where

$$\Delta_i^n\mathbf{X}^{\mathrm{trunc}}=\left(\Delta_i^nX_j^o\,\mathbf{1}\left(|\Delta_i^nX_j^o|\le v_{j,n}^{(2)}\right)\right)_{j=1,\dots,p}$$

and $v_{j,n}^{(2)}$, $j=1,\dots,p$, are the truncation parameters to handle the jumps. We utilize

$$v_{j,n}^{(2)}=C_{j,v}^{(2)}\sqrt{\log p} \tag{3.5}$$

for some large constants $C_{j,v}^{(2)}$, $j=1,\dots,p$.

Remark 3. As in (3.1), the truncation parameter $v_{j,n}^{(2)}$ controls the information for the noise covariances by detecting and truncating the jumps. To detect the jumps, the condition $v_{j,n}^{(2)}\ge C\sqrt{\log p}$ is required, which is different from the case of (3.1), since we only need to estimate the noise covariance matrix using the observed log-returns. We note that the $\log p$ term is required to bound the noise part with high probability. Conversely, we need $\left(\max_j v_{j,n}^{(2)}\right)^2\log p/n\to 0$ as $n,p\to\infty$ to satisfy the restricted eigenvalue condition. Thus, we choose sharp $v_{j,n}^{(2)}$, as in (3.5). It is worth noting that when the jump size is finite, truncation for the observed log-returns is not required. However, since we do not impose any assumption on the jump size process, the proposed truncation method is used to handle the heavy-tailedness of the jump sizes.

Then, the instantaneous coefficient estimator at time $i\Delta_n$ is defined as follows:

$$\widehat{\boldsymbol{\beta}}_{i\Delta_n} = \left(\widehat{\theta}_{i\Delta_n,j}\right)_{j=1,\dots,p},$$

where

$$\widehat{\boldsymbol{\theta}}_{i\Delta_n} = \left(\widehat{\theta}_{i\Delta_n,j}\right)_{j=1,\dots,p+r} = \underset{\|\boldsymbol{\theta}\|_1\le\rho}{\arg\min}\, \mathcal{L}_i(\boldsymbol{\theta}) + \eta\left\|\boldsymbol{\theta}\right\|_1, \tag{3.6}$$

$$\mathcal{L}_i(\boldsymbol{\theta}) = \frac{n}{2\phi k_1 k_2}\|\mathcal{Y}_i - \widehat{\mathbf{G}}_i\boldsymbol{\theta}\|_2^2 - \frac{n\zeta}{2\phi k_1}\boldsymbol{\theta}^\top\widehat{\mathbf{V}}\boldsymbol{\theta}, \quad \widehat{\mathbf{V}} = \left(\begin{array}{c|c} \widehat{\mathbf{V}}^X & \mathbf{0}_{p\times r} \\ \hline \mathbf{0}_{r\times p} & \mathbf{0}_{r\times r}\end{array}\right), \tag{3.7}$$

ρ satisfies $\rho \ge \|\boldsymbol{\theta}_0(i\Delta_n)\|_1$, $\eta > 0$ is the regularization parameter, $\mathcal{L}_i(\boldsymbol{\theta})$ is the empirical loss function, $\phi = \frac{1}{k_1}\sum_{\ell=0}^{k_1-1}\left\{g\left(\frac{\ell}{k_1}\right)\right\}^2$, and $\zeta = \sum_{l=0}^{k_1-1}\left\{g\left(\frac{l}{k_1}\right) - g\left(\frac{l+1}{k_1}\right)\right\}^2 = O\left(\frac{1}{k_1}\right)$. The tuning parameters ρ and η will be specified in Theorem 1. For the empirical loss function $\mathcal{L}_i(\boldsymbol{\theta})$ in (3.7), the deviation condition $\|\nabla\mathcal{L}_i(\boldsymbol{\theta}_0(i\Delta_n))\|_\infty = \frac{n}{\phi k_1 k_2}\left\|\widehat{\mathbf{G}}_i^\top\widehat{\mathbf{G}}_i\boldsymbol{\theta}_0(i\Delta_n) - \widehat{\mathbf{G}}_i^\top\mathcal{Y}_i - k_2\zeta\widehat{\mathbf{V}}\boldsymbol{\theta}_0(i\Delta_n)\right\|_{\max} \xrightarrow{p} 0$ is satisfied since the noise part in $\widehat{\mathbf{U}}_i^\top\widehat{\mathbf{U}}_i$ is adjusted by the noise covariance estimator $\widehat{\mathbf{V}}^X$ (see Proposition 4 in the Appendix). Additionally, the Hessian matrix of the empirical loss function has the following structure:

$$\nabla^2\mathcal{L}_i(\boldsymbol{\theta}) = \left(\begin{array}{c|c} \frac{n}{\phi k_1 k_2}\widehat{\mathbf{U}}_i^\top\widehat{\mathbf{U}}_i - \frac{n\zeta}{\phi k_1}\widehat{\mathbf{V}}^X & \frac{n}{\phi k_1 k_2}\widehat{\mathbf{U}}_i^\top\widehat{\mathbf{F}}_i \\ \hline \frac{n}{\phi k_1 k_2}\widehat{\mathbf{F}}_i^\top\widehat{\mathbf{U}}_i & \frac{n}{\phi k_1 k_2}\widehat{\mathbf{F}}_i^\top\widehat{\mathbf{F}}_i\end{array}\right).$$

We note that the upper left term has the same form as the pre-averaging realized volatility (PRV) (Christensen et al., 2010; Jacod et al., 2009), which can be one of the estimators for $\boldsymbol{\Sigma}_u(t) = \boldsymbol{\nu}_u(t)\boldsymbol{\nu}_u^\top(t)$. However, we cannot guarantee that $\nabla^2\mathcal{L}_i(\boldsymbol{\theta})$ is positive semidefinite due to the bias adjustment, which implies that the objective function $\mathcal{L}_i(\boldsymbol{\theta}) + \eta\left\|\boldsymbol{\theta}\right\|_1$ can be unbounded from below. To handle the unbounded problem, we impose a constraint on $\boldsymbol{\theta}$, such as $\|\boldsymbol{\theta}\|_1 \le \rho$, for the nonconvex optimization problem (3.6). Theorem 1 shows that the instantaneous coefficient estimator $\widehat{\boldsymbol{\beta}}_{i\Delta_n}$ is consistent when we choose appropriate values for ρ and η. To estimate the integrated coefficient, we can consider the integration of $\widehat{\boldsymbol{\beta}}_{i\Delta_n}$. However, $\widehat{\boldsymbol{\beta}}_{i\Delta_n}$ is biased due to the regularization, so their integration fails to enjoy the law of large number property. In other words, while the integration has the same convergence rate as $\widehat{\boldsymbol{\beta}}_{i\Delta_n}$, the convergence rate is not fast enough. To obtain a faster convergence rate, we apply the debiasing scheme to each value of $\widehat{\boldsymbol{\beta}}_{i\Delta_n}$ as follows. We first estimate the inverse instantaneous idiosyncratic volatility matrix $\boldsymbol{\Omega}(i\Delta_n) = \boldsymbol{\Sigma}_u^{-1}(i\Delta_n)$ based on the following constrained ℓ_1-minimization for inverse matrix estimation (CLIME) (Cai et al., 2011):

$$\widehat{\boldsymbol{\Omega}}_{i\Delta_n} = \arg\min \|\boldsymbol{\Omega}\|_1 \quad \text{s.t.} \quad \left\| \left(\frac{n}{\phi k_1 k_2} \widehat{\mathbf{U}}_i^\top \widehat{\mathbf{U}}_i - \frac{n\zeta}{\phi k_1} \widehat{\mathbf{V}}^X \right) \boldsymbol{\Omega} - \mathbf{I} \right\|_{\max} \leq \tau, \tag{3.8}$$

where τ is the tuning parameter that will be specified in Theorem 2. Using the inverse instantaneous volatility matrix estimator $\widehat{\boldsymbol{\Omega}}_{i\Delta_n}$, we adjust the instantaneous coefficient estimator $\widehat{\boldsymbol{\beta}}_{i\Delta_n}$ as follows:

$$\widetilde{\boldsymbol{\beta}}_{i\Delta_n} = \widehat{\boldsymbol{\beta}}_{i\Delta_n} + \frac{n}{\phi k_1 k_2} \widehat{\boldsymbol{\Omega}}_{i\Delta_n}^\top \left\{ \widehat{\mathbf{U}}_i^\top \mathcal{Y}_i - \left(\widehat{\mathbf{U}}_i^\top \mathcal{X}_i - k_2 \zeta \widehat{\mathbf{V}}^X \right) \widehat{\boldsymbol{\beta}}_{i\Delta_n} \right\}. \tag{3.9}$$

Note that in (3.9), we used $\frac{n}{\phi k_1 k_2} \widehat{\mathbf{U}}_i^\top \mathcal{X}_i - \frac{n\zeta}{\phi k_1} \widehat{\mathbf{V}}^X$ as the proxy for the instantaneous idiosyncratic volatility matrix at time $i\Delta_n$, instead of $\frac{n}{\phi k_1 k_2} \widehat{\mathbf{U}}_i^\top \widehat{\mathbf{U}}_i - \frac{n\zeta}{\phi k_1} \widehat{\mathbf{V}}^X$. This is because $\widehat{\mathbf{U}}_i^\top \left(\mathcal{Y}_i - \mathcal{X}_i \widehat{\boldsymbol{\beta}}_{i\Delta_n} \right)$ has fewer error terms than $\widehat{\mathbf{U}}_i^\top \left(\mathcal{Y}_i - \widehat{\mathbf{U}}_i \widehat{\boldsymbol{\beta}}_{i\Delta_n} \right)$. Then, we estimate the integrated coefficient as follows:

$$\widehat{I\beta} = \sum_{i=0}^{[1/(k_2\Delta_n)]-1} \widetilde{\boldsymbol{\beta}}_{ik_2\Delta_n} k_2 \Delta_n. \tag{3.10}$$

The debiased integrated coefficient estimator $\widehat{I\beta}$ benefits from the law of large number property and has a faster convergence rate than the integration of the instantaneous coefficient estimators. However, the bias adjustment leads to the non-sparse structure of the integrated coefficient estimator. To accommodate the sparse structure of the integrated coefficient, we employ the thresholding scheme as follows:

$$\widetilde{I\beta}_j = s(\widehat{I\beta}_j)\mathbf{1}\left(|\widehat{I\beta}_j| \geq h_n\right) \quad \text{and} \quad \widetilde{I\beta} = \left(\widetilde{I\beta}_j\right)_{j=1,\ldots,p}, \tag{3.11}$$

where $s(x)$ is the thresholding function satisfying $|s(x) - x| \leq h_n$ and h_n is a thresholding level that will be specified in Theorem 3. For the thresholding function $s(x)$, we usually employ the soft thresholding function $s(x) = x - \text{sign}(x)h_n$ or hard thresholding function $s(x) = x$. We utilize the hard thresholding function in the empirical study. We call this the Factor Adjusted Thresholded dEbiased Nonconvex LASSO (FATEN-LASSO) estimator. A summary of the FATEN-LASSO estimation procedure is presented in Appendix A.11. We will discuss the choice of the tuning parameters in Appendix A.1.

Remark 4. In this paper, we allow the time variation of processes, such as the dependent, covariate, coefficient, and factor loading matrix. The rank r is assumed to be constant over time. To allow the time variation of the rank r, the state heterogeneous structure (Chun and Kim, 2022) can be considered for the covariate process. That is, the rank r is the same within the same state, but may change when the state changes. However, extending this approach from the one-dimensional case, as in Chun and Kim (2022), to the high-dimensional case is not straightforward. Thus, we leave this issue for a future study.

3.2 Theoretical results

In this section, we show the asymptotic properties of the proposed FATEN-LASSO estimator. To investigate its asymptotic behaviors, the following assumptions are required.

Assumption 1.

(a) $\boldsymbol{\mu}(t)$, $\boldsymbol{\Sigma}_f(t) = \boldsymbol{\nu}_f(t)\boldsymbol{\nu}_f^\top(t)$, $\boldsymbol{\Sigma}_u(t)$, $\nu_z(t)$, $\boldsymbol{\beta}(t)$, $\boldsymbol{\mu}_\beta(t)$, $\boldsymbol{\Sigma}_\beta(t) = \boldsymbol{\nu}_\beta(t)\boldsymbol{\nu}_\beta^\top(t)$, $\mathbf{B}(t)$, *and* $\boldsymbol{\Sigma}_B(t) = \boldsymbol{\nu}_B(t)\boldsymbol{\nu}_B^\top(t)$ *are almost surely entry-wise bounded, and* $\|\boldsymbol{\Sigma}_u^{-1}(t)\|_1 \leq C$ *a.s.*

(b) The noises $\epsilon_j^X(t_i)$, $j = 1, \ldots, p$, and $\epsilon^Y(t_i)$ are sub-Gaussian with a bounded parameter.

(c) For $0 \le i \le [1/(k_2\Delta_n)] - 1$, $\mathbf{\Sigma}_f(ik_2\Delta_n)$ has distinct eigenvalues bounded away from 0 and $\|p^{-1}\mathbf{B}^\top(ik_2\Delta_n)\mathbf{B}(ik_2\Delta_n) - \mathbf{I}_r\|_2 \to 0$ as $p \to \infty$.

(d) $\sup_{0\le t\le 1} \|\mathbf{\Sigma}_u(t)\|_1 \le C$ and $\left\|\mathbf{V}^X\right\|_1 \le C$.

(e) For $\mathbf{b} = (1, \ldots, 1)^\top \in \mathbb{R}^p$, the random variable $\left(\mathbf{b}^\top \int_{i\Delta_n}^{(i+k_2)\Delta_n} d\mathbf{\Lambda}(t)dt\right)$ is sub-exponential with a bounded parameter.

(f) The volatility matrix processes, $\mathbf{\Sigma}_f(t) = (\Sigma_{f,ij}(t))_{i,j=1,\ldots,r}$ and $\mathbf{\Sigma}_u(t) = (\Sigma_{u,ij}(t))_{i,j=1,\ldots,p}$, satisfy the following Hölder condition:

$$\max_{1\le i,j\le r} |\Sigma_{f,ij}(t) - \Sigma_{f,ij}(s)| \le C\sqrt{|t-s|\log p} \ \ a.s.,$$
$$\max_{1\le i,j\le p} |\Sigma_{u,ij}(t) - \Sigma_{u,ij}(s)| \le C\sqrt{|t-s|\log p} \ \ a.s.$$

(g) $Cn^{c_1} \le p \le \exp(n^{c_2})$ for some constants $c_1, c_2 > 0$, and $s_p\left(n^{-1/8} + p^{-1/2}\right)(\log p)^{5/2} \to 0$ as $n, p \to \infty$.

(h) The inverse idiosyncratic volatility matrix process $\mathbf{\Sigma}_u^{-1}(t) = \mathbf{\Omega}(t) = (\omega_{ij}(t))_{i,j=1,\ldots,p}$ satisfies the following sparsity condition:

$$\sup_{0\le t\le 1} \max_{1\le i\le p} \sum_{j=1}^{p} |\omega_{ij}(t)|^q \le s_{\omega,p} \ \ a.s.,$$

where $q \in [0, 1)$ and $s_{\omega,p} \le (\log p)^{c_\omega}$ for a constant $c_\omega > 0$.

(i) For large n, $|I\beta_j| \ge 2h_n$ a.s. for all $j \in S_{I\beta}$, where h_n is defined in Theorem 3 and $S_{I\beta}$ is the support of $I\beta$.

Remark 5. Assumption 1(a) is the boundedness condition that implies the sub-Gaussian tails for the continuous processes, such as the latent factor process $\mathbf{f}(t)$, the idiosyncratic process $\mathbf{u}(t)$, and the coefficient process $\boldsymbol{\beta}(t)$. Sub-Gaussian assumptions are frequently used in high-dimensional statistics to obtain probability bounds. Similarly, in Assumption 1(b), we impose sub-Gaussianity for the noises. Assumption 1(c) is the identifiability condition, which is often used to estimate the latent factor and factor loading matrix (Aït-Sahalia and Xiu, 2017; Dai et al., 2019). Assumption 1(d) is the sparsity condition on the idiosyncratic volatility matrix and noise covariance matrix, which is required to harness the approximate factor model and to obtain the restricted eigenvalue condition for the LASSO-type estimator. To obtain the restricted eigenvalue condition, we also need the technical condition Assumption 1(e). Assumption 1(f) is the continuity condition, which is required to investigate the asymptotic properties of the estimators for the time-varying processes. This condition can be obtained with high probability when the volatility processes $\boldsymbol{\Sigma}_f(t)$ and $\boldsymbol{\Sigma}_u(t)$ follow continuous Itô diffusion models with bounded drift and volatility processes. In Assumption 1(h), we impose the sparse structure on the inverse idiosyncratic volatility matrix process to investigate the asymptotic behaviors of the CLIME estimator. Finally, Assumption 1(i) is imposed to investigate the sign consistency of the FATEN-LASSO estimator. We note that h_n converges to 0 as n goes to infinity.

The following theorem establishes the asymptotic behaviors of the instantaneous coefficient estimator $\widehat{\boldsymbol{\beta}}_{i\Delta_n}$.

Theorem 1. *Under the models* (2.1)–(2.4) *and Assumption 1(a)–(g), let* $k_1 = c_{k_1} n^{1/2}$ *and* $k_2 = c_{k_2} n^{3/4}$ *for some constants* c_{k_1} *and* c_{k_2}*. For any given positive constant* a*, choose* $\rho = C_{\rho,a} s_p$ *and* $\eta = C_{\eta,a}\left\{n^{-1/8} s_p (\log p)^2 + p^{-1/2} s_p \sqrt{\log p}\right\}$ *for some large constants* $C_{\rho,a}$ *and* $C_{\eta,a}$*. Then, for large* n*, we have*

$$\max_i \|\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0(i\Delta_n)\|_1 \le C s_p \eta \quad \text{and} \quad \max_i \|\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0(i\Delta_n)\|_2 \le C\sqrt{s_p}\eta, \tag{3.12}$$

with the probability at least $1 - p^{-a}$.

Remark 6. Theorem 1 shows that $\widehat{\boldsymbol{\beta}}_{i\Delta_n}$ has the convergence rate of $n^{-1/8}+p^{-1/2}$ with the sparsity level and log order terms. For each local regression, the number of observed log-returns is $Cn^{3/4}$, whereas the number of non-overlapping smoothed variables is $Cn^{1/4}$. Due to the cost of managing the noises, we are able to use only $Cn^{1/4}$ variables to estimate the instantaneous coefficient. Thus, the optimal convergence rate is expected to be $n^{-1/8}$, and the first term, $n^{-1/8}s_p^2\left(\log p\right)^2$, is the near-optimal convergence rate. The second term, $p^{-1/2}s_p^2\sqrt{\log p}$, comes from estimating the latent factor process, and $p^{-1/2}$ is the same rate as in Fan et al. (2020). When $n = O(p^4)$, the proposed instantaneous coefficient estimator can achieve the near-optimal convergence rate.

As discussed in Section 3.1, the instantaneous coefficient estimators are biased due to the regularization. Thus, the integration of the instantaneous coefficient estimators cannot benefit from the law of large number property. In other words, while the integration converges, the convergence rate is not fast enough. To handle this issue, we utilize the debiasing scheme and obtain the debiased integrated coefficient estimator, as outlined in (3.9) and (3.10). We establish the asymptotic property of the debiased integrated coefficient estimator in the following theorem.

Theorem 2. *Under the assumptions in Theorem 1 and Assumption 1(h), for any given positive constant* a*, choose* $\tau = C_{\tau,a}\left\{n^{-1/8}\left(\log p\right)^2 + p^{-1/2}\sqrt{\log p}\right\}$ *for some constant* $C_{\tau,a}$*. Then, we have, with the probability at least* $1-p^{-a}$*,*

$$\|\widehat{I\beta} - I\beta_0\|_{\max} \leq Cb_n, \tag{3.13}$$

where $b_n = s_p^2 n^{-1/4}\left(\log p\right)^4 + s_p s_{\omega,p} n^{(-2+q)/8}\left(\log p\right)^{4-2q} + p^{-1}s_p^2\log p + p^{(-2+q)/2}s_p s_{\omega,p}(\log p)^{(2-q)/2} + p^{-1/2}s_p s_{\omega,p} n^{(-1+q)/8}(\log p)^{(5-4q)/2}$.

Remark 7. Theorem 2 indicates that the debiased integrated coefficient estimator is consistent in terms of the max norm. When the inverse volatility matrix process satisfies the exact sparsity condition, that is, $q = 0$, the debiased integrated coefficient estimator has the convergence rate of $n^{-1/4}\left(s_p + s_{\omega,p}\right) s_p \left(\log p\right)^4 + p^{-1}\left(s_p + s_{\omega,p}\right) s_p \log p$. In contrast, we have the convergence rate of $n^{-1/8} s_p^2 \left(\log p\right)^2 + p^{-1/2} s_p^2 \sqrt{\log p}$ without a debiasing scheme. In high-dimensional statistics, the sparsity level is assumed to diverge relatively slowly, such as $\log p$. Thus, the debiased integrated coefficient estimator has the faster convergence rate than the integration of the instantaneous coefficient estimators.

Theorem 2 shows that the input-integrated coefficient estimator $\widehat{I\beta}$ performs well because of the debiasing scheme. Finally, to accommodate the sparse structure, we utilize the thresholding scheme and obtain the FATEN-LASSO estimator. The following theorem establishes the ℓ_1 convergence rate and sign consistency of the FATEN-LASSO estimator.

Theorem 3. *Under the assumptions in Theorem 2 and Assumption 1(i), for any given positive constant a, choose $h_n = C_{h,a} b_n$ for some constant $C_{h,a}$, where b_n is defined in Theorem 2. Then, we have, with probability at least $1 - p^{-a}$,*

$$\|\widetilde{I\beta} - I\beta_0\|_1 \le C s_p b_n \quad \textit{and} \tag{3.14}$$

$$\textit{sign}\left(\widetilde{I\beta}_j\right) = \textit{sign}\left(I\beta_{0,j}\right) \quad \textit{for all } j = 1, \ldots, p. \tag{3.15}$$

Theorem 3 shows the ℓ_1 norm error bound and sign consistency of the proposed FATEN-LASSO estimator. When the exact sparsity condition is satisfied, that is, $q = 0$, the FATEN-LASSO estimator has the convergence rate of $n^{-1/4}\left(s_p + s_{\omega,p}\right) s_p^2 \left(\log p\right)^4 + p^{-1}\left(s_p + s_{\omega,p}\right) s_p^2 \log p$. We note that in the presence of microstructure noises, $n^{-1/4}$ is the optimal convergence rate of the integrated coefficient estimator in the finite-dimensional setup. Thus, when $n = O(p^4)$, the FATEN-LASSO estimator achieves the optimal convergence rate with up to $\log p$ and sparsity level orders.

Remark 8. When the true rank r is 0, the FATEN-LASSO estimator has the optimal convergence rate with up to the orders of $\log p$, s_p, and $s_{\omega,p}$. We note that since we do not need to estimate the latent factor for $r = 0$, we do not have the p^{-1} term.

3.3 Implementation of the FATEN-LASSO estimation procedure

To implement the FATEN-LASSO estimation procedure, we need to solve the nonconvex optimization problem (3.6). However, it is generally hard to obtain the global minimizer of a nonconvex optimization problem in a polynomial time. To handle this issue, we employ the composite gradient descent method (Agarwal et al., 2012) as follows:

$$\widehat{\boldsymbol{\beta}}_{i\Delta_n}^{t+1} = \left(\widehat{\theta}_{i\Delta_n,j}^{t+1}\right)_{j=1,\dots,p},$$

where

$$\begin{aligned}
\widehat{\boldsymbol{\theta}}_{i\Delta_n}^{t+1} &= \left(\widehat{\theta}_{i\Delta_n,j}^{t+1}\right)_{j=1,\dots,p+r} \\
&= \underset{\|\boldsymbol{\theta}\|_1 \le \rho}{\arg\min} \left\{ \mathcal{L}_i\left(\widehat{\boldsymbol{\theta}}_{i\Delta_n}^{t}\right) + \langle \nabla \mathcal{L}_i(\widehat{\boldsymbol{\theta}}_{i\Delta_n}^{t}), \boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{i\Delta_n}^{t} \rangle + \alpha_2 \left\| \boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{i\Delta_n}^{t} \right\|_2^2 + \eta \left\|\boldsymbol{\theta}\right\|_1 \right\}, \quad (3.16)
\end{aligned}$$

$\widehat{\boldsymbol{\theta}}_{i\Delta_n}^{0}$ is the initial parameter, and $\alpha_2 > 0$ is defined in Proposition 5 in the Appendix. Then, we can obtain the following proposition.

Proposition 1. *Under the assumptions in Theorem 1, we have, with the probability at least* $1-p^{-a}$*,*

$$\max_i \|\widehat{\boldsymbol{\beta}}_{i\Delta_n}^{t} - \widehat{\boldsymbol{\beta}}_{i\Delta_n}\|_2 \le C\sqrt{s_p}\eta \quad (3.17)$$

for all $t \ge C\left\{\log\left(\frac{\phi_i(\widehat{\boldsymbol{\theta}}_{i\Delta_n}^{0}) - \phi_i(\widehat{\boldsymbol{\theta}}_{i\Delta_n})}{\|\widehat{\boldsymbol{\theta}}_{i\Delta_n} - \boldsymbol{\theta}_0(i\Delta_n)\|_2^2}\right) + \log_2\log_2\left(\frac{C s_p \eta}{\|\widehat{\boldsymbol{\theta}}_{i\Delta_n} - \boldsymbol{\theta}_0(i\Delta_n)\|_2^2}\right)\right\}$*, where* $\phi_i\left(\boldsymbol{\theta}\right) = \mathcal{L}_i(\boldsymbol{\theta}) + \eta \left\|\boldsymbol{\theta}\right\|_1$.

Proposition 1 shows that the ℓ_2 distance between the local minimizer $\widehat{\boldsymbol{\beta}}_{i\Delta_n}^{t}$ and global minimizer $\widehat{\boldsymbol{\beta}}_{i\Delta_n}$ has the same convergence rate as the statistical error of the global minimizer $\widehat{\boldsymbol{\beta}}_{i\Delta_n}$. That is, the local and global minimizers have the same convergence rate in terms of the ℓ_2 norm. Furthermore, the local minimizer $\widehat{\boldsymbol{\beta}}_{i\Delta_n}^{t}$ can be obtained in a polynomial time. Thus, the proposed FATEN-LASSO procedure is computationally feasible with theoretical guarantees.

4 A simulation study

In this section, we conducted a simulation study to check the finite sample performance of the FATEN-LASSO estimator. The data were generated with a frequency of $1/n^{all}$ based on the factor-based regression jump diffusion models in (2.1)–(2.3). The detailed simulation setup is presented in Appendix A.2. Noise-contaminated high-frequency observations were generated as follows:

$$Y^o(t_i) = Y(t_i) + \epsilon^Y(t_i) \quad \text{and} \quad \mathbf{X}^o(t_i) = \mathbf{X}(t_i) + \boldsymbol{\varepsilon}^X(t_i) \quad \text{for } i = 0, \dots, n,$$

where $\epsilon^Y(t_i)$ and $\epsilon_j^X(t_i)$ were obtained from an independent normal distribution with a mean of zero and a standard deviation of $0.05\sqrt{\int_0^1 \left[\boldsymbol{\beta}^\top(t)\mathbf{B}(t)\boldsymbol{\Sigma}_f(t)\mathbf{B}^\top(t)\boldsymbol{\beta}(t) + \boldsymbol{\beta}^\top(t)\boldsymbol{\Sigma}_u(t)\boldsymbol{\beta}(t) + \nu_z^2(t)\right] dt}$ and $0.05\sqrt{\left(\int_0^1 \left[\mathbf{B}(t)\boldsymbol{\Sigma}_f(t)\mathbf{B}^\top(t) + \boldsymbol{\Sigma}_u(t)\right] dt\right)_{jj}}$, respectively. We set $p = 200$, $s_p = [\log p]$, $r = 3$, $n^{all} = 23400$, and we varied n from 1170 to 23400. To obtain the FATEN-LASSO estimator, we employed the hard thresholding function $s(x) = x$ and implemented the tuning parameter choice procedure discussed in Appendix A.1. To determine the rank r, we first obtained the FATEN-LASSO estimator $\widetilde{I\beta}^r$ for $r = 0, 1, 2, 3, 4, 5$. Then, we chose r by minimizing the mean squared error (MSE) given by

$$\frac{1}{n - k_1 + 1} \sum_{i=0}^{n-k_1} \left(\Delta_i^n \widehat{Y}^{\text{trunc}} - \widetilde{I\beta}^{r\top} \Delta_i^n \widehat{\mathbf{X}}^{\text{trunc}}\right)^2. \tag{4.1}$$

For the comparison, we employed the Factor Adjusted Thresholded dEbiased Convex LASSO (FATEC-LASSO) estimator. It uses the same estimation procedure as the FATEN-LASSO estimator except for the bias adjustment for the noise covariance terms. Specifically, $\widehat{\boldsymbol{\theta}}_{i\Delta_n}^{\text{FATEC}}$ is calculated using the usual LASSO procedure as follows:

$$\widehat{\boldsymbol{\theta}}_{i\Delta_n}^{\text{FATEC}} = \arg\min \frac{n}{2\phi k_1 k_2} \left\| \mathcal{Y}_i - \widehat{\mathbf{G}}_i \boldsymbol{\theta} \right\|_2^2 + \eta^{\text{FATEC}} \left\| \boldsymbol{\theta} \right\|_1, \tag{4.2}$$

where the regularization parameter $\eta^{\text{FATEC}} = c_\eta^{\text{FATEC}} n^{-1/8} (\log p)^2$ and $c_\eta^{\text{FATEC}} \in [10^{-4}, 10^4]$ is chosen by minimizing the corresponding Bayesian information criterion (BIC). We also chose r for the FATEC-LASSO estimator using the same procedure as the FATEN-LASSO estimator. We note that the FATEC-LASSO estimator can partially handle the noises by using the smoothed variables. Furthermore, it can handle the strongly correlated structure of the covariate process by applying PCA. However, it cannot satisfy the deviation condition due to the bias from the noises, which leads to the non-consistency of the instantaneous coefficient estimator. We also considered the TED estimator (Kim et al., 2024), which can handle the time variation in the coefficient process and the curse of dimensionality. Specifically, with the observed log-returns, we first utilized the Dantzig selector (Candes and Tao, 2007) to obtain the instantaneous coefficient estimator. Then, we employed debiasing and truncation schemes to obtain the integrated coefficient estimator. The detailed estimation procedure is presented in Algorithm 1 in Kim et al. (2024). Since the TED estimator directly uses the observed log-returns, it cannot account for the noises and the factor structure. Finally, we employed the LASSO estimator (Tibshirani, 1996) as follows:

$$\widetilde{I\beta}^{\text{LASSO}} = \text{argmin}_{\boldsymbol{\beta}} \left\{ \sum_{i=0}^{n-1} \left(\Delta_{i+1}^n Y^{\text{trunc2}} - \left(\Delta_{i+1}^n \mathbf{X}^{\text{trunc2}} \right)^\top \boldsymbol{\beta} \right)^2 + \eta^{\text{LASSO}} \|\boldsymbol{\beta}\|_1 \right\}, \tag{4.3}$$

where $\Delta_i^n Y^{\text{trunc2}} = \Delta_i^n Y^o \mathbf{1}\left(|\Delta_i^n Y^o| \le w_n^{(3)} \right)$, $\Delta_i^n \mathbf{X}^{\text{trunc2}} = \left(\Delta_i^n X_j^o \mathbf{1}\left(|\Delta_i^n X_j^o| \le v_{j,n}^{(3)} \right) \right)_{j=1,\dots,p}$, and the regularization parameter $\eta^{\text{LASSO}} \in [10^{-4}, 10^4]$ was selected by minimizing the corresponding BIC. We choose

$$w_n^{(3)} = 3n^{-0.47}\sqrt{BV^Y} \quad \text{and} \quad v_{j,n}^{(3)} = 3n^{-0.47}\sqrt{BV_j},$$

where the bipower variations $BV^Y = \frac{\pi}{2}\sum_{i=2}^n |\Delta_{i-1}^n Y^o| \cdot |\Delta_i^n Y^o|$ and $BV_j = \frac{\pi}{2}\sum_{i=2}^n |\Delta_{i-1}^n X_j^o| \cdot |\Delta_i^n X_j^o|$. This choice of truncation parameters is often used in the literature (Aït-Sahalia et al., 2020; Aït-Sahalia and Xiu, 2019). We note that the LASSO estimator can handle the high dimensionality; however, it cannot account for the noises, factor structure, and time variation in the coefficient process. The average estimation errors under the max norm, ℓ_1 norm, and ℓ_2 norm were calculated through 500 iterations.

We first investigated the performance of the estimators for detecting non-zero integrated coefficients. Table 1 reports the average rates of false positives and false negatives of the FATEN-LASSO, FATEC-LASSO, TED, and LASSO estimators with $p = 200$ and $n = 1170, 7800, 23400$. As seen in Table 1, the errors of the proposed FATEN-LASSO estimator are usually decreasing as the sample size n increases and have the smallest values overall. This may be because the proposed FATEN-LASSO estimator can fully handle the noises in the data, strong dependence in the covariate process, and the time variation of the high-dimensional coefficient processes. The TED and LASSO estimators have low false positive rates but high false negative rates. This is because they often estimate the integrated coefficients as 0.

Table 1: Average rates of false positives and false negatives of the FATEN-LASSO, FATEC-LASSO, TED, and LASSO estimators for $p = 200$ and $n = 1170, 7800, 23400$.

	False positive rate			
	Estimator			
n	FATEN-LASSO	FATEC-LASSO	TED	LASSO
1170	0.000	0.000	0.000	0.001
7800	0.000	0.000	0.000	0.000
23400	0.000	0.000	0.025	0.000
	False negative rate			
	Estimator			
n	FATEN-LASSO	FATEC-LASSO	TED	LASSO
1170	0.680	0.938	1.000	0.847
7800	0.019	0.176	0.719	0.995
23400	0.004	0.038	0.714	1.000

Figure 2 plots the log max, ℓ_1, and ℓ_2 norm errors of the FATEN-LASSO, FATEC-LASSO, TED, and LASSO estimators with $n = 1170, 7800, 23400$. As seen in Figure 2, the estimation errors of the FATEN-LASSO estimator are decreasing as the sample size n increases. As expected, the FATEN-LASSO estimator outperforms other estimators for all error norms. This may be because only the FATEN-LASSO estimator can fully handle the factor structure in the covariate process, the microstructure noise of high-frequency data, and the time variation in the coefficient process. We note that the TED and LASSO estimators are not consistent. One possible explanation for this is that the proportion of the noise in log-returns increases as the sample size n increases. These results indicate that the proposed FATEN-LASSO estimator can help deal with the strongly correlated structure of the covariates, microstructure noises, and time-varying coefficient process when estimating high-dimensional integrated coefficients. In Appendix A.3, we reported the performance of the FATEN-LASSO and FATEC-LASSO estimators when the rank r is always selected as 0. FATEN-LASSO still shows the best performance. From this result, we can conclude that when the covariates are strongly correlated, the factor-adjusted scheme improves the performance. Details can be found in Appendix A.3.

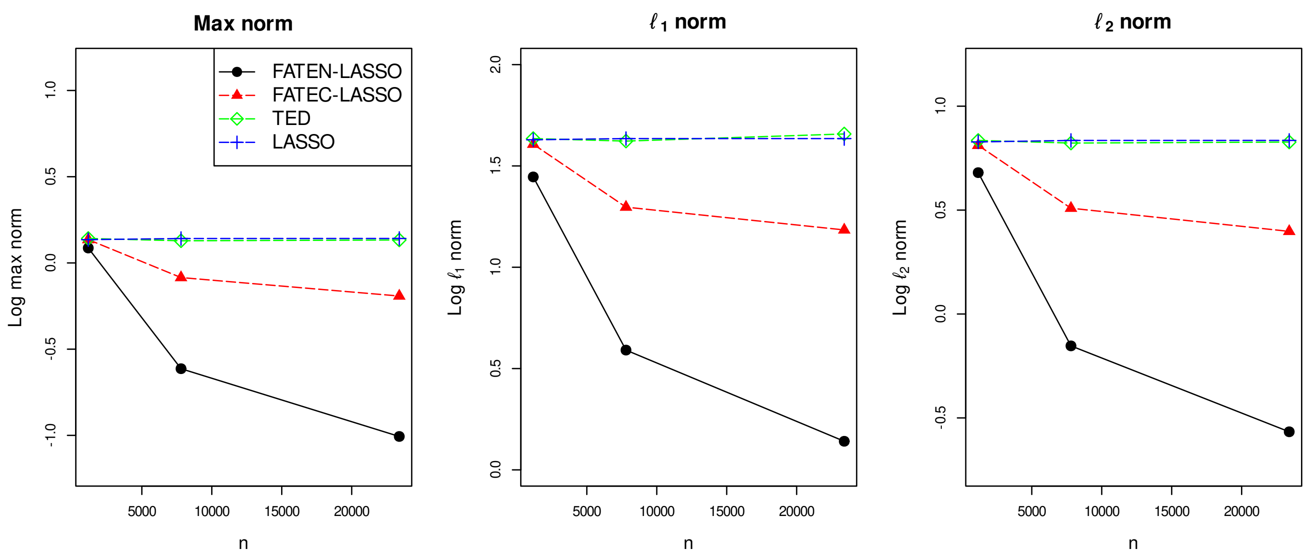

Figure 2: The log max, ℓ_1, and ℓ_2 norm error plots of the FATEN-LASSO (black dot), FATEC-LASSO (red triangle), TED (green diamond), and LASSO (blue plus) estimators for $p = 200$ and $n = 1170, 7800, 23400$.

5 An empirical study

We applied the proposed FATEN-LASSO estimator to real high-frequency trading data, collected from January 2013 to December 2019. We obtained stock price data from the End of Day website (https://eoddata.com/), futures price data from the FirstRate Data website (https://firstratedata .com/), and firm fundamentals from the Center for Research in Security Prices (CRSP)/Compustat

Merged Database. We collected 1-min log-price data using the previous tick scheme (Wang and Zou, 2010; Zhang, 2011) and excluded half trading days. For the dependent process, we considered five assets: Apple Inc. (AAPL), Berkshire Hathaway Inc. (BRK.B), General Motors Company (GM), Alphabet Inc. (GOOG), and Exxon Mobil Corporation (XOM). These assets have the largest market values in the following global industrial classification standards (GICS) sectors: information technology, financials, consumer discretionary, communication services, and energy. For the covariate process, we first constructed Fama-French five factors (Fama and French, 2015) and the momentum factor (Carhart, 1997) using 1-min high-frequency data. MKT, HML, SMB, RMW, CMA, and MOM denote the market, value, size, profitability, investment, and momentum factors, respectively. Then, we constructed 1-min high-frequency factor portfolios listed in Jensen et al. (2023), who grouped factors based on the hierarchical agglomerative clustering (Murtagh and Legendre, 2014). Among the groups in Jensen et al. (2023), we constructed all factors in the value, size, profitability, investment, and momentum groups. The group, symbol, description, and original publication source of the factor portfolios are listed in Table 4 in Appendix A.4. We note that the value, size, profitability, investment, and momentum groups have 18, 5, 11, 22, and 8 constituents, respectively. In summary, we used the five assets and 70 factors for the dependent and covariate processes, respectively. The detailed data processing procedure can be found in Appendix A.4.

When calculating the FATEN-LASSO estimator, we employed the tuning parameter choice procedure discussed in Section 4 and Appendix A.1. Moreover, we chose $k_2 = 390$, that is, we estimated instantaneous coefficients on a daily basis. To choose the rank r, we first obtained the FATEN-LASSO and FATEC-LASSO estimators for $r = 0, 1, 2, 3, 4, 5$. Then, for each estimator, we chose the value of r that minimized the corresponding MSE defined in (4.1). To select the tuning parameter c_h, we utilized the mean squared prediction error (MSPE) using the data in 2013. We first defined

$$\Lambda(c_h) = \frac{1}{55} \sum_{m=1}^{11} \sum_{s=1}^{5} \left\| \widetilde{I\beta}^{m,s}(c_h) - \widehat{I\beta}^{(m+1),s} \right\|_2^2,$$

where $\widetilde{I\beta}^{m,s}(c_h)$ is the FATEN-LASSO estimator obtained with the tuning parameter c_h, and $\widehat{I\beta}^{m,s}$ is the debiased integrated coefficient estimator from the mth month in 2013 and the sth stock. Then, we chose c_h by minimizing $\Lambda(c_h)$ over $c_h \in \{0.05l \mid 5 \leq l \leq 10, l \in \mathbb{Z}\}$. The result is $c_h = 0.25$. We implemented the same procedure for the FATEC-LASSO estimator and obtained the same result. We note that the stationarity condition on the coefficient process is reasonable, which motivates the above tuning parameter choice procedure. Then, we obtained the monthly integrated coefficients based on the FATEN-LASSO, FATEC-LASSO, TED, and LASSO estimation procedures for each of the five assets. The coefficients of the non-trading period were set as zero.

Table 2: Annual average in-sample and out-of-sample R^2 for the FATEN-LASSO, FATEC-LASSO, TED, and LASSO estimators over the five assets.

	In-sample R^2			
	Estimator			
	FATEN-LASSO	FATEC-LASSO	TED	LASSO
whole period	0.344	0.312	0.264	0.244
2013	0.322	0.318	0.201	0.195
2014	0.327	0.306	0.236	0.219
2015	0.322	0.256	0.293	0.257
2016	0.351	0.311	0.267	0.253
2017	0.295	0.278	0.186	0.185
2018	0.418	0.372	0.381	0.351
2019	0.372	0.344	0.283	0.250
	Out-of-sample R^2			
	Estimator			
	FATEN-LASSO	FATEC-LASSO	TED	LASSO
whole period	0.313	0.234	0.259	0.236
2014	0.274	0.213	0.205	0.188
2015	0.300	0.208	0.280	0.253
2016	0.314	0.216	0.258	0.237
2017	0.273	0.206	0.173	0.171
2018	0.387	0.317	0.369	0.337
2019	0.330	0.241	0.267	0.229

We first compared the performances of the FATEN-LASSO, FATEC-LASSO, TED, and LASSO estimators based on the monthly in-sample and out-of-sample R^2. The out-of-sample R^2 was obtained using the integrated coefficients from the previous month. For the out-of-sample R^2, we excluded the year 2013, since we chose the tuning parameters using the data from 2013. Then, we obtained the annual average R^2 across the five assets. Table 2 reports the annual average in-sample and out-of-sample R^2 for the FATEN-LASSO, FATEC-LASSO, TED, and LASSO estimators. From Table 2, we can see that the FATEN-LASSO estimator shows the best performance for all periods. This may be because the proportion of microstructure noises in the 1-min high-frequency data is not negligible, and only the FATEN-LASSO estimator can fully handle the noises, time-varying coefficient processes, and factor structure of the covariate processes. In Appendix A.5, we investigated the performance of the FATEN-LASSO and FATEC-LASSO estimators in the case where the rank r was always selected as 0. These results with zero rank are similar to the FATEN-LASSO and FATEC-LASSO estimators because the rank r was frequently chosen as zero. Details can be found in Appendix A.5.

Table 3: Monthly average of the non-zero frequency of the FATEN-LASSO, FATEC-LASSO, TED, and LASSO estimators over 70 factors and 84 months for the five assets.

	Non-zero frequency			
	Estimator			
	FATEN-LASSO	FATEC-LASSO	TED	LASSO
AAPL	25.214	47.809	3.261	4.273
BRK.B	28.690	53.130	3.750	4.738
GM	26.583	54.785	3.976	4.523
GOOG	23.821	50.047	4.797	4.583
XOM	26.940	52.011	5.047	6.714

Table 3 reports the monthly average of non-zero frequency of the FATEN-LASSO, FATEC-LASSO, TED, and LASSO estimators over 70 factors and 84 months for the five assets. We note that the term "non-zero frequency" refers to the number of coefficient entries that are non-zero out of 70 entries. As seen in Table 3, the FATEN-LASSO estimator is more sparse than the FATEC-LASSO estimator. Combining the results in Tables 2 and 3, we can conjecture that the proposed FATEN-LASSO estimator can better account for market dynamics than the FATEC-LASSO estimator with a simpler model. On the other hand, the TED and LASSO estimators are more sparse and exhibit lower R^2. This may be because they struggle to capture market signals effectively due to the noisy data, even though there are many factors in financial markets, as represented by the factor zoo (Jensen et al., 2023). In Appendix A.6, we investigated the integrated coefficient estimates obtained from the FATEN-LASSO procedure. The results show the time-varying property and sparsity of the coefficient process. Additionally, the findings support the multi-factor models (Asness et al., 2013; Carhart, 1997; Fama and French, 1992, 2015) and factor zoo analysis (Jensen et al., 2023). Details can be found in Appendix A.6.

6 Conclusion

In this paper, we developed a novel FATEN-LASSO estimation procedure that can accommodate the microstructure noise of high-frequency data, factor structure of the covariate process, and time variation in the high-dimensional coefficient process. To handle the noise and factor structure, we first smoothed the observed variables and applied PCA to the smoothed covariate process. Then, to estimate the instantaneous coefficient, we employed the nonconvex optimization with the smoothed variables. We showed that the proposed instantaneous coefficient estimator can handle the noises, factor structure, and the high-dimensional time-varying coefficient process with the sharp convergence rate. To handle the bias from the ℓ_1-regularization, we utilized the debiasing scheme and obtained the debiased integrated coefficient estimator using debiased instantaneous coefficient estimators. Then, we further regularized the debiased integrated coefficient estimator to account for the sparse structure of the coefficient process. We showed that the proposed FATEN-LASSO estimator achieves the sharp convergence rate. We still hold this property even if the factor structure in the covariate processes does not exist. In the empirical study, the FATEN-LASSO estimation procedure performs best overall in terms of both in-sample and out-of-sample R^2. This finding suggests that, when estimating high-dimensional integrated coefficients based on high-frequency data with a factor structure, the proposed FATEN-LASSO estimation method helps handle the time-varying property of the coefficient process as well as the microstructure noise in high-frequency data.

References

Agarwal, A., Negahban, S., and Wainwright, M. J. (2012). Fast global convergence of gradient methods for high-dimensional statistical recovery. *The Annals of Statistics*, 40(5):2452–2482.

Aït-Sahalia, Y., Fan, J., and Xiu, D. (2010). High-frequency covariance estimates with noisy and asynchronous financial data. *Journal of the American Statistical Association*, 105(492):1504–1517.

Aït-Sahalia, Y., Kalnina, I., and Xiu, D. (2020). High-frequency factor models and regressions. *Journal of Econometrics*, 216(1):86–105.

Aït-Sahalia, Y. and Xiu, D. (2017). Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *Journal of Econometrics*, 201(2):384–399.

Aït-Sahalia, Y. and Xiu, D. (2019). Principal component analysis of high-frequency data. *Journal of the American Statistical Association*, 114(525):287–303.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, 5(1):31–56.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Wu, J. (2005). A framework for exploring the macroeconomic determinants of systematic risk. *American Economic Review*, 95(2):398–404.

Andersen, T. G., Thyrsgaard, M., and Todorov, V. (2021). Recalcitrant betas: Intraday variation in the cross-sectional dispersion of systematic risk. *Quantitative Economics*, 12(2):647–682.

Anderson, C. W. and Garcia-Feijoo, L. (2006). Empirical evidence on capital investment, growth options, and security returns. *The Journal of Finance*, 61(1):171–194.

Asness, C. S., Moskowitz, T. J., and Pedersen, L. H. (2013). Value and momentum everywhere. *The Journal of Finance*, 68(3):929–985.

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

Bali, T. G., Cakici, N., and Whitelaw, R. F. (2011). Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics*, 99(2):427–446.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, 9(1):3–18.

Barbee Jr, W. C., Mukherji, S., and Raines, G. A. (1996). Do sales–price and debt–equity explain stock returns better than book–market and firm size? *Financial Analysts Journal*, 52(2):56–60.

Barigozzi, M., Cho, H., and Owens, D. (2024). Fnets: Factor-adjusted network estimation and forecasting for high-dimensional time series. *Journal of Business & Economic Statistics*, 42(3):890–902.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2011). Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics*, 162(2):149–169.

Barndorff-Nielsen, O. E. and Shephard, N. (2004). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica*, 72(3):885–925.

Basu, S. (1983). The relationship between earnings' yield, market value and return for nyse common stocks: Further evidence. *Journal of Financial Economics*, 12(1):129–156.

Belo, F. and Lin, X. (2012). The inventory growth spread. *The Review of Financial Studies*, 25(1):278–313.

Belo, F., Lin, X., and Bazdresch, S. (2014). Labor hiring, investment, and stock return predictability in the cross section. *Journal of Political Economy*, 122(1):129–177.

Bhandari, L. C. (1988). Debt/equity ratio and expected common stock returns: Empirical evidence. *The Journal of Finance*, 43(2):507–528.

Blitz, D., Huij, J., and Martens, M. (2011). Residual momentum. *Journal of Empirical Finance*, 18(3):506–521.

Bouchaud, J.-P., Krueger, P., Landier, A., and Thesmar, D. (2019). Sticky expectations and the profitability anomaly. *The Journal of Finance*, 74(2):639–674.

Boudoukh, J., Michaely, R., Richardson, M., and Roberts, M. R. (2007). On the importance of measuring payout yield: Implications for empirical asset pricing. *The Journal of Finance*, 62(2):877–915.

Bradshaw, M. T., Richardson, S. A., and Sloan, R. G. (2006). The relation between corporate financing activities, analysts' forecasts and stock returns. *Journal of Accounting and Economics*, 42(1-2):53–85.

Brennan, M. J., Chordia, T., and Subrahmanyam, A. (1998). Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. *Journal of Financial Economics*, 49(3):345–373.

Cai, T., Liu, W., and Luo, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607.

Campbell, J. Y., Hilscher, J., and Szilagyi, J. (2008). In search of distress risk. *The Journal of Finance*, 63(6):2899–2939.

Candes, E. and Tao, T. (2007). The dantzig selector: Statistical estimation when p is much larger than n. *The Annals of Statistics*, 35(6):2313–2351.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52(1):57–82.

Chan, L. K., Lakonishok, J., and Sougiannis, T. (2001). The stock market valuation of research and development expenditures. *The Journal of Finance*, 56(6):2431–2456.

Chen, D., Mykland, P. A., and Zhang, L. (2024). Realized regression with asynchronous and noisy high frequency and high dimensional data. *Journal of Econometrics*, 239(2):105446.

Chen, R. Y. (2018). Inference for volatility functionals of multivariate itô semimartingales observed with jump and noise. *arXiv preprint arXiv:1810.04725*.

Chordia, T., Subrahmanyam, A., and Anshuman, V. R. (2001). Trading activity and expected stock returns. *Journal of Financial Economics*, 59(1):3–32.

Christensen, K., Kinnebrock, S., and Podolskij, M. (2010). Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *Journal of Econometrics*, 159(1):116–133.

Chun, D. and Kim, D. (2022). State heterogeneity analysis of financial volatility using high-frequency financial data. *Journal of Time Series Analysis*, 43(1):105–124.

Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of Finance*, 66(4):1047–1108.

Cooper, M. J., Gulen, H., and Schill, M. J. (2008). Asset growth and the cross-section of stock returns. *the Journal of Finance*, 63(4):1609–1651.

Dai, C., Lu, K., and Xiu, D. (2019). Knowing factors or factor loadings, or neither? evaluating estimators of large covariance matrices with noisy and asynchronous data. *Journal of Econometrics*, 208(1):43–79.

Daniel, K. and Titman, S. (2006). Market reactions to tangible and intangible information. *The Journal of Finance*, 61(4):1605–1643.

De Bondt, W. F. and Thaler, R. (1985). Does the stock market overreact? *The Journal of Finance*, 40(3):793–805.

Dechow, P. M., Sloan, R. G., and Soliman, M. T. (2004). Implied equity duration: A new measure of equity risk. *Review of Accounting Studies*, 9:197–228.

Desai, H., Rajgopal, S., and Venkatachalam, M. (2004). Value-glamour and accruals mispricing: One anomaly or two? *The Accounting Review*, 79(2):355–385.

Dichev, I. D. (1998). Is the risk of bankruptcy a systematic risk? *the Journal of Finance*, 53(3):1131–1147.

Fairfield, P. M., Whisenant, J. S., and Yohn, T. L. (2003). Accrued earnings and growth: Implications for future profitability and market mispricing. *The Accounting Review*, 78(1):353–371.

Fama, E. and French, K. (1992). The cross-section of expected stock returns. *The Journal of Finance*, 47(2):427–465.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22.

Fan, J., Ke, Y., and Wang, K. (2020). Factor-adjusted regularized model selection. *Journal of Econometrics*, 216(1):71–85.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, 96(456):1348–1360.

Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(4):603–680.

Fan, J., Lou, Z., and Yu, M. (2024). Are latent factor regression and sparse regression adequate? *Journal of the American Statistical Association*, 119(546):1076–1088.

Ferson, W. E. and Harvey, C. R. (1999). Conditioning variables and the cross section of stock returns. *The Journal of Finance*, 54(4):1325–1360.

Frankel, R. and Lee, C. M. (1998). Accounting valuation, market expectation, and cross-sectional stock returns. *Journal of Accounting and Economics*, 25(3):283–319.

George, T. J. and Hwang, C.-Y. (2004). The 52-week high and momentum investing. *The Journal of Finance*, 59(5):2145–2176.

Harvey, C. R., Liu, Y., and Zhu, H. (2016). ... and the cross-section of expected returns. *The Review of Financial Studies*, 29(1):5–68.

Haugen, R. A. and Baker, N. L. (1996). Commonality in the determinants of expected stock returns. *Journal of Financial Economics*, 41(3):401–439.

Heston, S. L. and Sadka, R. (2008). Seasonality in the cross-section of stock returns. *Journal of Financial Economics*, 87(2):418–445.

Hirshleifer, D., Hou, K., Teoh, S. H., and Zhang, Y. (2004). Do investors overvalue firms with bloated balance sheets? *Journal of Accounting and Economics*, 38:297–331.

Hou, K., Xue, C., and Zhang, L. (2015). Digesting anomalies: An investment approach. *The Review of Financial Studies*, 28(3):650–705.

Hou, K., Xue, C., and Zhang, L. (2020). Replicating anomalies. *The Review of Financial Studies*, 33(5):2019–2133.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., and Vetter, M. (2009). Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic processes and their applications*, 119(7):2249–2276.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1):65–91.

Jensen, T. I., Kelly, B., and Pedersen, L. H. (2023). Is there a replication crisis in finance? *The Journal of Finance*, 78(5):2465–2518.

Kalnina, I. (2023). Inference for nonparametric high-frequency estimators with an application to time variation in betas. *Journal of Business & Economic Statistics*, 41(2):538–549.

Kim, D., Oh, M., and Shin, M. (2024). High-dimensional time-varying coefficient estimation. *arXiv preprint arXiv:2202.08419*.

Kneip, A. and Sarda, P. (2011). Factor models and variable selection in high-dimensional regression analysis. *The Annals of Statistics*, 39(5):2410–2447.

Lakonishok, J., Shleifer, A., and Vishny, R. W. (1994). Contrarian investment, extrapolation, and risk. *The Journal of Finance*, 49(5):1541–1578.

Litzenberger, R. H. and Ramaswamy, K. (1979). The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence. *Journal of Financial Economics*, 7(2):163–195.

Loh, P.-L. and Wainwright, M. J. (2012). High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity. *The Annals of Statistics*, 40(3):1637–1664.

Loughran, T. and Wellman, J. W. (2011). New evidence on the relation between the enterprise multiple and average stock returns. *Journal of Financial and Quantitative Analysis*, 46(6):1629–1650.

Lyandres, E., Sun, L., and Zhang, L. (2008). The new issues puzzle: Testing the investment-based explanation. *The Review of Financial Studies*, 21(6):2825–2855.

McLean, R. D. and Pontiff, J. (2016). Does academic research destroy stock return predictability? *The Journal of Finance*, 71(1):5–32.

Miller, M. H. and Scholes, M. S. (1982). Dividends and taxes: Some empirical evidence. *Journal of Political Economy*, 90(6):1118–1141.

Murtagh, F. and Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: which algorithms implement ward's criterion? *Journal of classification*, 31:274–295.

Mykland, P. A. and Zhang, L. (2009). Inference for continuous semimartingales observed at high frequency. *Econometrica*, 77(5):1403–1445.

Oh, M., Kim, D., and Wang, Y. (2024). Robust realized integrated beta estimator with application to dynamic analysis of integrated beta. *Journal of Econometrics*, page 105810.

Ortiz-Molina, H. and Phillips, G. M. (2014). Real asset illiquidity and the cost of capital. *Journal of Financial and Quantitative Analysis*, 49(1):1–32.

Penman, S. H., Richardson, S. A., and Tuna, I. (2007). The book-to-price effect in stock returns: accounting for leverage. *Journal of Accounting Research*, 45(2):427–467.

Piotroski, J. D. (2000). Value investing: The use of historical financial statement information to separate winners from losers. *Journal of Accounting Research*, pages 1–41.

Pontiff, J. and Woodgate, A. (2008). Share issuance and cross-sectional returns. *The Journal of Finance*, 63(2):921–945.

Reiß, M., Todorov, V., and Tauchen, G. (2015). Nonparametric test for a constant beta between itô semi-martingales based on high-frequency data. *Stochastic Processes and their Applications*, 125(8):2955–2988.

Richardson, S. A., Sloan, R. G., Soliman, M. T., and Tuna, I. (2005). Accrual reliability, earnings persistence and stock prices. *Journal of Accounting and Economics*, 39(3):437–485.

Rosenberg, B., Reid, K., and Lanstein, R. (1985). Persuasive evidence of market inefficiency. *Journal of Portfolio Management*, 11(3):9–16.

Shin, M. and Kim, D. (2023). Robust high-dimensional time-varying coefficient estimation. *arXiv preprint arXiv:2302.13658*.

Soliman, M. T. (2008). The use of dupont analysis by market participants. *The Accounting Review*, 83(3):823–853.

Stambaugh, R. F. and Yuan, Y. (2017). Mispricing factors. *The Review of Financial Studies*, 30(4):1270–1315.

Thomas, J. K. and Zhang, H. (2002). Inventory changes and future returns. *Review of Accounting Studies*, 7(2):163–187.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Wang, Y. and Zou, J. (2010). Vast volatility matrix estimation for high-frequency financial data. *The Annals of Statistics*, 38(2):943–978.

Xie, H. (2001). The mispricing of abnormal accruals. *The Accounting Review*, 76(3):357–373.

Zhang, L. (2011). Estimating covariation: Epps effect, microstructure noise. *Journal of Econometrics*, 160(1):33–47.

A Appendix

A.1 Discussion on tuning parameter selection

In this section, we discuss the process of selecting the tuning parameter for the FATEN-LASSO estimator. To obtain the smoothed variables, we choose

$$g(x) = x \wedge (1-x) \quad \text{and} \quad k_1 = \frac{1}{2} n^{1/2}.$$

We select $k_2 = n^{3/4}$ in the simulation study, and the selection of k_2 for the empirical study is described in Section 5. For the jump truncation parameters in (3.1) and (3.5), we use

$$w_n = 3\,\mathrm{sd}\left(\Delta_i^n \widehat{Y}\right), \quad v_{j,n} = 3\,\mathrm{sd}\left(\Delta_i^n \widehat{X}_j\right), \quad \text{and} \quad v_{j,n}^{(2)} = 3\,\mathrm{sd}\left(\Delta_i^n X_j^o\right), \tag{A.1}$$

where sd represents the sample standard deviation. After Step 2 in Algorithm 1 in Section A.11, each column of $\mathcal{Y}_i$ and $\widehat{\mathbf{G}}_i$ is standardized to have a mean of zero and a variance of 1. For $j = 1, \ldots, p$, the j-th entry of $\Delta_i^n \widehat{\mathbf{X}}^{\mathrm{trunc}}$ is divided by the standard deviation for the j-th column of $\widehat{\mathbf{G}}_i$. We conduct re-scaling after obtaining the FATEN-LASSO estimator. For the nonconvex optimization, we implement the composite gradient descent method (3.16) with 10^3 updates. We set $\widehat{\boldsymbol{\theta}}_0^0 = (0, \ldots, 0)^\top \in \mathbb{R}^{p+r}$ and $\widehat{\boldsymbol{\theta}}_{i\Delta_n}^0 = \widehat{\boldsymbol{\theta}}_{(i-k_2)\Delta_n}$ for $i \geq k_2$. That is, we set the initial parameter as the previous estimator. We choose α_2 as the largest eigenvalue of $\dfrac{n}{\phi k_1 k_2} \widehat{\mathbf{G}}_i^\top \widehat{\mathbf{G}}_i$ and set $\rho = 10$. Furthermore, we select

$$\eta = c_\eta n^{-1/8} (\log p)^2, \quad \tau = c_\tau n^{-1/8} (\log p)^2, \quad \text{and} \quad h_n = c_h n^{-1/4}, \tag{A.2}$$

where c_η, c_τ, and c_h are the tuning parameters. We choose $c_\eta \in [10^{-4}, 10^4]$ by minimizing the corresponding Bayesian information criterion (BIC). We select $c_\tau \in [10^{-2}, 10^2]$, which minimizes the following loss function:

$$\mathrm{tr}\left[\left(\left\{\frac{n}{\phi k_1 k_2}\widehat{\mathbf{U}}_i^\top \widehat{\mathbf{U}}_i - \frac{n\zeta}{\phi k_1}\widehat{\mathbf{V}}^X\right\}\widehat{\boldsymbol{\Omega}}_{i\Delta_n} - \mathbf{I}_p\right)^2\right],$$

where $\mathbf{I}_p$ is the p-dimensional identity matrix. In the simulation study, we choose c_h as 1. For the empirical study, c_h is selected by minimizing the corresponding mean squared prediction error (MSPE). The specific procedure is described in Section 5.

A.2 Simulation setup

We considered the following factor-based regression jump diffusion models in (2.1)–(2.3).

$$\begin{aligned}
&dY(t) = \boldsymbol{\beta}^\top(t)d\mathbf{X}^c(t) + dZ(t) + J^y(t)d\Lambda^y(t), \quad d\mathbf{X}(t) = d\mathbf{X}^c(t) + d\mathbf{X}^J(t),\\
&d\mathbf{X}^c(t) = \mathbf{B}(t)d\mathbf{f}(t) + d\mathbf{u}(t), \quad d\mathbf{X}^J(t) = \mathbf{J}(t)d\boldsymbol{\Lambda}(t), \quad d\mathbf{f}(t) = \boldsymbol{\nu}_f(t)d\mathbf{W}_f(t),\\
&d\mathbf{u}(t) = \boldsymbol{\nu}_u(t)d\mathbf{W}_u(t), \quad \text{and} \quad dZ(t) = \nu_z(t)dW_z(t),
\end{aligned}$$

where the jump sizes $J^Y(t)$ and $J_j(t)$ were generated from the independent normal distribution with a mean of zero and a standard deviation of 0.05, and $\Lambda^y(t)$ and $\Lambda_j(t)$ were generated using Poisson processes with intensities of 15 and $(10,\dots,10)^\top$, respectively. The initial values $Y(0)$ and $X_j(0)$ were set as zero, and $\nu_z(t)$ was generated from the following Ornstein–Uhlenbeck process:

$$d\nu_z(t) = 4\left(0.12 - \nu_z(t)\right)dt + 0.03dW_z(t),$$

where $\nu_z(0) = 0.15$ and $W_z(t)$ is an independent Brownian motion. For the volatility processes $\boldsymbol{\nu}_f(t)$ and $\boldsymbol{\nu}_u(t)$, we first generated the following Ornstein–Uhlenbeck processes:

$$d\xi_f(t) = 3\left(0.25 - \xi_f(t)\right)dt + 0.05dW_f^\xi(t),$$

$$d\xi_u(t) = 5\left(0.5 - \xi_u(t)\right)dt + 0.1 dW_u^{\xi}(t),$$

where $\xi_f(0) = 0.25$, $\xi_u(0) = 0.45$, and $W_f^{\xi}(t)$ and $W_u^{\xi}(t)$ are r-dimensional and p-dimensional independent Brownian motions, respectively. Then, $\boldsymbol{\nu}_f(t)$ and $\boldsymbol{\nu}_u(t)$ were taken to be the Cholesky decompositions of $\mathbf{\Sigma}_f(t) = (\Sigma_{f,ij}(t))_{1\le i,j\le r}$ and $\mathbf{\Sigma}_u(t) = (\Sigma_{u,ij}(t))_{1\le i,j\le p}$, respectively, where $\Sigma_{f,ij}(t) = \xi_f(t)0.5^{|i-j|}$ and $\Sigma_{u,ij}(t) = 0.3\xi_u(t)\mathbf{I}_p$. For the coefficient process $\boldsymbol{\beta}(t)$, we considered the following model:

$$d\boldsymbol{\beta}(t) = \boldsymbol{\mu}_{\beta}(t)dt + \boldsymbol{\nu}_{\beta}(t)d\mathbf{W}_{\beta}(t),$$

where $\boldsymbol{\mu}_{\beta}(t) = \left(\mu_{\beta,1}(t), \ldots, \mu_{\beta,p}(t)\right)^{\top}$ is a drift process, $\boldsymbol{\nu}_{\beta}(t) = \left(\nu_{\beta,ij}(t)\right)_{1\le i,j\le p}$ is an instantaneous volatility matrix process, and $\mathbf{W}_{\beta}(t)$ is a p-dimensional independent Brownian motion. To generate $\boldsymbol{\nu}_{\beta}(t)$, we first generated the following Ornstein–Uhlenbeck process:

$$d\varphi_{\beta}(t) = 3\left(0.25 - \varphi_{\beta}(t)\right)dt + 0.05 dW_{\beta}^{\varphi}(t),$$

where $\varphi_{\beta}(0) = 0.15$ and $W_{\beta}^{\varphi}(t)$ is an independent Brownian motion. Then, we set $\left(\nu_{\beta,ij}(t)\right)_{1\le i,j\le s_p}$ as $\varphi_{\beta}(t)\mathbf{I}_{s_p}$, where $\mathbf{I}_{s_p}$ is the s_p-dimensional identity matrix. For $j = 1, \ldots, s_p$, we took $\beta_j(0) = 1$ and $\mu_{j,\beta}(t) = 0.05$ for $0 \le t \le 1$, whereas we set $\beta_j(t) = 0$ for $j = s_p + 1, \ldots, p$. For the factor loading matrix process $\mathbf{B}(t)$, we employed the following model:

$$d\text{vec}\left(\mathbf{B}(t)\right) = \boldsymbol{\mu}_B(t)dt + \boldsymbol{\nu}_B(t)d\mathbf{W}_B(t),$$

where $\boldsymbol{\mu}_B(t) = \left(\mu_{B,1}(t), \ldots, \mu_{B,pr}(t)\right)^{\top}$ is a drift process, $\boldsymbol{\nu}_B(t) = \left(\nu_{B,ij}(t)\right)_{1\le i,j\le pr}$ is an instantaneous volatility matrix process, and $\mathbf{W}_B(t)$ is a pr-dimensional independent Brownian motion. To obtain $\boldsymbol{\nu}_B(t)$, we first employed the following Ornstein–Uhlenbeck process:

$$d\varphi_B(t) = 3\left(0.07 - \varphi_B(t)\right) dt + 0.02 dW_B^{\varphi}(t),$$

where $\varphi_B(0) = 0.1$ and $W_B^{\varphi}(t)$ is an independent Brownian motion. Then, we set $(\nu_{B,ij}(t))_{1\le i,j\le pr}$ as $\varphi_B(t)\mathbf{I}_{pr}$, where $\mathbf{I}_{pr}$ is the pr-dimensional identity matrix. The initial value $B_{ij}(0)$ is obtained from an independent normal distribution with a mean of zero and a standard deviation of 0.3, and we set $\mu_{B,ij}(t) = 0.025$ for $1 \le i \le p$, $1 \le j \le r$, and $0 \le t \le 1$.

A.3 Simulation study for rank $r = 0$ selection

In this section, we conducted a simulation study to check the performance of the FATEN-LASSO and FATEC-LASSO estimators when the rank r was always selected as 0. We denote the TEN-LASSO and TEC-LASSO estimators as the FATEN-LASSO and FATEC-LASSO estimators, respectively, when r is set to 0. Thus, the TEN-LASSO and TEC-LASSO estimators can be considered special cases of the FATEN-LASSO and FATEC-LASSO estimators. Figure 3 depicts the log max, ℓ_1, and ℓ_2 norm errors of the FATEN-LASSO, FATEC-LASSO, TEN-LASSO, and TEC-LASSO estimators with $n = 1170, 7800, 23400$. From Figure 3, we find that the FATEN-LASSO estimator outperforms the other estimators. This may be because only the FATEN-LASSO estimator can handle both the factor structure in the covariate process and the noise in data.

A.4 Data processing procedure for the empirical study

In this section, we discuss the data processing procedure for the factors used in Section 5. For each of the Fama-French five factors (Fama and French, 2015) and the momentum factor (Carhart, 1997), we first obtained the monthly portfolio constituents among the stocks traded on the NYSE, NASDAQ, and AMEX. Specifically, we obtained MKT as the return of a value-weighted portfolio

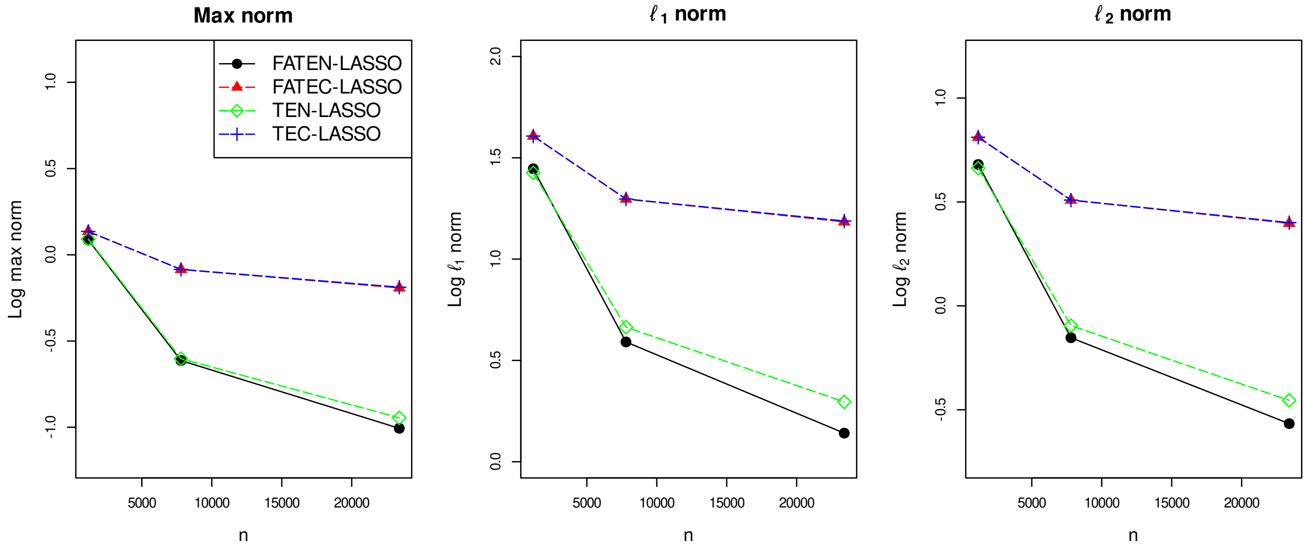

Figure 3: The log max, ℓ_1, and ℓ_2 norm error plots of the FATEN-LASSO (black dot), FATEC-LASSO (red triangle), TEN-LASSO (green diamond), and TEC-LASSO (blue plus) estimators for $p = 200$ and $n = 1170, 7800, 23400$.

of all assets, and the other factors were calculated as follows:

$$HML = (SH + BH)/2 - (SL + BL)/2,$$
$$SMB = (SH + SM + SL)/3 - (BH + BM + BL)/3,$$
$$RMW = (SR + BR)/2 - (SW + BW)/2,$$
$$CMA = (SC + BC)/2 - (SA + BA)/2,$$
$$MOM = (SU + BU)/2 - (SD + BD)/2,$$

where small (S) and big (B) portfolios consist of assets with small and big market equities, respectively, and we classified high (H), medium (M), and low (L) portfolios based on the ratio of book equity to market equity. In addition, robust (R), neutral (N), and weak (W) portfolios were classified by profitability, and we obtained the constituents of conservative (C), neutral (N), and aggressive (A) portfolios using the investment data. Finally, up (U), flat (F), and down (D) portfolios were classified by the momentum of the return. The details of this process can be found in Aït-Sahalia et al. (2020); Kim et al. (2024). Then, using 1-min high-frequency data, we calculated the portfolio returns as follows:

$$WRet_{d,i} = \frac{\sum_{s=1}^{N_d} w_{d,i}^s \times Ret_{d,i}^s}{\sum_{s=1}^{N_d} w_{d,i}^s},$$

where $WRet_{d,i}$ represents the portfolio return for the dth day and ith time interval, N_d is the number of assets in the portfolio for the dth day (superscript s represents the sth asset of the portfolio), and $w_{d,i}^s$ is defined by

$$w_{d,i}^s = w_d^s \times \prod_{l=0}^{i-1} \left(1 + Ret_{d,l}^s\right),$$

where w_d^s represents the market capitalization obtained using the close price of the sth stock on the $(d-1)$th day, and $Ret_{d,0}^s$ denotes the overnight return from the $(d-1)$th day to the dth day. Similarly, we obtained 1-min high-frequency factor portfolios in Table 4 below. We first obtained the monthly constituents for the high and low portfolios based on the terciles for each signal. Then, each factor portfolio was constructed by subtracting the return of the high portfolio from the return of the low portfolio. The detailed process can be found in Jensen et al. (2023).

A.5 Empirical study for rank $r = 0$ selection

In this section, we conducted an empirical study to investigate the performance of the TEN-LASSO and TEC-LASSO estimators. Table 5 presents the annual average in-sample and out-of-sample R^2 for the FATEN-LASSO, FATEC-LASSO, TEN-LASSO, and TEC-LASSO estimators. From Table 5, we can see that the FATEN-LASSO and TEN-LASSO estimators perform similarly. The FATEC-LASSO and TEC-LASSO estimators also show similar results. These results may be because the effect of the common factor is not significant compared to the estimation error for the rank r, and the rank r is selected as 0 for 72.6% and 96.4% of the whole months for the FATEN-LASSO and

Table 4: The group, symbol, description, and citation for 64 factor portfolios used in Section 5. We used the same factor symbols as those in Jensen et al. (2023).

Group	Symbol	Description	Citation
Value	at_me	Assets-to-market	Fama and French (1992)
	be_me	Book-to-market equity	Rosenberg et al. (1985)
	bev_mev	Book-to-market enterprise value	Penman et al. (2007)
	chcsho_12m	Net stock issues	Pontiff and Woodgate (2008)
	debt_me	Debt-to-market	Bhandari (1988)
	div12m_me	Dividend yield	Litzenberger and Ramaswamy (1979)
	ebitda_mev	Ebitda-to-market enterprise value	Loughran and Wellman (2011)
	eq_dur	Equity duration	Dechow et al. (2004)
	eqnetis_at	Net equity issuance	Bradshaw et al. (2006)
	eqnpo_12m	Equity net payout	Daniel and Titman (2006)
	eqnpo_me	Net payout yield	Boudoukh et al. (2007)
	eqpo_me	Payout yield	Boudoukh et al. (2007)
	fcf_me	Free cash flow-to-price	Lakonishok et al. (1994)
	ival_me	Intrinsic value-to-market	Frankel and Lee (1998)
	netis_at	Net total issuance	Bradshaw et al. (2006)
	ni_me	Earnings-to-price	Basu (1983)
	ocf_me	Operating cash flow-to-market	Desai et al. (2004)
	sale_me	Sales-to-market	Barbee Jr et al. (1996)
Size	ami_126d	Amihud Measure	Amihud (2002)
	dolvol_126d	Dollar trading volume	Brennan et al. (1998)
	market_equity	Market Equity	Banz (1981)
	prc	Price per share	Miller and Scholes (1982)
	rd_me	R&D-to-market	Chan et al. (2001)
Profitability	dolvol_var_126d	Coefficient of variation for dollar trading volume	Chordia et al. (2001)
	ebit_bev	Return on net operating assets	Soliman (2008)
	ebit_sale	Profit margin	Soliman (2008)
	f_score	Pitroski F-score	Piotroski (2000)
	ni_be	Return on equity	Haugen and Baker (1996)
	niq_be	Quarterly return on equity	Hou et al. (2015)
	o_score	Ohlson O-score	Dichev (1998)
	ocf_at	Operating cash flow to assets	Bouchaud et al. (2019)
	ope_be	Operating profits-to-book equity	Fama and French (2015)
	ope_bel1	Operating profits-to-lagged book equity	Fama and French (2015)
	turnover_var_126d	Coefficient of variation for share turnover	Chordia et al. (2001)
Investment	aliq_at	Liquidity of book assets	Ortiz-Molina and Phillips (2014)
	at_gr1	Asset Growth	Cooper et al. (2008)
	be_gr1a	Change in common equity	Richardson et al. (2005)
	capx_gr1	CAPEX growth (1 year)	Xie (2001)
	capx_gr2	CAPEX growth (2 years)	Anderson and Garcia-Feijoo (2006)
	capx_gr3	CAPEX growth (3 years)	Anderson and Garcia-Feijoo (2006)
	coa_gr1a	Change in current operating assets	Richardson et al. (2005)
	col_gr1a	Change in current operating liabilities	Richardson et al. (2005)
	emp_gr1	Hiring rate	Belo et al. (2014)
	inv_gr1	Inventory growth	Belo and Lin (2012)
	inv_gr1a	Inventory change	Thomas and Zhang (2002)
	lnoa_gr1a	Change in long-term net operating assets	Fairfield et al. (2003)
	mispricing_mgmt	Mispricing factor: Management	Stambaugh and Yuan (2017)
	ncoa_gr1a	Change in noncurrent operating assets	Richardson et al. (2005)
	nncoa_gr1a	Change in net noncurrent operating assets	Richardson et al. (2005)
	noa_gr1a	Change in net operating assets	Hirshleifer et al. (2004)
	ppeinv_gr1a	Change PPE and Inventory	Lyandres et al. (2008)
	ret_60_12	Long-term reversal	De Bondt and Thaler (1985)
	sale_gr1	Sales Growth (1 year)	Lakonishok et al. (1994)
	sale_gr3	Sales Growth (3 years)	Lakonishok et al. (1994)
	saleq_gr1	Sales growth (1 quarter)	Fama and French (1992)
	seas_2_5na	Years 2-5 lagged returns, nonannual	Heston and Sadka (2008)
Momentum	prc_highprc_252d	Current price to high price over last year	George and Hwang (2004)
	resff3_6_1	Residual momentum t-6 to t-1	Blitz et al. (2011)
	resff3_12_1	Residual momentum t-12 to t-1	Blitz et al. (2011)
	ret_3_1	Price momentum t-3 to t-1	Jegadeesh and Titman (1993)
	ret_6_1	Price momentum t-6 to t-1	Jegadeesh and Titman (1993)
	ret_9_1	Price momentum t-9 to t-1	Jegadeesh and Titman (1993)
	ret_12_1	Price momentum t-12 to t-1	Jegadeesh and Titman (1993)
	seas_1_1na	Year 1-lagged return, nonannual	Heston and Sadka (2008)

FATEC-LASSO estimators, respectively. Additionally, the factor adjustment does not significantly change the non-zero frequency of the estimator, as shown in Table 6.

Table 5: Annual average in-sample and out-of-sample R^2 for the FATEN-LASSO, FATEC-LASSO, TEN-LASSO, and TEC-LASSO estimators over the five assets.

	In-sample R^2			
	Estimator			
	FATEN-LASSO	FATEC-LASSO	TEN-LASSO	TEC-LASSO
whole period	0.344	0.312	0.345	0.312
2013	0.322	0.318	0.324	0.320
2014	0.327	0.306	0.328	0.306
2015	0.322	0.256	0.325	0.256
2016	0.351	0.311	0.354	0.313
2017	0.295	0.278	0.296	0.278
2018	0.418	0.372	0.418	0.372
2019	0.372	0.344	0.374	0.342
	Out-of-sample R^2			
	Estimator			
	FATEN-LASSO	FATEC-LASSO	TEN-LASSO	TEC-LASSO
whole period	0.313	0.234	0.315	0.234
2014	0.274	0.213	0.276	0.211
2015	0.300	0.208	0.305	0.209
2016	0.314	0.216	0.315	0.218
2017	0.273	0.206	0.273	0.207
2018	0.387	0.317	0.393	0.318
2019	0.330	0.241	0.330	0.241

Table 6: Monthly average of the non-zero frequency of the FATEN-LASSO, FATEC-LASSO, TEN-LASSO, and TEC-LASSO estimators over 70 factors and 84 months for the five assets.

	Non-zero frequency			
	Estimator			
	FATEN-LASSO	FATEC-LASSO	TEN-LASSO	TEC-LASSO
AAPL	25.214	47.809	27.904	48.785
BRK.B	28.690	53.130	30.333	53.154
GM	26.583	54.785	26.797	54.785
GOOG	23.821	50.047	28.083	50.714
XOM	26.940	52.011	29.178	52.690

A.6 Empirical study for the FATEN-LASSO estimates

In this section, we explore the integrated coefficient estimates from the FATEN-LASSO procedure. Figure 4 plots the monthly integrated coefficients from the FATEN-LASSO estimator for the five assets and 70 factors. Figure 5 shows the non-zero frequency of the FATEN-LASSO estimator for the six groups, which consist of the Fama-French five-factors and momentum factor group, value group, size group, profitability group, investment group, and momentum group. Figures 4 and 5 present the time variation and sparsity of the coefficient process. We find that all groups, except for the size group, had non-zero integrated coefficients for most periods. The size group contains the smallest number of factors (five). This finding is in line with results from multi-factor models (Asness et al., 2013; Carhart, 1997; Fama and French, 1992, 2015) and factor zoo analysis (Jensen et al., 2023). To investigate the coefficient behaviors of the frequent factors, in Figure 6, we draw the integrated coefficient estimates for the three most frequent factors. AAPL has MKT, aliq_at (liquidity of book assets), and inv_gr1 (inventory growth); BRK.B has MKT, ope_be (operating profits-to-book equity), and ope_bel1 (operating profits-to-lagged book equity); GM has MKT, ebit_sale (profit margin), and rd_me (R&D-to-market); GOOG has MKT, aliq_at, and capx_gr1 (CAPEX growth (1 year)); and XOM has MKT, f_score (Pitroski F-score), and turnover_var_126d (coefficient of variation for share turnover). We observed that the MKT factor most frequently had non-zero integrated coefficients for all five assets. This result is also consistent with the results of multi-factor models.

A.7 Proof of Theorem 1

Without loss of generality, it suffices to show the statement for fixed i. For simplicity, we denote the true $\boldsymbol{\beta}(i\Delta_n)$, $\mathbf{B}(i\Delta_n)$, $\boldsymbol{\gamma}(i\Delta_n)$, and $\boldsymbol{\theta}(i\Delta_n)$ by $\boldsymbol{\beta}_0 = (\beta_{0,j})_{j=1,\ldots,p}$, $\mathbf{B}_0 = (B_{0,ij})_{1\le i\le p, 1\le j\le r}$, $\boldsymbol{\gamma}_0 = (\gamma_{0,j})_{j=1,\ldots,r}$, and $\boldsymbol{\theta}_0 = (\theta_{0,j})_{j=1,\ldots,p+r}$, respectively. Let

$$\mathcal{Y}_i^c = \begin{pmatrix} \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right)\Delta_{i+l+1}^n Y^c \\ \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right)\Delta_{i+l+2}^n Y^c \\ \vdots \\ \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right)\Delta_{i+l+k_2-k_1+1}^n Y^c \end{pmatrix}, \quad \mathcal{X}_i^c = \begin{pmatrix} \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right)\Delta_{i+l+1}^n \mathbf{X}^{c\top} \\ \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right)\Delta_{i+l+2}^n \mathbf{X}^{c\top} \\ \vdots \\ \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right)\Delta_{i+l+k_2-k_1+1}^n \mathbf{X}^{c\top} \end{pmatrix},$$

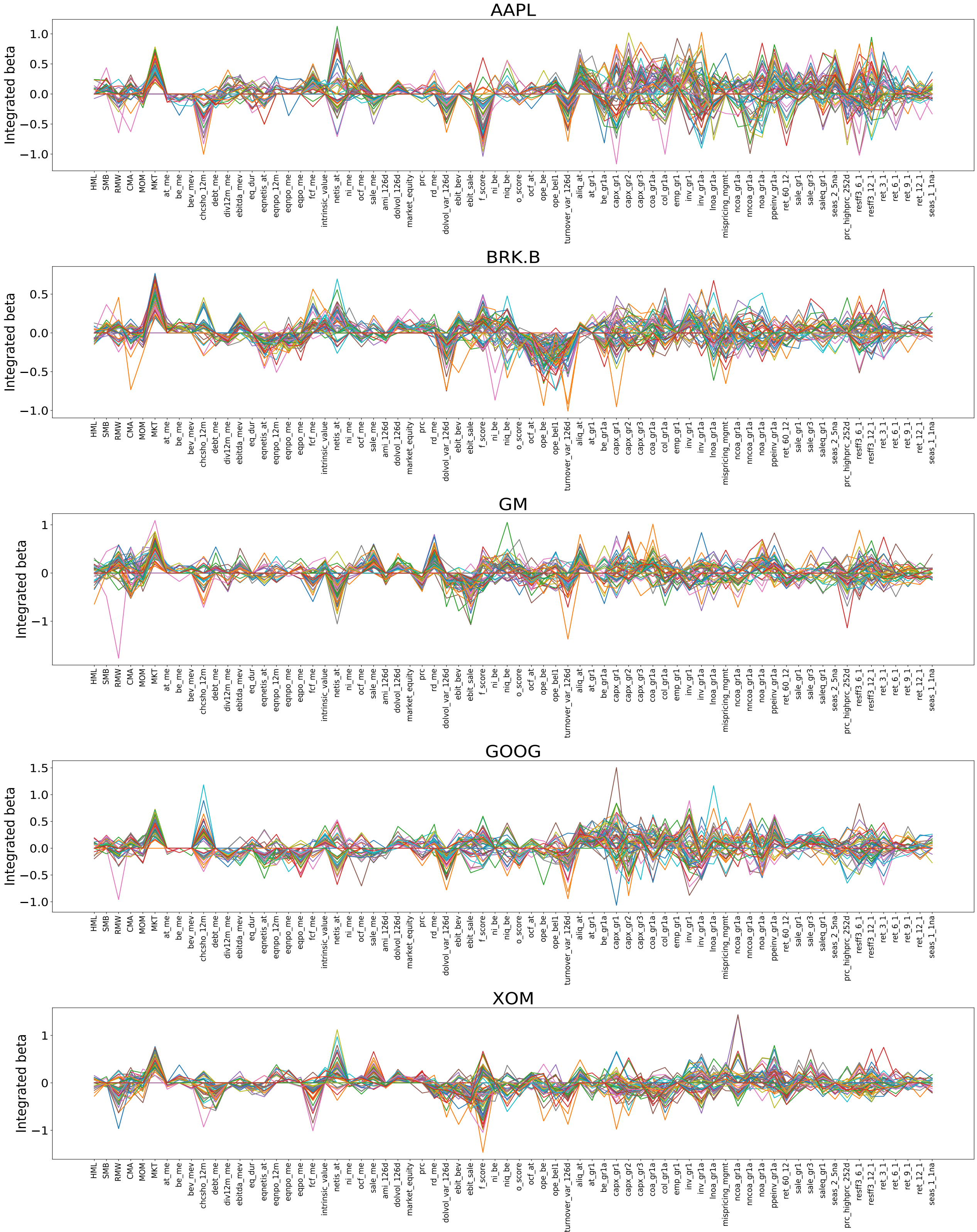

Figure 4: Monthly integrated coefficient estimates from the FATEN-LASSO procedure for the five assets and 70 factors. Each line indicates the 70 integrated coefficient estimates from each month.

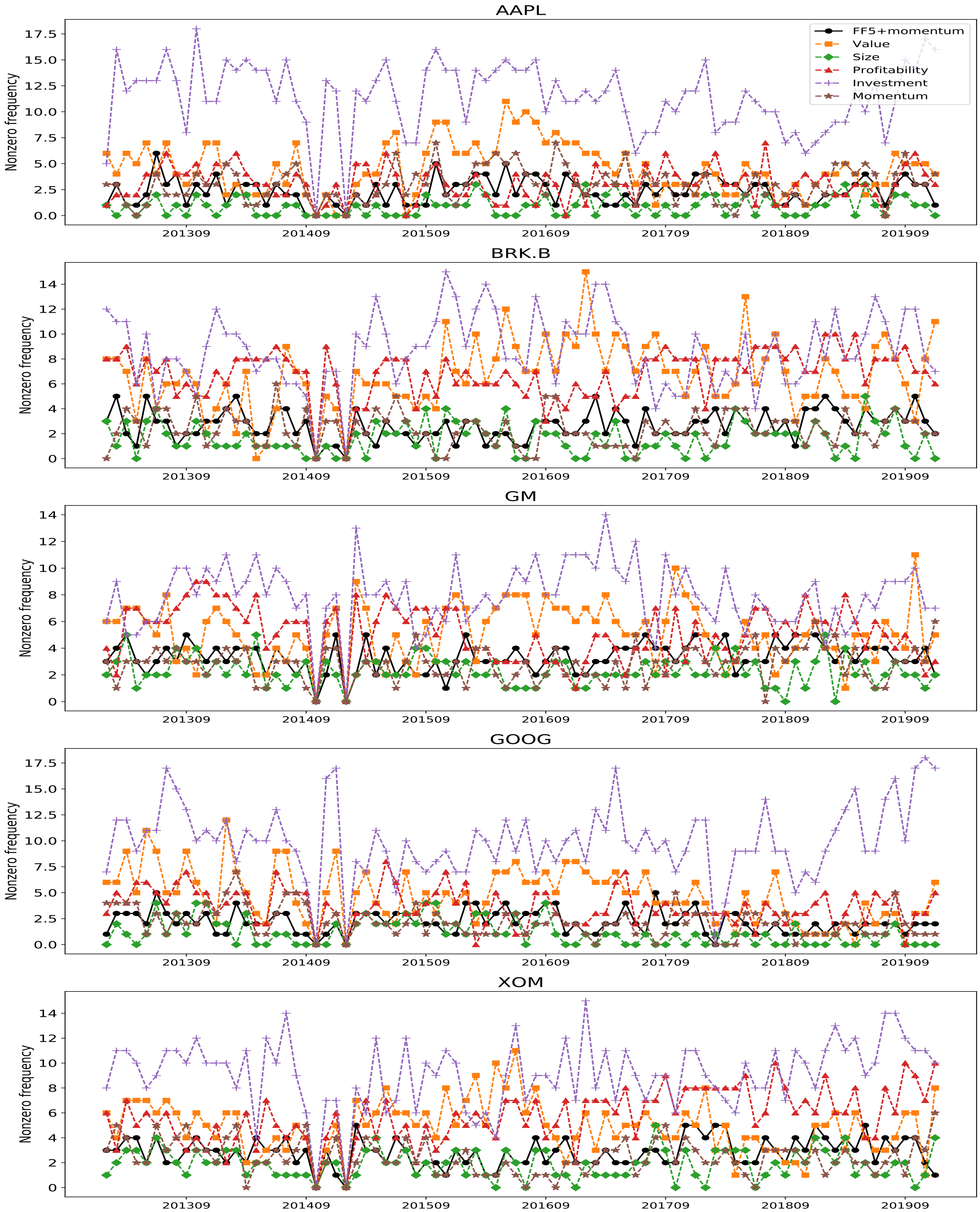

Figure 5: Nonzero frequency of the monthly integrated coefficient estimates from the FATEN-LASSO procedure for the five assets and six groups, which consist of the Fama-French five-factors and momentum factor group, value group, size group, profitability group, investment group, and momentum group.

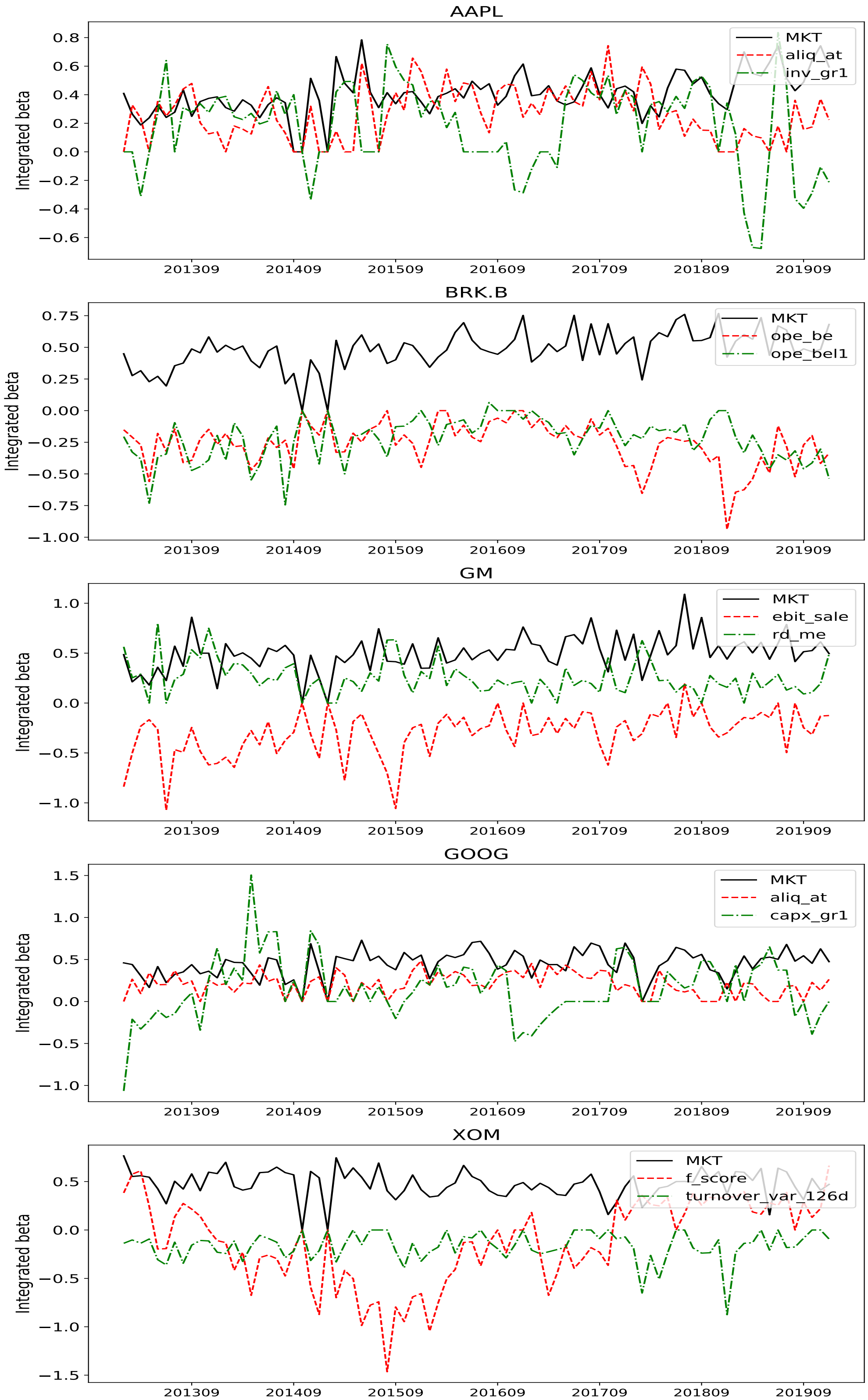

Figure 6: Monthly integrated coefficient estimates from the FATEN-LASSO procedure for the three most frequent factors among the 70 factors for the five assets.

$$\widetilde{\mathcal{X}}_i = \begin{pmatrix} \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right) \int_{(i+l)\Delta_n}^{(i+l+1)\Delta_n} (\boldsymbol{\beta}(t) - \boldsymbol{\beta}_0)^\top d\mathbf{X}^c(t) \\ \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right) \int_{(i+l+1)\Delta_n}^{(i+l+2)\Delta_n} (\boldsymbol{\beta}(t) - \boldsymbol{\beta}_0)^\top d\mathbf{X}^c(t) \\ \vdots \\ \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right) \int_{(i+l+k_2-k_1)\Delta_n}^{(i+l+k_2-k_1+1)\Delta_n} (\boldsymbol{\beta}(t) - \boldsymbol{\beta}_0)^\top d\mathbf{X}^c(t) \end{pmatrix},$$

$$\mathcal{Z}_i = \begin{pmatrix} \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right) \Delta_{i+l+1}^n Z \\ \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right) \Delta_{i+l+2}^n Z \\ \vdots \\ \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right) \Delta_{i+l+k_2-k_1+1}^n Z \end{pmatrix},$$

$$\widetilde{\mathbf{F}}_i = \begin{pmatrix} \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right) \left[\int_{(i+l)\Delta_n}^{(i+l+1)\Delta_n} (\mathbf{B}(t) - \mathbf{B}_0) d\mathbf{f}(t)\right]^\top \\ \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right) \left[\int_{(i+l+1)\Delta_n}^{(i+l+2)\Delta_n} (\mathbf{B}(t) - \mathbf{B}_0) d\mathbf{f}(t)\right]^\top \\ \vdots \\ \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right) \left[\int_{(i+l+k_2-k_1)\Delta_n}^{(i+l+k_2-k_1+1)\Delta_n} (\mathbf{B}(t) - \mathbf{B}_0) d\mathbf{f}(t)\right]^\top \end{pmatrix}.$$

Then, we have

$$\mathcal{Y}_i^c = \mathcal{X}_i^c \boldsymbol{\beta}_0 + \mathcal{Z}_i + \widetilde{\mathcal{X}}_i \quad \text{and} \quad \mathcal{X}_i^c = \mathbf{F}_i \mathbf{B}_0^\top + \mathbf{U}_i + \widetilde{\mathbf{F}}_i.$$

Also, let

$$\mathcal{E}_i^Y = \begin{pmatrix} \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right) \Delta_{i+l+1}^n \epsilon^Y \\ \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right) \Delta_{i+l+2}^n \epsilon^Y \\ \vdots \\ \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right) \Delta_{i+l+k_2-k_1+1}^n \epsilon^Y \end{pmatrix}, \quad \mathcal{E}_i^X = \begin{pmatrix} \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right) \Delta_{i+l+1}^n \left(\boldsymbol{\varepsilon}^X\right)^\top \\ \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right) \Delta_{i+l+2}^n \left(\boldsymbol{\varepsilon}^X\right)^\top \\ \vdots \\ \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right) \Delta_{i+l+k_2-k_1+1}^n \left(\boldsymbol{\varepsilon}^X\right)^\top \end{pmatrix},$$

$$
\mathcal{Y}_i' = \begin{pmatrix} \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right) \left\{\Delta_{i+l+1}^n Y^c + \Delta_{i+l+1}^n \epsilon^Y\right\} \mathbf{1}\left(|\Delta_i^n \widehat{Y}| \le w_n\right) \\ \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right) \left\{\Delta_{i+l+2}^n Y^c + \Delta_{i+l+2}^n \epsilon^Y\right\} \mathbf{1}\left(|\Delta_{i+1}^n \widehat{Y}| \le w_n\right) \\ \vdots \\ \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right) \left\{\Delta_{i+l+k_2-k_1+1}^n Y^c + \Delta_{i+l+k_2-k_1+1}^n \epsilon^Y\right\} \mathbf{1}\left(|\Delta_{i+k_2-k_1}^n \widehat{Y}| \le w_n\right) \end{pmatrix},
$$

$$
\mathcal{X}_i' = \begin{pmatrix} \left(\Delta_i^n \widetilde{\mathbf{X}}^{\text{trunc}}\right)^\top \\ \left(\Delta_{i+1}^n \widetilde{\mathbf{X}}^{\text{trunc}}\right)^\top \\ \vdots \\ \left(\Delta_{i+k_2-k_1}^n \widetilde{\mathbf{X}}^{\text{trunc}}\right)^\top \end{pmatrix},
$$

where

$$
\Delta_i^n \widetilde{\mathbf{X}}^{\text{trunc}} = \left(\sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right) \left\{\Delta_{i+l+1}^n X_j^c + \Delta_{i+l+1}^n \epsilon_j^X\right\} \mathbf{1}\left(|\Delta_i^n \widehat{X}_j| \le v_{j,n}\right) \right)_{j=1,\dots,p}.
$$

Proposition 2. *Under the assumptions in Theorem 1, with the probability at least $1 - p^{-3-a}$, we have*

$$
\left\| \frac{n}{\phi k_1 k_2} \mathbf{F}_i^\top \mathbf{F}_i - \mathbf{\Sigma}_{0,f}(i\Delta_n) \right\|_{\max} \le C n^{-1/8} \sqrt{\log p}, \tag{A.3}
$$

$$
\left\| \frac{n}{\phi k_1 k_2} \mathbf{U}_i^\top \mathbf{U}_i - \mathbf{\Sigma}_{0,u}(i\Delta_n) \right\|_{\max} \le C n^{-1/8} \sqrt{\log p}, \tag{A.4}
$$

$$
\left\| \frac{n}{\phi k_1 k_2} \left(\mathcal{E}_i^X\right)^\top \mathcal{E}_i^X - \frac{n\zeta}{\phi k_1} \mathbf{V}^X \right\|_{\max} \le C n^{-1/8} \sqrt{\log p}, \tag{A.5}
$$

$$
\left\| \frac{n}{\phi k_1 k_2} \mathbf{F}_i^\top \mathbf{U}_i \right\|_{\max} \le C n^{-1/8} \sqrt{\log p}, \tag{A.6}
$$

$$
\left\| \frac{n}{\phi k_1 k_2} \mathbf{F}_i^\top \mathcal{E}_i^X \right\|_{\max} \le C n^{-1/8} \sqrt{\log p}, \tag{A.7}
$$

$$
\left\| \frac{n}{\phi k_1 k_2} \mathbf{U}_i^\top \mathcal{E}_i^X \right\|_{\max} \le C n^{-1/8} \sqrt{\log p}, \tag{A.8}
$$

$$
\left\| \frac{n}{\phi k_1 k_2} \mathcal{X}_i^\top \mathcal{X}_i - \frac{n}{\phi k_1 k_2} \mathbf{B}_0 \mathbf{F}_i^\top \mathbf{F}_i \mathbf{B}_0^\top - \mathbf{\Sigma}_{0,u}(i\Delta_n) - \frac{n\zeta}{\phi k_1} \mathbf{V}^X \right\|_{\max} \le C n^{-1/8} (\log p)^{3/2}, \tag{A.9}
$$

$$\left\| \frac{n\zeta}{\phi k_1} \left(\widehat{\mathbf{V}}^X - \mathbf{V}^X \right) \right\|_{\max} \leq C n^{-1/2} \sqrt{\log p}. \tag{A.10}$$

Proof of Proposition 2. Consider (A.3). Without loss of generality, we assume that $k_2 = k_1(L+1)-1$ for some $L \in \mathbb{N}$. Let $\Delta_i^n \bar{f}_j = \sum_{l=0}^{k_1-1} g\left(\frac{l}{k_1}\right) \Delta_{i+l+1}^n f_j$ and $\boldsymbol{\Sigma}_{0,f}(t) = (\Sigma_{0,f,jm}(t))_{j,m=1,\dots,p}$. Define

$$T_{jm}(i\Delta_n) = \sum_{l=0}^{k_1-1} \left\{ g\left(\frac{l}{k_1}\right) \right\}^2 \mathrm{E}\left\{ \int_{(i+l)\Delta_n}^{(i+l+1)\Delta_n} \Sigma_{0,f,jm}(t) dt \middle| i\Delta_n \right\}.$$

Then, we have

$$\mathrm{E}\left[\Delta_i^n \bar{f}_j \Delta_i^n \bar{f}_m \middle| \mathcal{F}_{i\Delta_n} \right] = T_{jm}(i\Delta_n) \text{ a.s.}$$

Also, we have

$$\begin{aligned}
&\left| \frac{n}{\phi k_1 k_2} \sum_{k=0}^{k_2-k_1} \Delta_{i+k}^n \bar{f}_j \Delta_{i+k}^n \bar{f}_m - \int_{i\Delta_n}^{(i+k_2)\Delta_n} \Sigma_{0,f,jm}(t) dt / (k_2 \Delta_n) \right| \\
&\leq \left| \frac{n}{\phi k_1 k_2} \sum_{k=0}^{k_2-k_1} \left[\Delta_{i+k}^n \bar{f}_j \Delta_{i+k}^n \bar{f}_m - T_{jm}((i+k)\Delta_n) \right] \right| \\
&\quad + \left| \frac{n}{\phi k_1 k_2} \sum_{k=0}^{k_2-k_1} T_{jm}((i+k)\Delta_n) - \int_{i\Delta_n}^{(i+k_2)\Delta_n} \Sigma_{0,f,jm}(t) dt / (k_2 \Delta_n) \right|.
\end{aligned}$$

Consider the first term. We have, for $x = 0, \dots, k_1 - 1$,

$$\begin{aligned}
&\left| \frac{n}{\phi k_1 k_2} \sum_{k=0}^{k_2-k_1} \left[\Delta_{i+k}^n \bar{f}_j \Delta_{i+k}^n \bar{f}_m - T_{jm}((i+k)\Delta_n) \right] \right| \\
&\leq \sum_{x=0}^{k_1-1} \left| \frac{n}{\phi k_1 k_2} \sum_{k=0}^{L} \left[\Delta_{i+k_1 k+x}^n \bar{f}_j \Delta_{i+k_1 k+x}^n \bar{f}_m - T_{jm}((i+k_1 k + x)\Delta_n) \right] \right|.
\end{aligned}$$

Note that $\Delta_i^n \bar{f}_j$ has the sub-Gaussian tail with the order of $n^{-1/4}$. Thus, by Bernstein's inequality for martingales, we have

$$\Pr\left\{ \left| \frac{n}{\phi k_1 k_2} \sum_{k=0}^{L} \left[\Delta_{i+k_1 k+x}^n \bar{f}_j \Delta_{i+k_1 k+x}^n \bar{f}_m - T_{jm}((i+k_1 k+x)\Delta_n) \right] \right| \leq C n^{-5/8} \sqrt{\log p} \right\} \geq 1 - p^{-9-a}.$$

Then, by Assumption 1(g), we have, for large n,

$$\Pr\left\{\left|\frac{n}{\phi k_1 k_2}\sum_{k=0}^{k_2-k_1}\left[\Delta_{i+k}^n \bar{f}_j \Delta_{i+k}^n \bar{f}_m - T_{jm}((i+k)\Delta_n)\right]\right| \leq C n^{-1/8}\sqrt{\log p}\right\} \geq 1 - p^{-6-a}.$$

Consider the second term. By the boundedness of $\Sigma_{0,f,jm}(t)$, we have

$$\begin{aligned}
&\left|\frac{n}{\phi k_1 k_2}\sum_{k=0}^{k_2-k_1} T_{jm}((i+k)\Delta_n) - \int_{i\Delta_n}^{(i+k_2)\Delta_n}\Sigma_{0,f,jm}(t)dt/(k_2\Delta_n)\right| \\
&= \left|\frac{n}{\phi k_1 k_2}\sum_{l=0}^{k_1-1}\left\{g\left(\frac{l}{k_1}\right)\right\}^2\left(\sum_{k=0}^{k_2-k_1}\left[\mathrm{E}\left\{\int_{(i+k+l)\Delta_n}^{(i+k+l+1)\Delta_n}\Sigma_{0,f,jm}(t)dt\middle|(i+k)\,\Delta_n\right\}\right]\right.\right. \\
&\qquad\qquad\left.\left. - \int_{i\Delta_n}^{(i+k_2)\Delta_n}\Sigma_{0,f,jm}(t)dt\right)\right| \\
&\leq \left|\frac{n}{\phi k_1 k_2}\sum_{l=0}^{k_1-1}\left\{g\left(\frac{l}{k_1}\right)\right\}^2\sum_{k=0}^{k_2-k_1}\left[\mathrm{E}\left\{\int_{(i+k+l)\Delta_n}^{(i+k+l+1)\Delta_n}\Sigma_{0,f,jm}(t)dt\middle|(i+k)\,\Delta_n\right\}\right.\right. \\
&\qquad\qquad\left.\left. - \int_{(i+k+l)\Delta_n}^{(i+k+l+1)\Delta_n}\Sigma_{0,f,jm}(t)dt\right]\right| + Cn^{-1/4}.
\end{aligned}$$

Note that $\sum_{k=0}^{k_2-k_1}\left[\mathrm{E}\left\{\int_{(i+k+l)\Delta_n}^{(i+k+l+1)\Delta_n}\Sigma_{0,f,jm}(t)dt\middle|(i+k)\,\Delta_n\right\} - \int_{(i+k+l)\Delta_n}^{(i+k+l+1)\Delta_n}\Sigma_{0,f,jm}(t)dt\right]$ is the sum of $l+1$ martingales. Hence, using the Azuma-Hoeffding inequality for each martingale, we can show for all $0 \leq l \leq k_1 - 1$,

$$\begin{aligned}
&\Pr\left\{\left|\sum_{k=0}^{k_2-k_1}\left[\mathrm{E}\left\{\int_{(i+k+l)\Delta_n}^{(i+k+l+1)\Delta_n}\Sigma_{0,f,jm}(t)dt\middle|(i+k)\,\Delta_n\right\} - \int_{(i+k+l)\Delta_n}^{(i+k+l+1)\Delta_n}\Sigma_{0,f,jm}(t)dt\right]\right|\right. \\
&\qquad\qquad\left. \leq Cn^{-3/8}\sqrt{\log p}\right\} \\
&\geq 1 - p^{-9-a},
\end{aligned}$$

which implies

$$\Pr\left\{\left|\frac{n}{\phi k_1 k_2}\sum_{k=0}^{k_2-k_1} T_{jm}((i+k)\Delta_n) - \int_{i\Delta_n}^{(i+k_2)\Delta_n}\Sigma_{0,f,jm}(t)dt/(k_2\Delta_n)\right| \leq Cn^{-1/8}\sqrt{\log p}\right\} \geq 1 - p^{-6-a}$$

for large n. Thus, we have

$$\Pr\left\{\left\|\frac{n}{\phi k_1 k_2}\mathbf{F}_i^\top \mathbf{F}_i - \int_{i\Delta_n}^{(i+k_2)\Delta_n} \mathbf{\Sigma}_{0,f}(t)dt/(k_2\Delta_n)\right\|_{\max} \leq Cn^{-1/8}\sqrt{\log p}\right\} \geq 1-2p^{-4-a}.$$

Then, by Assumption 1(f), we have

$$\Pr\left\{\left\|\frac{n}{\phi k_1 k_2}\mathbf{F}_i^\top \mathbf{F}_i - \mathbf{\Sigma}_{0,f}(i\Delta_n)\right\|_{\max} \leq Cn^{-1/8}\sqrt{\log p}\right\} \geq 1-2p^{-4-a}. \quad \text{(A.11)}$$

Similarly, we can show that (A.4)–(A.8) hold with the probability at least $1-p^{-4-a}$.

Now, it is enough to show that (A.9) and (A.10) hold with the probability at least $1-Cp^{-4-a}$. Consider (A.9). We have

$$\begin{aligned}
&\left\|\frac{n}{\phi k_1 k_2}\mathcal{X}_i^\top \mathcal{X}_i - \frac{n}{\phi k_1 k_2}\mathbf{B}_0\mathbf{F}_i^\top \mathbf{F}_i \mathbf{B}_0^\top - \mathbf{\Sigma}_{0,u}(i\Delta_n) - \frac{n\zeta}{\phi k_1}\mathbf{V}^X\right\|_{\max} \\
&\leq \frac{n}{\phi k_1 k_2}\left\|\mathcal{X}_i^\top \mathcal{X}_i - \left(\mathcal{X}_i^c + \mathcal{E}_i^X\right)^\top \left(\mathcal{X}_i^c + \mathcal{E}_i^X\right)\right\|_{\max} \\
&\quad + \frac{n}{\phi k_1 k_2}\left\|\left(\mathcal{X}_i^c + \mathcal{E}_i^X\right)^\top \left(\mathcal{X}_i^c + \mathcal{E}_i^X\right) - \left(\mathbf{F}_i\mathbf{B}_0^\top + \mathbf{U}_i + \mathcal{E}_i^X\right)^\top \left(\mathbf{F}_i\mathbf{B}_0^\top + \mathbf{U}_i + \mathcal{E}_i^X\right)\right\|_{\max} \\
&\quad + \frac{n}{\phi k_1 k_2}\left\|\left(\mathbf{F}_i\mathbf{B}_0^\top\right)^\top \left(\mathbf{U}_i + \mathcal{E}_i^X\right) + \mathbf{U}_i^\top \left(\mathbf{F}_i\mathbf{B}_0^\top + \mathcal{E}_i^X\right) + \left(\mathcal{E}_i^X\right)^\top \left(\mathbf{F}_i\mathbf{B}_0^\top + \mathbf{U}_i\right)\right\|_{\max} \\
&\quad + \left\|\frac{n}{\phi k_1 k_2}\mathbf{U}_i^\top \mathbf{U}_i - \mathbf{\Sigma}_{0,u}(i\Delta_n)\right\|_{\max} + \left\|\frac{n}{\phi k_1 k_2}\left(\mathcal{E}_i^X\right)^\top \mathcal{E}_i^X - \frac{n\zeta}{\phi k_1}\mathbf{V}^X\right\|_{\max} \\
&= (I) + (II) + (III) + (IV) + (V).
\end{aligned} \quad \text{(A.12)}$$

For some large constant $C > 0$, define

$$\begin{aligned}
Q_1 &= \left\{\max_i \|\mathcal{Y}_i^c\|_\infty \leq Cs_p\sqrt{\frac{k_1 \log p}{n}}\right\} \cap \left\{\max_i \left\|\mathcal{E}_i^Y\right\|_\infty \leq C\sqrt{\frac{\log p}{k_1}}\right\}, \\
Q_2 &= \left\{\max_i \|\mathcal{X}_i^c\|_{\max} \leq C\sqrt{\frac{k_1 \log p}{n}}\right\} \cap \left\{\max_i \|\mathbf{F}_i\|_{\max} \leq C\sqrt{\frac{k_1 \log p}{n}}\right\} \\
&\quad \cap \left\{\max_i \|\mathbf{U}_i\|_{\max} \leq C\sqrt{\frac{k_1 \log p}{n}}\right\} \cap \left\{\max_i \|\mathcal{Z}_i\|_{\max} \leq C\sqrt{\frac{k_1 \log p}{n}}\right\} \\
&\quad \cap \left\{\max_i \left\|\mathcal{E}_i^X\right\|_{\max} \leq C\sqrt{\frac{\log p}{k_1}}\right\},
\end{aligned}$$

$$
\begin{aligned}
Q_3 &= \left\{ \int_0^1 d\Lambda^Y(t) \leq C \log p \right\} \cap \left\{ \max_j \int_0^1 d\Lambda_j(t) \leq C \log p \right\}, \\
Q_4 &= \left\{ \sum_{i=0}^{n-k_1} \mathbf{1}\left(|\Delta_i^n \widehat{Y}| > w_n \right) \leq C k_1 \log p \right\} \cap \left\{ \max_j \sum_{i=0}^{n-k_1} \mathbf{1}\left(|\Delta_i^n \widehat{X}_j| > v_{j,n} \right) \leq C k_1 \log p \right\}, \\
Q_5 &= \left\{ \max_i \left\| \widetilde{\mathcal{X}}_i \right\|_\infty \leq C s_p n^{-3/8} \log p \right\} \cap \left\{ \max_i \left\| \widetilde{\mathbf{F}}_i \right\|_{\max} \leq C n^{-3/8} \log p \right\}.
\end{aligned}
$$

We note that the variables related to the dependent process are also considered to avoid repetition in the proof. From Assumption 1(a),(b), we can show

$$
\Pr\left(Q_1 \cap Q_2 \right) \geq 1 - p^{-4-a}.
$$

By the boundedness of the intensity process, we have

$$
\Pr\left(Q_3 \right) \geq 1 - p^{-4-a}.
$$

Under $Q_1 \cap Q_2 \cap Q_3$, we have, for large n,

$$
\sum_{i=0}^{n-k_1} \mathbf{1}\left(|\Delta_i^n \widehat{Y}| > w_n \right) \leq C k_1 \log p \quad \text{and} \quad \max_j \sum_{i=0}^{n-k_1} \mathbf{1}\left(|\Delta_i^n \widehat{X}_j| > v_{j,n} \right) \leq C k_1 \log p.
$$

Consider Q_5. By Assumption 1(a), $\sum_{j=1}^p |\beta_j(t) - \beta_{0,j}|$ has the sub-Gaussian tail. Thus, we have

$$
\begin{aligned}
&\Pr\left\{ \sup_{t \in [i\Delta_n, (i+k_2)\Delta_n]} \sum_{j=1}^p |\beta_j(t) - \beta_{0,j}| \geq C s_p n^{-1/8} \sqrt{\log p} \right\} \\
&\leq C \Pr\left\{ \sum_{j=1}^p |\beta_j((i+k_2)\Delta_n) - \beta_{0,j}| \geq C s_p n^{-1/8} \sqrt{\log p} \right\} \\
&\leq p^{-10-a}.
\end{aligned}
$$

Let

$$
E = \left\{ \max_i \sup_{t \in [i\Delta_n, (i+k_2)\Delta_n]} \sum_{j=1}^p |\beta_j(t) - \beta_{0,j}| \geq C s_p n^{-1/8} \sqrt{\log p} \right\}.
$$

Since

$$\Pr\left\{E\right\} \le p^{-5-a}$$

for large n, we have

$$\begin{aligned}
&\Pr\left\{\max_i \left\|\widetilde{\mathcal{X}}_i\right\|_\infty \ge C s_p n^{-3/8} \log p\right\} \\
&\le \Pr\left\{\max_i \left\|\widetilde{\mathcal{X}}_i\right\|_\infty \ge C s_p n^{-3/8} \log p, E^c\right\} + p^{-5-a} \\
&\le 2p^{-5-a}.
\end{aligned}$$

Similarly, we can show

$$\Pr\left\{\max_i \left\|\widetilde{\mathbf{F}}_i\right\|_{\max} \ge C n^{-3/8} \log p\right\} \le 2p^{-5-a},$$

which implies

$$\Pr\left(Q_5\right) \ge 1 - p^{-4-a}.$$

Thus, we have

$$\Pr\left(Q_1 \cap Q_2 \cap Q_3 \cap Q_4 \cap Q_5\right) \ge 1 - 3p^{-4-a}. \tag{A.13}$$

From (A.13), we have, with the probability at least $1 - 3p^{-4-a}$,

$$\begin{aligned}
(I) &\le \frac{n}{\phi k_1 k_2} \left\|\mathcal{X}_i^\top \mathcal{X}_i - \mathcal{X}_i'^\top \mathcal{X}_i'\right\|_{\max} + \frac{n}{\phi k_1 k_2} \left\|\mathcal{X}_i'^\top \mathcal{X}_i' - \left(\mathcal{X}_i^c + \mathcal{E}_i^X\right)^\top \left(\mathcal{X}_i^c + \mathcal{E}_i^X\right)\right\|_{\max} \\
&\le C n^{-1/4} \left(\log p\right)^2, \qquad \text{(A.14)} \\
(II) &\le \frac{2n}{\phi k_1 k_2} \left\|\left(\mathbf{F}_i \mathbf{B}_0^\top + \mathbf{U}_i + \mathcal{E}_i^X\right)^\top \widetilde{\mathbf{F}}_i\right\|_{\max} + \frac{n}{\phi k_1 k_2} \left\|\widetilde{\mathbf{F}}_i^\top \widetilde{\mathbf{F}}_i\right\|_{\max} \\
&\le C n^{-1/8} \left(\log p\right)^{3/2}. \qquad \text{(A.15)}
\end{aligned}$$

For (III), note that the elements of $\mathbf{F}_i$, $\mathbf{U}_i$, and $\mathcal{E}_i^X$ have sub-Gaussian tails. Thus, from Bernstein's inequality for martingales, we have

$$\Pr\left\{(III) \leq Cn^{-1/8}\sqrt{\log p}\right\} \geq 1 - p^{-4-a}. \tag{A.16}$$

Consider (IV) and (V). Similar to the proofs of (A.11), we can show

$$\Pr\left\{(IV) + (V) \leq Cn^{-1/8}\sqrt{\log p}\right\} \geq 1 - p^{-4-a}. \tag{A.17}$$

Combining (A.12) and (A.14)–(A.17), we have, with the probability at least $1 - 5p^{-4-a}$,

$$\left\| \frac{n}{\phi k_1 k_2} \mathcal{X}_i^\top \mathcal{X}_i - \frac{n}{\phi k_1 k_2} \mathbf{B}_0 \mathbf{F}_i^\top \mathbf{F}_i \mathbf{B}_0^\top - \mathbf{\Sigma}_{0,u}(i\Delta_n) - \frac{n\zeta}{\phi k_1} \mathbf{V}^X \right\|_{\max} \leq Cn^{-1/8} (\log p)^{3/2}. \tag{A.18}$$

Similarly, we can show

$$\Pr\left\{ \left\| \frac{n\zeta}{\phi k_1} \left(\widehat{\mathbf{V}}^X - \mathbf{V}^X \right) \right\|_{\max} \leq Cn^{-1/2}\sqrt{\log p} \right\} \geq 1 - p^{-4-a}. \tag{A.19}$$

■

Proposition 3. *Under the assumptions in Theorem 1, there exists a r by r matrix $\mathbf{H}_i$ such that with the probability at least $1 - p^{-2-a}$,*

$$\left\| \widehat{\mathbf{B}}_{i\Delta_n} - \mathbf{B}_0 \mathbf{H}_i \right\|_{\max} \leq C \left\{ n^{-1/8} (\log p)^{3/2} + p^{-1/2} \right\}, \tag{A.20}$$

$$\|\mathbf{H}_i\|_2 \leq C, \quad \|\mathbf{H}_i^{-1}\|_2 \leq C, \tag{A.21}$$

$$\left\| \widehat{\mathbf{F}}_i - \mathbf{F}_i \left(\mathbf{H}_i^\top \right)^{-1} \right\|_2 \leq C \left\{ (\log p)^{3/2} + p^{-1/2} n^{1/8} \right\}. \tag{A.22}$$

Proof of Proposition 3. For $j = 1, \ldots, r$, let $\lambda_{i,j}$ be the j-th largest eigenvalue of $\frac{n}{\phi k_1 k_2} \mathcal{X}_i^\top \mathcal{X}_i$

and $\xi_{i,j}$ be its corresponding eigenvector. Define

$$\mathbf{\Lambda}_i = \mathrm{Diag}\left(\lambda_{i,1}, \ldots, \lambda_{i,r}\right) \quad \text{and} \quad \mathbf{H}_i = \frac{n}{\phi k_1 k_2} \mathbf{F}_i^\top \mathbf{F}_i \mathbf{B}_0^\top \widehat{\mathbf{B}}_{i\Delta_n} \mathbf{\Lambda}_i^{-1}.$$

By (3.2), we have

$$\widehat{\mathbf{B}}_{i\Delta_n} = \sqrt{p}\left(\widehat{\xi}_{i,1}, \ldots, \widehat{\xi}_{i,r}\right) \quad \text{and} \quad \widehat{\mathbf{F}}_i = p^{-1} \mathcal{X}_i \widehat{\mathbf{B}}_{i\Delta_n}.$$

Then, we have

$$\frac{n}{\phi k_1 k_2} \mathcal{X}_i^\top \mathcal{X}_i \widehat{\mathbf{B}}_{i\Delta_n} = \widehat{\mathbf{B}}_{i\Delta_n} \mathbf{\Lambda}_i, \quad \widehat{\mathbf{B}}_{i\Delta_n}^\top \widehat{\mathbf{B}}_{i\Delta_n} = p\mathbf{I}_r, \quad \text{and} \quad \widehat{\mathbf{F}}_i^\top \widehat{\mathbf{F}}_i = \frac{\phi k_1 k_2}{np} \mathbf{\Lambda}_i.$$

We first investigate $\mathbf{\Lambda}_i$. By the Weyl's theorem and (A.9), for $1 \leq j \leq r$, we have, with the probability at least $1 - p^{-3-a}$,

$$\begin{aligned}
&\left|\lambda_{i,j} - \lambda_j \left(\frac{n}{\phi k_1 k_2} \mathbf{B}_0 \mathbf{F}_i^\top \mathbf{F}_i \mathbf{B}_0^\top\right)\right| \\
&\leq \left\|\frac{n}{\phi k_1 k_2} \mathcal{X}_i^\top \mathcal{X}_i - \frac{n}{\phi k_1 k_2} \mathbf{B}_0 \mathbf{F}_i^\top \mathbf{F}_i \mathbf{B}_0^\top\right\|_2 \\
&\leq \left\|\frac{n}{\phi k_1 k_2} \mathcal{X}_i^\top \mathcal{X}_i - \frac{n}{\phi k_1 k_2} \mathbf{B}_0 \mathbf{F}_i^\top \mathbf{F}_i \mathbf{B}_0^\top - \mathbf{\Sigma}_{0,u}(i\Delta_n) - \frac{n\zeta}{\phi k_1} \mathbf{V}^X\right\|_2 + \left\|\mathbf{\Sigma}_{0,u}(i\Delta_n) + \frac{n\zeta}{\phi k_1} \mathbf{V}^X\right\|_2 \\
&\leq Cpn^{-1/8} \left(\log p\right)^{3/2} + \left\|\mathbf{\Sigma}_{0,u}(i\Delta_n)\right\|_1 + \left\|\frac{n\zeta}{\phi k_1} \mathbf{V}^X\right\|_1 \\
&\leq Cpn^{-1/8} \left(\log p\right)^{3/2} + C,
\end{aligned}$$

where the last inequality is due to Assumption 1(d). Also, by the Weyl's theorem, for $1 \leq j \leq r$ and large n, we have, with the probability at least $1 - p^{-3-a}$,

$$\begin{aligned}
&\left|p^{-1} \lambda_j \left(\frac{n}{\phi k_1 k_2} \sqrt{\mathbf{F}_i^\top \mathbf{F}_i} \mathbf{B}_0^\top \mathbf{B}_0 \sqrt{\mathbf{F}_i^\top \mathbf{F}_i}\right) - \lambda_j \left(\mathbf{\Sigma}_{0,f}(i\Delta_n)\right)\right| \\
&\leq \left|p^{-1} \lambda_j \left(\frac{n}{\phi k_1 k_2} \sqrt{\mathbf{F}_i^\top \mathbf{F}_i} \mathbf{B}_0^\top \mathbf{B}_0 \sqrt{\mathbf{F}_i^\top \mathbf{F}_i}\right) - \lambda_j \left(\frac{n}{\phi k_1 k_2} \mathbf{F}_i^\top \mathbf{F}_i\right)\right| \\
&\quad + \left|\lambda_j \left(\frac{n}{\phi k_1 k_2} \mathbf{F}_i^\top \mathbf{F}_i\right) - \lambda_j \left(\mathbf{\Sigma}_{0,f}(i\Delta_n)\right)\right|
\end{aligned}$$

$$
\begin{aligned}
&\leq \left\| \frac{n}{\phi k_1 k_2} \mathbf{F}_i^\top \mathbf{F}_i \right\|_2 \times \left\| p^{-1} \mathbf{B}_0^\top \mathbf{B}_0 - \mathbf{I}_r \right\|_2 + \left\| \frac{n}{\phi k_1 k_2} \mathbf{F}_i^\top \mathbf{F}_i - \mathbf{\Sigma}_{0,f}(i\Delta_n) \right\|_2 \\
&\leq \frac{1}{2} \lambda_r \left(\mathbf{\Sigma}_{0,f}(i\Delta_n) \right),
\end{aligned}
$$

where the last inequality is from (A.3) and Assumption 1(c). Note that the non-zero eigenvalues of $\frac{n}{\phi k_1 k_2} \mathbf{B}_0 \mathbf{F}_i^\top \mathbf{F}_i \mathbf{B}_0^\top$ and $\frac{n}{\phi k_1 k_2} \sqrt{\mathbf{F}_i^\top \mathbf{F}_i} \mathbf{B}_0^\top \mathbf{B}_0 \sqrt{\mathbf{F}_i^\top \mathbf{F}_i}$ are the same. Thus, for large n, we have, with the probability at least $1 - 2p^{-3-a}$,

$$
\|\mathbf{\Lambda}_i^{-1}\|_{\max} \leq C p^{-1}. \tag{A.23}
$$

Consider (A.20). We have, with the probability at least $1 - 3p^{-3-a}$,

$$
\begin{aligned}
&\left\| \widehat{\mathbf{B}}_{i\Delta_n} - \mathbf{B}_0 \mathbf{H}_i \right\|_{\max} \\
&= \left\| \frac{n}{\phi k_1 k_2} \mathcal{X}_i^\top \mathcal{X}_i \widehat{\mathbf{B}}_{i\Delta_n} \mathbf{\Lambda}_i^{-1} - \frac{n}{\phi k_1 k_2} \mathbf{B}_0 \mathbf{F}_i^\top \mathbf{F}_i \mathbf{B}_0^\top \widehat{\mathbf{B}}_{i\Delta_n} \mathbf{\Lambda}_i^{-1} \right\|_{\max} \\
&\leq \left\| \left(\frac{n}{\phi k_1 k_2} \mathcal{X}_i^\top \mathcal{X}_i - \frac{n}{\phi k_1 k_2} \mathbf{B}_0 \mathbf{F}_i^\top \mathbf{F}_i \mathbf{B}_0^\top - \mathbf{\Sigma}_{0,u}(i\Delta_n) - \frac{n\zeta}{\phi k_1} \mathbf{V}^X \right) \widehat{\mathbf{B}}_{i\Delta_n} \mathbf{\Lambda}_i^{-1} \right\|_{\max} \\
&\quad + \left\| \left(\mathbf{\Sigma}_{0,u}(i\Delta_n) + \frac{n\zeta}{\phi k_1} \mathbf{V}^X \right) \widehat{\mathbf{B}}_{i\Delta_n} \mathbf{\Lambda}_i^{-1} \right\|_{\max} \\
&\leq \left\| \frac{n}{\phi k_1 k_2} \mathcal{X}_i^\top \mathcal{X}_i - \frac{n}{\phi k_1 k_2} \mathbf{B}_0 \mathbf{F}_i^\top \mathbf{F}_i \mathbf{B}_0^\top - \mathbf{\Sigma}_{0,u}(i\Delta_n) - \frac{n\zeta}{\phi k_1} \mathbf{V}^X \right\|_{\max} \times \left\| \widehat{\mathbf{B}}_{i\Delta_n} \right\|_1 \times \left\| \mathbf{\Lambda}_i^{-1} \right\|_{\max} \\
&\quad + \left\| \mathbf{\Sigma}_{0,u}(i\Delta_n) + \frac{n\zeta}{\phi k_1} \mathbf{V}^X \right\|_\infty \times \left\| \widehat{\mathbf{B}}_{i\Delta_n} \right\|_{\max} \times \left\| \mathbf{\Lambda}_i^{-1} \right\|_{\max} \\
&\leq C \left\{ n^{-1/8} (\log p)^{3/2} + p^{-1/2} \right\},
\end{aligned} \tag{A.24}
$$

where the last inequality is from Assumption 1(d), (A.9), and (A.23).

For (A.21), we have, with the probability at least $1 - 3p^{-3-a}$,

$$
\|\mathbf{H}_i\|_2 \leq \| \frac{n}{\phi k_1 k_2} \mathbf{F}_i^\top \mathbf{F}_i \|_2 \times \|\mathbf{B}_0\|_2 \times \|\widehat{\mathbf{B}}_{i\Delta_n}\|_2 \times \|\mathbf{\Lambda}_i^{-1}\|_2 \leq C, \tag{A.25}
$$

where the last inequality is from (A.3) and (A.23). Then, by (A.24) and (A.25), for large n, we have, with the probability at least $1-6p^{-3-a}$,

$$\begin{aligned}
\|\mathbf{H}_i^\top \mathbf{H}_i - \mathbf{I}_r\|_2 &\leq \|\mathbf{H}_i^\top \mathbf{H}_i - p^{-1}\mathbf{H}_i^\top \mathbf{B}_0^\top \mathbf{B}_0 \mathbf{H}_i\|_2 + \|p^{-1}\mathbf{H}_i^\top \mathbf{B}_0^\top \mathbf{B}_0 \mathbf{H}_i - \mathbf{I}_r\|_2 \\
&\leq \|\mathbf{H}_i\|_2^2 \times \left\|p^{-1}\mathbf{B}_0^\top \mathbf{B}_0 - \mathbf{I}_r\right\|_2 + p^{-1}\left\|\mathbf{H}_i^\top \mathbf{B}_0^\top \mathbf{B}_0 \mathbf{H}_i - \widehat{\mathbf{B}}_{i\Delta_n}^\top \widehat{\mathbf{B}}_{i\Delta_n}\right\|_2 \\
&\leq \frac{1}{3} + p^{-1}\left\|\widehat{\mathbf{B}}_{i\Delta_n} - \mathbf{B}_0\mathbf{H}_i\right\|_2 \times \|\mathbf{B}_0\mathbf{H}_i\|_2 + p^{-1}\left\|\widehat{\mathbf{B}}_{i\Delta_n} - \mathbf{B}_0\mathbf{H}_i\right\|_2 \times \|\mathbf{B}_0\|_2 \\
&\leq \frac{1}{2}. \qquad \text{(A.26)}
\end{aligned}$$

Thus, by the Weyl's theorem, we have, with the probability at least $1-6p^{-3-a}$,

$$\lambda_r\left(\mathbf{H}_i^\top \mathbf{H}_i\right) \geq \frac{1}{2}, \quad \det(\mathbf{H}_i) \geq C, \quad \text{and} \quad \left\|\mathbf{H}_i^{-1}\right\|_2 \leq C. \qquad \text{(A.27)}$$

Consider (A.22). We have

$$\begin{aligned}
&\left\|\widehat{\mathbf{F}}_i - \mathbf{F}_i\left(\mathbf{H}_i^\top\right)^{-1}\right\|_2 \\
&= \left\|p^{-1}\mathcal{X}_i\widehat{\mathbf{B}}_{i\Delta_n} - \mathbf{F}_i\left(\mathbf{H}_i^\top\right)^{-1}\right\|_2 \\
&\leq \left\|p^{-1}\left(\mathcal{X}_i - \mathcal{X}_i^c - \mathcal{E}_i^X\right)\widehat{\mathbf{B}}_{i\Delta_n}\right\|_2 + \left\|p^{-1}\widetilde{\mathbf{F}}_i\widehat{\mathbf{B}}_{i\Delta_n}\right\|_2 \\
&\quad + \left\|p^{-1}\left(\mathbf{F}_i\mathbf{B}_0^\top + \mathbf{U}_i + \mathcal{E}_i^X\right)\widehat{\mathbf{B}}_{i\Delta_n} - \mathbf{F}_i\left(\mathbf{H}_i^\top\right)^{-1}\right\|_2 \\
&= (I) + (II) + (III). \qquad \text{(A.28)}
\end{aligned}$$

For (I), let w_k be a kth row vector of $\mathcal{X}_i - \mathcal{X}_i^c - \mathcal{E}_i^X$. Note that

$$\sup_{\boldsymbol{\beta}\in\mathbb{R}^p, \|\boldsymbol{\beta}\|_2\leq 1} |w_k\boldsymbol{\beta}| \leq C \max_j v_{j,n}\sqrt{\sum_{j=1}^p \mathbf{1}\left(|\Delta_{i+k-1}^n \widehat{X}_j| > v_{j,n}\right)}$$

under Q_2. Thus, under $Q_2 \cap Q_4$, we have

$$(I) \leq Cp^{-1/2}\max_j v_{j,n}\sqrt{\sum_{i=1}^{k_2-k_1+1}\sum_{j=1}^p \mathbf{1}\left(|\Delta_{i+k-1}^n \widehat{X}_j| > v_{j,n}\right)}$$

$$
\begin{aligned}
&\leq \quad Cp^{-1/2}\max_j v_{j,n}\sqrt{pk_1\log p} \\
&\leq \quad C\log p,
\end{aligned}
$$

which implies

$$
\Pr\{(I)\leq C\log p\}\geq 1-p^{-3-a}. \tag{A.29}
$$

For (II), under Q_5, we have

$$
(II)\leq p^{-1}\left\|\widetilde{\mathbf{F}}_i\right\|_2\times\left\|\widehat{\mathbf{B}}_{i\Delta_n}\right\|_2\leq C\log p.
$$

Thus, we have

$$
\Pr\{(II)\leq C\log p\}\geq 1-p^{-3-a}. \tag{A.30}
$$

Consider (III). We have

$$
\begin{aligned}
&\left\|p^{-1}\left(\mathbf{F}_i\mathbf{B}_0^\top+\mathbf{U}_i+\mathcal{E}_i^X\right)\widehat{\mathbf{B}}_{i\Delta_n}-\mathbf{F}_i\left(\mathbf{H}_i^\top\right)^{-1}\right\|_2 \\
&=\left\|p^{-1}\left(\mathbf{F}_i\mathbf{B}_0^\top+\mathbf{U}_i+\mathcal{E}_i^X\right)\widehat{\mathbf{B}}_{i\Delta_n}-p^{-1}\mathbf{F}_i\left(\mathbf{H}_i^\top\right)^{-1}\widehat{\mathbf{B}}_{i\Delta_n}^\top\widehat{\mathbf{B}}_{i\Delta_n}\right\|_2 \\
&=\left\|p^{-1}\mathbf{F}_i\left(\mathbf{B}_0^\top\widehat{\mathbf{B}}_{i\Delta_n}-\left(\mathbf{H}_i^\top\right)^{-1}\widehat{\mathbf{B}}_{i\Delta_n}^\top\widehat{\mathbf{B}}_{i\Delta_n}\right)+p^{-1}\left(\mathbf{U}_i+\mathcal{E}_i^X\right)\widehat{\mathbf{B}}_{i\Delta_n}\right\|_2 \\
&\leq\left\|p^{-1}\mathbf{F}_i\left(\mathbf{H}_i^\top\right)^{-1}\left(\mathbf{H}_i^\top\mathbf{B}_0^\top-\widehat{\mathbf{B}}_{i\Delta_n}^\top\right)\widehat{\mathbf{B}}_{i\Delta_n}\right\|_2+\left\|p^{-1}\left(\mathbf{U}_i+\mathcal{E}_i^X\right)\left(\widehat{\mathbf{B}}_{i\Delta_n}-\mathbf{B}_0\mathbf{H}_i\right)\right\|_2 \\
&\quad+\left\|p^{-1}\left(\mathbf{U}_i+\mathcal{E}_i^X\right)\mathbf{B}_0\mathbf{H}_i\right\|_2 \\
&=(III)^{(1)}+(III)^{(2)}+(III)^{(3)}
\end{aligned}
$$

By (A.3), (A.24), and (A.27), we have, with the probability at least $1-10p^{-3-a}$,

$$
(III)^{(1)}\leq p^{-1}\left\|\mathbf{F}_i\right\|_2\times\left\|\mathbf{H}_i^{-1}\right\|_2\times\left\|\widehat{\mathbf{B}}_{i\Delta_n}-\mathbf{B}_0\mathbf{H}_i\right\|_2\times\left\|\widehat{\mathbf{B}}_{i\Delta_n}\right\|_2\leq C\left\{(\log p)^{3/2}+p^{-1/2}n^{1/8}\right\}.
$$

Also, by (A.4), (A.5), and (A.24), we have, with the probability at least $1 - 4p^{-3-a}$,

$$\begin{aligned}
(III)^{(2)} &\leq p^{-1} \left\| \mathbf{U}_i + \mathcal{E}_i^X \right\|_2 \times \left\| \widehat{\mathbf{B}}_{i\Delta_n} - \mathbf{B}_0 \mathbf{H}_i \right\|_2 \\
&\leq C \left\{ n^{-1/16} (\log p)^{7/4} + p^{-1} n^{1/8} + p^{-1/2} n^{1/16} (\log p)^{1/4} \right\}.
\end{aligned}$$

Consider $(III)^{(3)}$. By Assumption 1(b), the random variable $\left(\sqrt{n} \mathbf{v}^\top \Delta_i^n \left(\mathbf{u} + \boldsymbol{\varepsilon}^X \right) \middle| \mathcal{F}_{(i-1)\Delta_n} \right)$ is sub-Gaussian with bounded parameter for any unit vector $\mathbf{v} \in \mathbb{R}^p$. Thus, each element of $\left(\mathbf{U}_i + \mathcal{E}_i^X \right) \mathbf{v}$ has the sub-Gaussian tail with the order of $n^{-1/4}$. Then, using the Bernstein's inequality for martingales, we can show, for any unit vector $\mathbf{v} \in \mathbb{R}^p$,

$$\begin{aligned}
&\Pr \left\{ \left| \mathbf{v}^\top \left(\frac{n}{\phi k_1 k_2} \left(\mathbf{U}_i + \mathcal{E}_i^X \right)^\top \left(\mathbf{U}_i + \mathcal{E}_i^X \right) - \boldsymbol{\Sigma}_{0,u}(i\Delta_n) - \frac{n\zeta}{\phi k_1} \mathbf{V}^X \right) \mathbf{v} \right| \leq C n^{-1/8} \sqrt{\log p} \right\} \\
&\geq 1 - p^{-3-a}. \qquad \text{(A.31)}
\end{aligned}$$

Thus, we have, with the probability at least $1 - p^{-3-a}$,

$$\begin{aligned}
(III)^{(3)} &\leq p^{-1} \sqrt{ \left\| \mathbf{B}_0^\top \left(\mathbf{U}_i + \mathcal{E}_i^X \right)^\top \left(\mathbf{U}_i + \mathcal{E}_i^X \right) \mathbf{B}_0 \right\|_\infty } \\
&\leq p^{-1} \sqrt{ \left\| \mathbf{B}_0^\top \left\{ \left(\mathbf{U}_i + \mathcal{E}_i^X \right)^\top \left(\mathbf{U}_i + \mathcal{E}_i^X \right) - \frac{\phi k_1 k_2}{n} \boldsymbol{\Sigma}_{0,u}(i\Delta_n) - k_2 \zeta \mathbf{V}^X \right\} \mathbf{B}_0 \right\|_\infty } \\
&\quad + p^{-1} \sqrt{ \left\| \mathbf{B}_0^\top \left(\frac{\phi k_1 k_2}{n} \boldsymbol{\Sigma}_{0,u}(i\Delta_n) + k_2 \zeta \mathbf{V}^X \right) \mathbf{B}_0 \right\|_\infty } \\
&\leq C p^{-1/2} n^{1/16} (\log p)^{1/4} + p^{-1} \sqrt{ \left\| \mathbf{B}_0 \right\|_\infty \times \left\| \mathbf{B}_0 \right\|_1 \times \left\| \frac{\phi k_1 k_2}{n} \boldsymbol{\Sigma}_{0,u}(i\Delta_n) + k_2 \zeta \mathbf{V}^X \right\|_\infty } \\
&\leq C p^{-1/2} n^{1/8},
\end{aligned}$$

which implies

$$\Pr \left((III) \leq C \left\{ (\log p)^{3/2} + p^{-1/2} n^{1/8} \right\} \right) \geq 1 - 15 p^{-3-a}. \qquad \text{(A.32)}$$

Combining (A.28)–(A.30) and (A.32), we have

$$\Pr\left(\left\|\widehat{\mathbf{F}}_i-\mathbf{F}_i\left(\mathbf{H}_i^{\top}\right)^{-1}\right\|_2\leq C\left\{(\log p)^{3/2}+p^{-1/2}n^{1/8}\right\}\right)\geq 1-17p^{-3-a}. \tag{A.33}$$

Then, the statement can be shown by (A.24), (A.25), (A.27), and (A.33). ■

Proposition 4. *(Deviation condition) Under the assumptions in Theorem 1, we have, with the probability at least* $1-p^{-1-a}$,

$$\|\nabla\mathcal{L}_i(\boldsymbol{\theta}_0)\|_\infty\leq\eta/2. \tag{A.34}$$

Proof of Proposition 4. We have

$$\begin{aligned}\|\nabla\mathcal{L}_i(\boldsymbol{\theta}_0)\|_\infty &\leq \left\|\left(\frac{n}{\phi k_1k_2}\widehat{\mathbf{U}}_i^{\top}\widehat{\mathbf{U}}_i-\frac{n\zeta}{\phi k_1}\widehat{\mathbf{V}}^X-\boldsymbol{\Sigma}_{0,u}(i\Delta_n)\right)\boldsymbol{\beta}_0\right\|_\infty\\ &\quad+\left\|\frac{n}{\phi k_1k_2}\widehat{\mathbf{U}}_i^{\top}\mathcal{Y}_i-\boldsymbol{\Sigma}_{0,u}(i\Delta_n)\boldsymbol{\beta}_0\right\|_\infty\\ &\quad+\left\|\frac{n}{\phi k_1k_2}\widehat{\mathbf{F}}_i^{\top}\widehat{\mathbf{F}}_i\mathbf{H}_i^{\top}\mathbf{B}_0^{\top}\boldsymbol{\beta}_0-\frac{n}{\phi k_1k_2}\widehat{\mathbf{F}}_i^{\top}\mathcal{Y}_i\right\|_\infty\\ &\quad+\left\|\frac{n}{\phi k_1k_2}\widehat{\mathbf{U}}_i^{\top}\widehat{\mathbf{F}}_i\mathbf{H}_i^{\top}\mathbf{B}_0^{\top}\boldsymbol{\beta}_0\right\|_\infty+\left\|\frac{n}{\phi k_1k_2}\widehat{\mathbf{F}}_i^{\top}\widehat{\mathbf{U}}_i\boldsymbol{\beta}_0\right\|_\infty\\ &= (I)+(II)+(III)+(IV)+(V).\end{aligned} \tag{A.35}$$

For (I), we have

$$\begin{aligned}&\left\|\frac{n}{\phi k_1k_2}\widehat{\mathbf{U}}_i^{\top}\widehat{\mathbf{U}}_i-\frac{n\zeta}{\phi k_1}\widehat{\mathbf{V}}^X-\boldsymbol{\Sigma}_{0,u}(i\Delta_n)\right\|_{\max}\\ &\leq\frac{n}{\phi k_1k_2}\left\|\widehat{\mathbf{U}}_i^{\top}\widehat{\mathbf{U}}_i-\left(\mathcal{X}_i-\mathbf{F}_i\mathbf{B}_0^{\top}\right)^{\top}\left(\mathcal{X}_i-\mathbf{F}_i\mathbf{B}_0^{\top}\right)\right\|_{\max}\\ &\quad+\frac{n}{\phi k_1k_2}\left\|\left(\mathcal{X}_i-\mathbf{F}_i\mathbf{B}_0^{\top}\right)^{\top}\left(\mathcal{X}_i-\mathbf{F}_i\mathbf{B}_0^{\top}\right)-\left(\mathcal{X}_i^c+\mathcal{E}_i^X-\mathbf{F}_i\mathbf{B}_0^{\top}\right)^{\top}\left(\mathcal{X}_i^c+\mathcal{E}_i^X-\mathbf{F}_i\mathbf{B}_0^{\top}\right)\right\|_{\max}\\ &\quad+\frac{n}{\phi k_1k_2}\left\|\left(\mathcal{X}_i^c+\mathcal{E}_i^X-\mathbf{F}_i\mathbf{B}_0^{\top}\right)^{\top}\left(\mathcal{X}_i^c+\mathcal{E}_i^X-\mathbf{F}_i\mathbf{B}_0^{\top}\right)\right.\\ &\qquad\qquad\left.-\left(\mathcal{X}_i^c-\mathbf{F}_i\mathbf{B}_0^{\top}\right)^{\top}\left(\mathcal{X}_i^c-\mathbf{F}_i\mathbf{B}_0^{\top}\right)-k_2\zeta\mathbf{V}^X\right\|_{\max}\\ &\quad+\frac{n}{\phi k_1k_2}\left\|\left(\mathcal{X}_i^c-\mathbf{F}_i\mathbf{B}_0^{\top}\right)^{\top}\left(\mathcal{X}_i^c-\mathbf{F}_i\mathbf{B}_0^{\top}\right)-\mathbf{U}_i^{\top}\mathbf{U}_i\right\|_{\max}+\left\|\frac{n}{\phi k_1k_2}\mathbf{U}_i^{\top}\mathbf{U}_i-\boldsymbol{\Sigma}_{0,u}(i\Delta_n)\right\|_{\max}\end{aligned}$$

$$
\begin{aligned}
&+\left\|\frac{n\zeta}{\phi k_1}\left(\widehat{\mathbf{V}}^X-\mathbf{V}^X\right)\right\|_{\max} \\
&=(I)^{(1)}+(I)^{(2)}+(I)^{(3)}+(I)^{(4)}+(I)^{(5)}+(I)^{(6)}. \qquad (A.36)
\end{aligned}
$$

Consider $(I)^{(1)}$. Let $\mathcal{X}_i-\mathbf{F}_i\mathbf{B}_0^\top=(a_{jl})_{1\le j\le k_2-k_1+1,1\le l\le p}$ and $\mathcal{X}_i-\widehat{\mathbf{F}}_i\widehat{\mathbf{B}}_{i\Delta_n}^\top=(\widehat{a}_{jl})_{1\le j\le k_2-k_1+1,1\le l\le p}$. For $1\le j\le k_2-k_1+1$, let $\mathbf{f}_j$ and $\widehat{\mathbf{f}}_j$ be j-th column of $\mathbf{F}_i^\top$ and $\widehat{\mathbf{F}}_i^\top$, respectively. For $1\le l\le p$, let $\mathbf{b}_l$ and $\widehat{\mathbf{b}}_l$ be l-th column of $\mathbf{B}_0^\top$ and $\widehat{\mathbf{B}}_{i\Delta_n}^\top$, respectively. We have

$$
\begin{aligned}
a_{jl}-\widehat{a}_{jl} &= \widehat{\mathbf{f}}_j^\top\widehat{\mathbf{b}}_l-\mathbf{f}_j^\top\mathbf{b}_l \\
&= \left(\widehat{\mathbf{f}}_j^\top-\mathbf{f}_j^\top\left(\mathbf{H}_i^\top\right)^{-1}\right)\mathbf{H}_i^\top\mathbf{b}_l+\mathbf{f}_j^\top\left(\mathbf{H}_i^\top\right)^{-1}\left(\widehat{\mathbf{b}}_l-\mathbf{H}_i^\top\mathbf{b}_l\right) \\
&\quad +\left(\widehat{\mathbf{f}}_j^\top-\mathbf{f}_j^\top\left(\mathbf{H}_i^\top\right)^{-1}\right)\left(\widehat{\mathbf{b}}_l-\mathbf{H}_i^\top\mathbf{b}_l\right).
\end{aligned}
$$

Thus, by the Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
&\sum_{j=1}^{k_2-k_1+1}\left(\widehat{a}_{jl}-a_{jl}\right)^2 \\
&\le 3\sum_{j=1}^{k_2-k_1+1}\left\{\left(\widehat{\mathbf{f}}_j^\top-\mathbf{f}_j^\top\left(\mathbf{H}_i^\top\right)^{-1}\right)\mathbf{H}_i^\top\mathbf{b}_l\right\}^2+3\sum_{j=1}^{k_2-k_1+1}\left\{\mathbf{f}_j^\top\left(\mathbf{H}_i^\top\right)^{-1}\left(\widehat{\mathbf{b}}_l-\mathbf{H}_i^\top\mathbf{b}_l\right)\right\}^2 \\
&\quad +3\sum_{j=1}^{k_2-k_1+1}\left\{\left(\widehat{\mathbf{f}}_j^\top-\mathbf{f}_j^\top\left(\mathbf{H}_i^\top\right)^{-1}\right)\left(\widehat{\mathbf{b}}_l-\mathbf{H}_i^\top\mathbf{b}_l\right)\right\}^2 \\
&=(A)+(B)+(C).
\end{aligned}
$$

For (A), we have

$$
\begin{aligned}
(A) &\le 3\sum_{j=1}^{k_2-k_1+1}\mathbf{b}_l^\top\mathbf{H}_i\left(\widehat{\mathbf{f}}_j^\top-\mathbf{f}_j^\top\left(\mathbf{H}_i^\top\right)^{-1}\right)^\top\left(\widehat{\mathbf{f}}_j^\top-\mathbf{f}_j^\top\left(\mathbf{H}_i^\top\right)^{-1}\right)\mathbf{H}_i^\top\mathbf{b}_l \\
&= 3\mathbf{b}_l^\top\mathbf{H}_i\left(\widehat{\mathbf{F}}_i-\mathbf{F}_i\left(\mathbf{H}_i^\top\right)^{-1}\right)^\top\left(\widehat{\mathbf{F}}_i-\mathbf{F}_i\left(\mathbf{H}_i^\top\right)^{-1}\right)\mathbf{H}_i^\top\mathbf{b}_l \\
&\le 3\lambda_{\max}\left(\mathbf{H}_i\left(\widehat{\mathbf{F}}_i-\mathbf{F}_i\left(\mathbf{H}_i^\top\right)^{-1}\right)^\top\left(\widehat{\mathbf{F}}_i-\mathbf{F}_i\left(\mathbf{H}_i^\top\right)^{-1}\right)\mathbf{H}_i^\top\right)\mathbf{b}_l^\top\mathbf{b}_l \\
&\le C\left\|\widehat{\mathbf{F}}_i-\mathbf{F}_i\left(\mathbf{H}_i^\top\right)^{-1}\right\|_2^2\times\left\|\mathbf{H}_i\right\|_2^2.
\end{aligned}
$$

Similarly, we can show

$$\begin{aligned}(B) &\leq C\left\|\mathbf{F}_i\right\|_2^2 \times \left\|\mathbf{H}_i^{-1}\right\|_2^2 \times \left\|\widehat{\mathbf{B}}_{i\Delta_n} - \mathbf{B}_0\mathbf{H}_i\right\|_{\max}^2, \\ (C) &\leq C\left\|\widehat{\mathbf{F}}_i - \mathbf{F}_i\left(\mathbf{H}_i^{\top}\right)^{-1}\right\|_2^2 \times \left\|\widehat{\mathbf{B}}_{i\Delta_n} - \mathbf{B}_0\mathbf{H}_i\right\|_{\max}^2.\end{aligned}$$

Then, by (A.3) and Proposition 3, we have, with the probability at least $1 - 2p^{-2-a}$,

$$\max_{1\leq l\leq p}\sum_{j=1}^{k_2-k_1+1}\left(\widehat{a}_{jl} - a_{jl}\right)^2 \leq C\left\{(\log p)^3 + p^{-1}n^{1/4}\right\}.$$

Note that by (A.13), we have, with the probability at least $1 - p^{-2-a}$,

$$\max_{1\leq j\leq k_2-k_1+1, 1\leq l\leq p}\left|a_{jl}\right| \leq Cn^{-1/4}\sqrt{\log p}.$$

Thus, from the Cauchy-Schwarz inequality, we have, with the probability at least $1 - 3p^{-2-a}$,

$$\begin{aligned}&\max_{1\leq l,m\leq p}\left|\sum_{j=1}^{k_2-k_1+1}\widehat{a}_{jl}\widehat{a}_{jm} - \sum_{j=1}^{k_2-k_1+1}a_{jl}a_{jm}\right| \\ &\leq \max_{1\leq l,m\leq p}\left|\sum_{j=1}^{k_2-k_1+1}\left(\widehat{a}_{jl} - a_{jl}\right)\left(\widehat{a}_{jm} - a_{jm}\right)\right| + 2\max_{1\leq l,m\leq p}\left|\sum_{j=1}^{k_2-k_1+1}a_{jl}\left(\widehat{a}_{jm} - a_{jm}\right)\right| \\ &\leq \max_{1\leq l\leq p}\sum_{j=1}^{k_2-k_1+1}\left(\widehat{a}_{jl} - a_{jl}\right)^2 + 2\sqrt{\max_{1\leq l\leq p}\sum_{j=1}^{k_2-k_1+1}a_{jl}^2 \times \max_{1\leq l\leq p}\sum_{j=1}^{k_2-k_1+1}\left(\widehat{a}_{jl} - a_{jl}\right)^2} \\ &\leq C\left\{n^{1/8}\left(\log p\right)^2 + p^{-1/2}n^{1/4}\sqrt{\log p}\right\},\end{aligned}$$

which implies

$$\Pr\left((I)^{(1)} \leq C\left\{n^{-1/8}\left(\log p\right)^2 + p^{-1/2}\sqrt{\log p}\right\}\right) \geq 1 - 3p^{-2-a}. \qquad \text{(A.37)}$$

Consider $(I)^{(2)}$. By (A.13), we have, with the probability at least $1-p^{-2-a}$,

$$\begin{aligned}(I)^{(2)} &\leq \frac{n}{\phi k_1 k_2}\left\|\left(\mathcal{X}_i-\mathbf{F}_i\mathbf{B}_0^\top\right)^\top\left(\mathcal{X}_i-\mathbf{F}_i\mathbf{B}_0^\top\right)-\left(\mathcal{X}_i'-\mathbf{F}_i\mathbf{B}_0^\top\right)^\top\left(\mathcal{X}_i'-\mathbf{F}_i\mathbf{B}_0^\top\right)\right\|_{\max} \\ &\quad+\frac{n}{\phi k_1 k_2}\left\|\left(\mathcal{X}_i'-\mathbf{F}_i\mathbf{B}_0^\top\right)^\top\left(\mathcal{X}_i'-\mathbf{F}_i\mathbf{B}_0^\top\right)-\left(\mathcal{X}_i^c+\mathcal{E}_i^X-\mathbf{F}_i\mathbf{B}_0^\top\right)^\top\left(\mathcal{X}_i^c+\mathcal{E}_i^X-\mathbf{F}_i\mathbf{B}_0^\top\right)\right\|_{\max} \\ &\leq Cn^{-1/4}\left(\log p\right)^2. \end{aligned} \tag{A.38}$$

For $(I)^{(3)}$, note that the elements of $\mathcal{X}_i^c$, $\mathbf{F}_i$, and $\mathcal{E}_i^X$ have sub-Gaussian tails. Hence, by Bernstein's inequality for martingales, we have

$$\Pr\left\{\frac{n}{\phi k_1 k_2}\left\|\left(\mathcal{X}_i^c-\mathbf{F}_i\mathbf{B}_0^\top\right)^\top\mathcal{E}_i^X\right\|_{\max}\leq Cn^{-1/8}\sqrt{\log p}\right\}\geq 1-p^{-2-a}.$$

Then, by (A.5), we have, with the probability at least $1-2p^{-2-a}$,

$$\begin{aligned}(I)^{(3)} &\leq \frac{2n}{\phi k_1 k_2}\left\|\left(\mathcal{X}_i^c-\mathbf{F}_i\mathbf{B}_0^\top\right)^\top\mathcal{E}_i^X\right\|_{\max}+\frac{n}{\phi k_1 k_2}\left\|\left(\mathcal{E}_i^X\right)^\top\mathcal{E}_i^X-k_2\zeta\mathbf{V}^X\right\|_{\max} \\ &\leq Cn^{-1/8}\sqrt{\log p}. \end{aligned} \tag{A.39}$$

Consider $(I)^{(4)}$. By (A.13), we have, with the probability at least $1-p^{-2-a}$,

$$\begin{aligned}(I)^{(4)} &\leq \frac{2n}{\phi k_1 k_2}\left\|\widetilde{\mathbf{F}}_i^\top\mathbf{U}_i\right\|_{\max}+\frac{n}{\phi k_1 k_2}\left\|\widetilde{\mathbf{F}}_i^\top\widetilde{\mathbf{F}}_i\right\|_{\max} \\ &\leq Cn^{-1/8}\left(\log p\right)^{3/2}. \end{aligned} \tag{A.40}$$

For $(I)^{(5)}$ and $(I)^{(6)}$, by Proposition 2, we have

$$\Pr\left\{(I)^{(5)}+(I)^{(6)}\leq Cn^{-1/8}\sqrt{\log p}\right\}\geq 1-p^{-2-a}. \tag{A.41}$$

Combining (A.36)–(A.41), we have, with the probability at least $1 - 8p^{-2-a}$,

$$\left\| \frac{n}{\phi k_1 k_2} \widehat{\mathbf{U}}_i^\top \widehat{\mathbf{U}}_i - \frac{n\zeta}{\phi k_1} \widehat{\mathbf{V}}^X - \mathbf{\Sigma}_{0,u}(i\Delta_n) \right\|_{\max} \leq C \left\{ n^{-1/8} (\log p)^2 + p^{-1/2} \sqrt{\log p} \right\},$$

which implies

$$\Pr\left((I) \leq C \left\{ s_p n^{-1/8} (\log p)^2 + p^{-1/2} s_p \sqrt{\log p} \right\} \right) \geq 1 - 8p^{-2-a}. \tag{A.42}$$

Consider (II). We have

$$\begin{aligned}
& (II) \\
\leq & \frac{n}{\phi k_1 k_2} \left\| \left(\widehat{\mathbf{F}}_i \widehat{\mathbf{B}}_{i\Delta_n}^\top - \mathbf{F}_i \mathbf{B}_0^\top \right)^\top \mathcal{Y}_i \right\|_\infty \\
& + \frac{n}{\phi k_1 k_2} \left\| \left(\mathcal{X}_i - \mathbf{F}_i \mathbf{B}_0^\top \right)^\top \mathcal{Y}_i - \left(\mathcal{X}_i^c + \mathcal{E}_i^X - \mathbf{F}_i \mathbf{B}_0^\top \right)^\top \left(\mathcal{Y}_i^c + \mathcal{E}_i^Y \right) \right\|_\infty \\
& + \frac{n}{\phi k_1 k_2} \left\| \left(\mathcal{X}_i^c + \mathcal{E}_i^X - \mathbf{F}_i \mathbf{B}_0^\top \right)^\top \left(\mathcal{Y}_i^c + \mathcal{E}_i^Y \right) - \left(\mathbf{U}_i + \mathcal{E}_i^X \right)^\top \left(\mathbf{U}_i \boldsymbol{\beta}_0 + \mathbf{F}_i \mathbf{B}_0^\top \boldsymbol{\beta}_0 + \mathcal{Z}_i + \mathcal{E}_i^Y \right) \right\|_\infty \\
& + \frac{n}{\phi k_1 k_2} \left\| \left(\mathbf{U}_i + \mathcal{E}_i^X \right)^\top \left(\mathbf{U}_i \boldsymbol{\beta}_0 + \mathbf{F}_i \mathbf{B}_0^\top \boldsymbol{\beta}_0 + \mathcal{Z}_i + \mathcal{E}_i^Y \right) - \mathbf{U}_i^\top \mathbf{U}_i \boldsymbol{\beta}_0 \right\|_\infty \\
& + \left\| \frac{n}{\phi k_1 k_2} \mathbf{U}_i^\top \mathbf{U}_i \boldsymbol{\beta}_0 - \mathbf{\Sigma}_{0,u}(i\Delta_n) \boldsymbol{\beta}_0 \right\|_\infty \\
= & (II)^{(1)} + (II)^{(2)} + (II)^{(3)} + (II)^{(4)} + (II)^{(5)}.
\end{aligned} \tag{A.43}$$

For $(II)^{(1)}$, by the Cauchy-Schwarz inequality, we have, with the probability at least $1 - 2p^{-2-a}$,

$$\begin{aligned}
(II)^{(1)} & \leq C n^{-1/4} \sqrt{ \max_{1 \leq l \leq p} \sum_{j=1}^{k_2 - k_1 + 1} \left(\widehat{a}_{jl} - a_{jl} \right)^2 \times (k_2 - k_1 + 1) \left\| \mathcal{Y}_i \right\|_\infty^2 } \\
& \leq C \left\{ s_p n^{-1/8} (\log p)^2 + p^{-1/2} s_p \sqrt{\log p} \right\}.
\end{aligned} \tag{A.44}$$

For $(II)^{(2)}$, by (A.13), we have, with the probability at least $1 - p^{-2-a}$,

$$(II)^{(2)} \leq \frac{n}{\phi k_1 k_2} \left\| \left(\mathcal{X}_i - \mathbf{F}_i \mathbf{B}_0^\top \right)^\top \left(\mathcal{Y}_i - \mathcal{Y}_i' \right) \right\|_\infty + \frac{n}{\phi k_1 k_2} \left\| \left(\mathcal{X}_i - \mathbf{F}_i \mathbf{B}_0^\top \right)^\top \left(\mathcal{Y}_i' - \mathcal{Y}_i^c - \mathcal{E}_i^Y \right) \right\|_\infty$$

$$
\begin{aligned}
&+\frac{n}{\phi k_1 k_2}\left\|\left(\mathcal{X}_i-\mathcal{X}_i^{\prime}\right)^{\top}\left(\mathcal{Y}_i^c+\mathcal{E}_i^Y\right)\right\|_{\infty}+\frac{n}{\phi k_1 k_2}\left\|\left(\mathcal{X}_i^{\prime}-\mathcal{X}_i^c-\mathcal{E}_i^X\right)^{\top}\left(\mathcal{Y}_i^c+\mathcal{E}_i^Y\right)\right\|_{\infty} \\
\leq\ & C s_p n^{-1/4}(\log p)^2 . \qquad \text{(A.45)}
\end{aligned}
$$

Also, by (A.13), we have, with the probability at least $1-p^{-2-a}$,

$$
\begin{aligned}
(II)^{(3)} \leq\ & \frac{n}{\phi k_1 k_2}\left\|\left(\mathbf{U}_i+\mathcal{E}_i^X\right)^{\top}\left(\widetilde{\mathbf{F}}_i \boldsymbol{\beta}_0+\widetilde{\mathcal{X}}_i\right)\right\|_{\infty}+\frac{n}{\phi k_1 k_2}\left\|\widetilde{\mathbf{F}}_i^{\top}\left(\mathbf{U}_i \boldsymbol{\beta}_0+\mathbf{F}_i \mathbf{B}_0^{\top} \boldsymbol{\beta}_0+\mathcal{Z}_i+\mathcal{E}_i^Y\right)\right\|_{\infty} \\
&+\frac{n}{\phi k_1 k_2}\left\|\widetilde{\mathbf{F}}_i^{\top}\left(\widetilde{\mathbf{F}}_i \boldsymbol{\beta}_0+\widetilde{\mathcal{X}}_i\right)\right\|_{\infty} \\
\leq\ & C s_p n^{-1/8}(\log p)^{3/2} . \qquad \text{(A.46)}
\end{aligned}
$$

Consider $(II)^{(4)}$. Note that the elements of $\mathbf{U}_i$, $\mathbf{F}_i$, $\mathcal{Z}_i$, $\mathcal{E}_i^Y$, and $\mathcal{E}_i^X$ have sub-Gaussian tails. Hence, by Bernstein's inequality for martingales, we have

$$
\operatorname{Pr}\left\{(II)^{(4)} \leq C s_p n^{-1/8} \sqrt{\log p}\right\} \geq 1-p^{-2-a}. \qquad \text{(A.47)}
$$

Consider $(II)^{(5)}$. By (A.4), we have, with the probability at least $1-p^{-2-a}$,

$$
\begin{aligned}
(II)^{(5)} \leq\ & \left\|\frac{n}{\phi k_1 k_2} \mathbf{U}_i^{\top} \mathbf{U}_i-\boldsymbol{\Sigma}_{0,u}(i \Delta_n)\right\|_{\max } \times\left\|\boldsymbol{\beta}_0\right\|_1 \\
\leq\ & C s_p n^{-1/8} \sqrt{\log p}. \qquad \text{(A.48)}
\end{aligned}
$$

Combining (A.43)–(A.48), we have

$$
\operatorname{Pr}\left((II) \leq C\left\{s_p n^{-1/8}(\log p)^2+p^{-1/2} s_p \sqrt{\log p}\right\}\right) \geq 1-6 p^{-2-a}. \qquad \text{(A.49)}
$$

Also, similar to the proofs of (A.42) and (A.49), we can show

$$
\operatorname{Pr}\left((III) \leq C\left\{s_p n^{-1/8}(\log p)^2+p^{-1/2} s_p \sqrt{\log p}\right\}\right) \geq 1-2 p^{-2-a}. \qquad \text{(A.50)}
$$

Consider (IV). We have

$$\begin{aligned}\frac{n}{\phi k_1 k_2}\left\|\widehat{\mathbf{U}}_i^\top \widehat{\mathbf{F}}_i \mathbf{H}_i^\top\right\|_{\max} &\leq \frac{n}{\phi k_1 k_2}\left\|\mathbf{U}_i^\top \mathbf{F}_i\right\|_{\max} + \frac{n}{\phi k_1 k_2}\left\|\widehat{\mathbf{U}}_i^\top\left(\widehat{\mathbf{F}}_i \mathbf{H}_i^\top - \mathbf{F}_i\right)\right\|_{\max} \\ &\quad + \frac{n}{\phi k_1 k_2}\left\|\left(\widehat{\mathbf{U}}_i - \mathbf{U}_i\right)^\top \mathbf{F}_i\right\|_{\max} \\ &= (IV)^{(1)} + (IV)^{(2)} + (IV)^{(3)}.\end{aligned}$$

For $(IV)^{(1)}$, by Proposition 2, we have

$$\Pr\left\{(IV)^{(1)} \leq C n^{-1/8}\sqrt{\log p}\right\} \geq 1 - p^{-2-a}.$$

For $(IV)^{(2)}$, by Proposition 3, we have, with the probability at least $1 - p^{-2-a}$,

$$\left\|\widehat{\mathbf{U}}_i\right\|_{\max} \leq C n^{-1/4}\sqrt{\log p} \quad \text{and} \quad \left\|\widehat{\mathbf{F}}_i \mathbf{H}_i^\top - \mathbf{F}_i\right\|_F \leq C\left\{(\log p)^{3/2} + p^{-1/2} n^{1/8}\right\}.$$

Thus, by the Cauchy-Schwarz inequality, we have

$$\Pr\left((IV)^{(2)} \leq C\left\{n^{-1/8}(\log p)^2 + p^{-1/2}\sqrt{\log p}\right\}\right) \geq 1 - p^{-2-a}.$$

Consider $(IV)^{(3)}$. We have

$$(IV)^{(3)} \leq \frac{n}{\phi k_1 k_2}\left\|\left(\mathbf{F}_i \mathbf{B}_0^\top - \widehat{\mathbf{F}}_i \widehat{\mathbf{B}}_{i\Delta_n}^\top + \widetilde{\mathbf{F}}_i + \mathcal{X}_i - \mathcal{X}_i^c - \mathcal{E}_i^X\right)^\top \mathbf{F}_i\right\|_{\max} + \frac{n}{\phi k_1 k_2}\left\|\left(\mathcal{E}_i^X\right)^\top \mathbf{F}_i\right\|_{\max}.$$

For the first term, recall that

$$\Pr\left(\max_{1 \leq l \leq p} \sum_{j=1}^{k_2 - k_1 + 1} \left(\widehat{a}_{jl} - a_{jl}\right)^2 \leq C\left\{(\log p)^3 + p^{-1} n^{1/4}\right\}\right) \geq 1 - 2p^{-2-a}.$$

Thus, by (A.13) and Cauchy-Schwarz inequality, we have, with the probability at least $1-3p^{-2-a}$,

$$\frac{n}{\phi k_1 k_2}\left\|\left(\mathbf{F}_i\mathbf{B}_0^\top-\widehat{\mathbf{F}}_i\widehat{\mathbf{B}}_{i\Delta_n}^\top+\widetilde{\mathbf{F}}_i+\mathcal{X}_i-\mathcal{X}_i^c-\mathcal{E}_i^X\right)^\top\mathbf{F}_i\right\|_{\max}\leq C\left\{n^{-1/8}(\log p)^2+p^{-1/2}\sqrt{\log p}\right\}.$$

For the second term, by Proposition 2, we have

$$\Pr\left\{\frac{n}{\phi k_1 k_2}\left\|\left(\mathcal{E}_i^X\right)^\top\mathbf{F}_i\right\|_{\max}\leq Cn^{-1/8}\sqrt{\log p}\right\}\geq 1-p^{-2-a}.$$

Thus, we have

$$\Pr\left((IV)^{(3)}\leq C\left\{n^{-1/8}(\log p)^2+p^{-1/2}\sqrt{\log p}\right\}\right)\geq 1-4p^{-2-a},$$

which implies

$$\Pr\left(\frac{n}{\phi k_1 k_2}\left\|\widehat{\mathbf{U}}_i^\top\widehat{\mathbf{F}}_i\mathbf{H}_i^\top\right\|_{\max}\leq C\left\{n^{-1/8}(\log p)^2+p^{-1/2}\sqrt{\log p}\right\}\right)\geq 1-6p^{-2-a}$$

and

$$\Pr\left((IV)\leq C\left\{s_p n^{-1/8}(\log p)^2+p^{-1/2}s_p\sqrt{\log p}\right\}\right)\geq 1-6p^{-2-a}. \tag{A.51}$$

Similarly, we can show

$$\Pr\left((V)\leq C\left\{s_p n^{-1/8}(\log p)^2+p^{-1/2}s_p\sqrt{\log p}\right\}\right)\geq 1-6p^{-2-a}. \tag{A.52}$$

Then, (A.34) is obtained by (A.35), (A.42), and (A.49)–(A.52). ■

Proposition 5. *(RE condition) Under the assumptions in Theorem 1, there exist positive constants*

α_1, α_2, *and* κ *such that, with the probability at least* $1 - p^{-1-a}$,

$$\boldsymbol{\theta}^\top \nabla^2 \mathcal{L}_i(\boldsymbol{\theta}) \boldsymbol{\theta} \geq \alpha_1 \|\boldsymbol{\theta}\|_2^2 - \kappa \left(n^{-1/4} + p^{-1}\right) (\log p)^{9/2} \|\boldsymbol{\theta}\|_1^2 \quad \textit{for all } \boldsymbol{\theta} \in \mathbb{R}^{p+r}, \tag{A.53}$$

$$\boldsymbol{\theta}^\top \nabla^2 \mathcal{L}_i(\boldsymbol{\theta}) \boldsymbol{\theta} \leq \alpha_2 \|\boldsymbol{\theta}\|_2^2 + \kappa \left(n^{-1/4} + p^{-1}\right) (\log p)^{9/2} \|\boldsymbol{\theta}\|_1^2 \quad \textit{for all } \boldsymbol{\theta} \in \mathbb{R}^{p+r}. \tag{A.54}$$

Proof of Proposition 5. The drift term $\boldsymbol{\mu}(t)$ has a negligible order comparing with the Brownian motion term. Thus, for simplicity, we assume that $\boldsymbol{\mu}(t) = 0$ for $0 \leq t \leq 1$ without loss of generality. For any parameter $s \geq 1$, define

$$\mathbb{K}_1^s = \left\{ \mathbf{x} \in \mathbb{R}^{p+r} \mid \|\mathbf{x}\|_2 \leq 1, \|\mathbf{x}\|_0 \leq s \right\} \quad \text{and} \quad \mathbb{K}_2^s = \left\{ \mathbf{x} \in \mathbb{R}^{p} \mid \|\mathbf{x}\|_2 \leq 1, \|\mathbf{x}\|_0 \leq s \right\}.$$

Also, define

$$\mathbb{B}^r = \left\{ \mathbf{x} \in \mathbb{R}^{r} \mid \|\mathbf{x}\|_2 \leq 1 \right\} \quad \text{and} \quad \boldsymbol{\Sigma}_{0,g}(i\Delta_n) = \left(\begin{array}{c|c} \boldsymbol{\Sigma}_{0,u}(i\Delta_n) & \mathbf{0}_{p\times r} \\ \hline \mathbf{0}_{r\times p} & \boldsymbol{\Lambda}_{0,f}(i\Delta_n) \end{array} \right),$$

where $\boldsymbol{\Lambda}_{0,f}(i\Delta_n) = \mathrm{Diag}\left(\lambda_{0,f,1}(i\Delta_n), \ldots, \lambda_{0,f,r}(i\Delta_n)\right)$ and $\lambda_{0,f,j}(i\Delta_n)$ is the j-th largest eigenvalue of $\boldsymbol{\Sigma}_{0,f}(i\Delta_n)$ for $j = 1, \ldots, r$. We note that by Assumption 1(a),(c),(d), the eigenvalues of $\boldsymbol{\Sigma}_{0,g}(i\Delta_n)$ are bounded above and below. Since

$$\nabla^2 \mathcal{L}_i(\boldsymbol{\theta}) = \left(\begin{array}{c|c} \frac{n}{\phi k_1 k_2} \widehat{\mathbf{U}}_i^\top \widehat{\mathbf{U}}_i - \frac{n\zeta}{\phi k_1} \widehat{\mathbf{V}}^X & \frac{n}{\phi k_1 k_2} \widehat{\mathbf{U}}_i^\top \widehat{\mathbf{F}}_i \\ \hline \frac{n}{\phi k_1 k_2} \widehat{\mathbf{F}}_i^\top \widehat{\mathbf{U}}_i & \frac{n}{\phi k_1 k_2} \widehat{\mathbf{F}}_i^\top \widehat{\mathbf{F}}_i \end{array} \right),$$

we have

$$\begin{aligned} &\sup_{\boldsymbol{\theta} \in \mathbb{K}_1^{s+r}} \left| \boldsymbol{\theta}^\top \left(\nabla^2 \mathcal{L}_i(\boldsymbol{\theta}) - \boldsymbol{\Sigma}_{0,g}(i\Delta_n) \right) \boldsymbol{\theta} \right| \\ &\leq \sup_{\boldsymbol{\beta} \in \mathbb{K}_2^s, \boldsymbol{\gamma} \in \mathbb{B}^r} \left| \left(\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top\right) \left(\nabla^2 \mathcal{L}_i(\boldsymbol{\theta}) - \boldsymbol{\Sigma}_{0,g}(i\Delta_n) \right) \left(\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top\right)^\top \right| \end{aligned}$$

$$
\begin{aligned}
&\leq \sup_{\boldsymbol{\beta}\in\mathbb{K}_2^s}\left|\boldsymbol{\beta}^\top\left(\frac{n}{\phi k_1 k_2}\widehat{\mathbf{U}}_i^\top\widehat{\mathbf{U}}_i-\boldsymbol{\Sigma}_{0,u}(i\Delta_n)-\frac{n\zeta}{\phi k_1}\mathbf{V}^X\right)\boldsymbol{\beta}\right| \\
&\quad+\sup_{\boldsymbol{\beta}\in\mathbb{K}_2^s}\left|\boldsymbol{\beta}^\top\left(\frac{n\zeta}{\phi k_1}\widehat{\mathbf{V}}^X-\frac{n\zeta}{\phi k_1}\mathbf{V}^X\right)\boldsymbol{\beta}\right|+\sup_{\boldsymbol{\beta}\in\mathbb{K}_2^s,\boldsymbol{\gamma}\in\mathbb{B}^r}\left|\frac{2n}{\phi k_1 k_2}\boldsymbol{\beta}^\top\widehat{\mathbf{U}}_i^\top\widehat{\mathbf{F}}_i\boldsymbol{\gamma}\right| \\
&\quad+\sup_{\boldsymbol{\gamma}\in\mathbb{B}^r}\left|\boldsymbol{\gamma}^\top\left(\frac{n}{\phi k_1 k_2}\widehat{\mathbf{F}}_i^\top\widehat{\mathbf{F}}_i-\boldsymbol{\Sigma}_{0,f}(i\Delta_n)\right)\boldsymbol{\gamma}\right| \\
&=(I)+(II)+(III)+(IV). \qquad\qquad (A.55)
\end{aligned}
$$

For (I), we have

$$
\begin{aligned}
(I) \quad \leq \quad & \sup_{\boldsymbol{\beta}\in\mathbb{K}_2^s}\left|\boldsymbol{\beta}^\top\left(\frac{n}{\phi k_1 k_2}\left(\mathbf{U}_i+\mathcal{E}_i^X\right)^\top\left(\mathbf{U}_i+\mathcal{E}_i^X\right)-\boldsymbol{\Sigma}_{0,u}(i\Delta_n)-\frac{n\zeta}{\phi k_1}\mathbf{V}^X\right)\boldsymbol{\beta}\right| \\
&+\sup_{\boldsymbol{\beta}\in\mathbb{K}_2^s}\frac{n}{\phi k_1 k_2}\left|\boldsymbol{\beta}^\top\left(\widehat{\mathbf{U}}_i^\top\widehat{\mathbf{U}}_i-\left(\mathbf{U}_i+\mathcal{E}_i^X\right)^\top\left(\mathbf{U}_i+\mathcal{E}_i^X\right)\right)\boldsymbol{\beta}\right| \\
= \quad & (I)^{(1)}+(I)^{(2)}. \qquad\qquad (A.56)
\end{aligned}
$$

Consider $(I)^{(1)}$. By Assumption 1(d), the random variable $\left(\sqrt{n}\mathbf{v}^\top\Delta_i^n\mathbf{u}\middle|\mathcal{F}_{(i-1)\Delta_n}\right)$ is sub-Gaussian with bounded parameter for any unit vector $\mathbf{v}\in\mathbb{R}^p$. Thus, each element of $\left(\mathbf{U}_i+\mathcal{E}_i^X\right)\boldsymbol{\beta}$ has the sub-Gaussian tail with the order of $n^{-1/4}\left\|\boldsymbol{\beta}\right\|_2$. Then, using the Bernstein's inequality for martingales, we can show, for any fixed unit vector $\boldsymbol{\beta}\in\mathbb{R}^p$,

$$
\begin{aligned}
&\Pr\left\{\left|\boldsymbol{\beta}^\top\left(\frac{n}{\phi k_1 k_2}\left(\mathbf{U}_i+\mathcal{E}_i^X\right)^\top\left(\mathbf{U}_i+\mathcal{E}_i^X\right)-\boldsymbol{\Sigma}_{0,u}(i\Delta_n)-\frac{n\zeta}{\phi k_1}\mathbf{V}^X\right)\boldsymbol{\beta}\right|\leq\frac{\lambda_{\min}\left\{\boldsymbol{\Sigma}_{0,g}\left(i\Delta_n\right)\right\}}{216}\right\} \\
&\geq 1-\exp\left(-c_3 n^{1/4}\right) \qquad\qquad (A.57)
\end{aligned}
$$

for some constant $c_3>0$. For any subset $U\subset\{1,\ldots,p\}$, define

$$
A_U=\left\{\boldsymbol{\beta}\in\mathbb{R}^p \;\middle|\; \|\boldsymbol{\beta}\|_2\leq 1,\,\mathrm{supp}\left(\boldsymbol{\beta}\right)\subset U\right\}.
$$

By (A.57) and discretization argument in Lemma 15 (Loh and Wainwright, 2012), for any A_U with

$|U| \le s$, we have, with the probability at least $1 - 9^s \exp\left(-c_3 n^{1/4}\right)$,

$$\sup_{\boldsymbol{\beta} \in A_U} \left| \boldsymbol{\beta}^\top \left(\frac{n}{\phi k_1 k_2} \left(\mathbf{U}_i + \mathcal{E}_i^X\right)^\top \left(\mathbf{U}_i + \mathcal{E}_i^X\right) - \boldsymbol{\Sigma}_{0,u}(i\Delta_n) - \frac{n\zeta}{\phi k_1} \mathbf{V}^X \right) \boldsymbol{\beta} \right| \le \frac{\lambda_{\min}\left\{\boldsymbol{\Sigma}_{0,g}\left(i\Delta_n\right)\right\}}{216}.$$

Note that $\mathbb{K}_2^s = \cup_{|U| \le s} A_U$ and $\binom{p}{s} \le p^s$. Hence, we have

$$\Pr\left\{ (I)^{(1)} \le \frac{\lambda_{\min}\left\{\boldsymbol{\Sigma}_{0,g}\left(i\Delta_n\right)\right\}}{216} \right\} \ge 1 - \exp\left(-c_3 n^{1/4} + s \log 9p\right). \tag{A.58}$$

Consider $(I)^{(2)}$. We have

$$\widehat{\mathbf{U}}_i - \mathbf{U}_i - \mathcal{E}_i^X = \mathcal{X}_i - \mathcal{X}_i^c - \mathcal{E}_i^X + \widetilde{\mathbf{F}}_i + \mathbf{F}_i \mathbf{B}_0^\top - \widehat{\mathbf{F}}_i \widehat{\mathbf{B}}_{i\Delta_n}^\top.$$

Thus, by the Cauchy-Schwarz inequality, we have

$$(I)^{(2)} \le \left(2\sqrt{D_1} + \sqrt{D_2} + \sqrt{D_3} + \sqrt{D_4}\right) \left(\sqrt{D_2} + \sqrt{D_3} + \sqrt{D_4}\right), \tag{A.59}$$

where

$$\begin{aligned}
D_1 &= \sup_{\boldsymbol{\beta} \in \mathbb{K}_2^s} \frac{n}{\phi k_1 k_2} \left| \boldsymbol{\beta}^\top \left(\mathbf{U}_i + \mathcal{E}_i^X\right)^\top \left(\mathbf{U}_i + \mathcal{E}_i^X\right) \boldsymbol{\beta} \right|, \\
D_2 &= \sup_{\boldsymbol{\beta} \in \mathbb{K}_2^s} \frac{n}{\phi k_1 k_2} \left| \boldsymbol{\beta}^\top \left(\mathcal{X}_i - \mathcal{X}_i^c - \mathcal{E}_i^X\right)^\top \left(\mathcal{X}_i - \mathcal{X}_i^c - \mathcal{E}_i^X\right) \boldsymbol{\beta} \right|, \\
D_3 &= \sup_{\boldsymbol{\beta} \in \mathbb{K}_2^s} \frac{n}{\phi k_1 k_2} \left| \boldsymbol{\beta}^\top \widetilde{\mathbf{F}}_i^\top \widetilde{\mathbf{F}}_i \boldsymbol{\beta} \right|, \\
D_4 &= \sup_{\boldsymbol{\beta} \in \mathbb{K}_2^s} \frac{n}{\phi k_1 k_2} \left| \boldsymbol{\beta}^\top \left(\mathbf{F}_i \mathbf{B}_0^\top - \widehat{\mathbf{F}}_i \widehat{\mathbf{B}}_{i\Delta_n}^\top\right)^\top \left(\mathbf{F}_i \mathbf{B}_0^\top - \widehat{\mathbf{F}}_i \widehat{\mathbf{B}}_{i\Delta_n}^\top\right) \boldsymbol{\beta} \right|.
\end{aligned}$$

For D_1, by Assumption 1(d), each element of $\mathbf{U}_i + \mathcal{E}_i^X$ has the sub-Gaussian tail with the order of

$n^{-1/4}$. Thus, we can show

$$\Pr\left\{\sup_{\|\boldsymbol{\beta}\|_2\leq 1}\left\|\left(\mathbf{U}_i+\mathcal{E}_i^X\right)\boldsymbol{\beta}\right\|_\infty \leq Cn^{-1/4}\sqrt{\log p}\right\}\geq 1-p^{-2-a},$$

which implies

$$\Pr\left\{D_1 \leq C\log p\right\}\geq 1-p^{-2-a}. \tag{A.60}$$

Consider D_2. For some large constant $C>0$, let

$$A_1=\left\{\sum_{j=1}^{p}\int_{i\Delta_n}^{(i+k_2)\Delta_n} d\Lambda_j(t)\leq C\log p\right\},$$
$$A_2=\left\{\sum_{j=1}^{p}\sum_{k=1}^{k_2-k_1+1}\mathbf{1}\left(|\Delta_{i+k-1}^n\widehat{X}_j|>v_{j,n}\right)\leq Ck_1\log p\right\}.$$

By Assumption 1(e), similar to the proofs of (A.13), we can show

$$\Pr\left(Q_2\cap A_1\cap A_2\right)\geq 1-p^{-2-a}.$$

Also, let w_k be a kth row vector of $\mathcal{X}_i-\mathcal{X}_i^c-\mathcal{E}_i^X$. Under Q_2, we have

$$\sup_{\|\boldsymbol{\beta}\|_2\leq 1}|w_k\boldsymbol{\beta}|\leq C\max_j v_{j,n}\sqrt{\sum_{j=1}^{p}\mathbf{1}\left(|\Delta_{i+k-1}^n\widehat{X}_j|>v_{j,n}\right)}.$$

Thus, we have, with the probability at least $1-p^{-2-a}$,

$$D_2\leq C\left(\max_j v_{j,n}\right)^2 n^{1/4}\log p\leq Cn^{-1/4}\left(\log p\right)^2. \tag{A.61}$$

For D_3, by (A.13), we have

$$\Pr\left\{D_3\leq Cn^{-1/4}\left(\log p\right)^2 s\right\}\geq 1-p^{-2-a}. \tag{A.62}$$

Consider D_4. By the proofs of (A.37), we have

$$\Pr\left(\left\|\left(\mathbf{F}_i\mathbf{B}_0^\top - \widehat{\mathbf{F}}_i\widehat{\mathbf{B}}_{i\Delta_n}^\top\right)^\top\left(\mathbf{F}_i\mathbf{B}_0^\top - \widehat{\mathbf{F}}_i\widehat{\mathbf{B}}_{i\Delta_n}^\top\right)\right\|_{\max} \le C\left\{(\log p)^3 + p^{-1}n^{1/4}\right\}\right) \ge 1 - 2p^{-2-a},$$

which implies

$$\Pr\left(D_4 \le C\left\{n^{-1/4}(\log p)^3 s + p^{-1}s\right\}\right) \ge 1 - 2p^{-2-a}. \tag{A.63}$$

Combining (A.59)–(A.63), we have

$$\Pr\left\{(I)^{(2)} \le C\left(n^{-1/8}(\log p)^2\sqrt{s} + n^{-1/4}(\log p)^3 s + p^{-1}s\right)\right\} \ge 1 - 5p^{-2-a}. \tag{A.64}$$

Then, by (A.56), (A.58), and (A.64), we have

$$\Pr\left\{(I) \le \frac{\lambda_{\min}\left\{\boldsymbol{\Sigma}_{0,g}(i\Delta_n)\right\}}{216} + C\left(n^{-1/8}(\log p)^2\sqrt{s} + n^{-1/4}(\log p)^3 s + p^{-1}s\right)\right\}$$
$$\ge 1 - \exp\left(-c_3 n^{1/4} + s\log 9p\right) - 5p^{-2-a}.$$

Choose

$$s = \frac{1}{(\log p)^{9/2}\left(n^{-1/4} + p^{-1}\right)}. \tag{A.65}$$

We have, for large n,

$$C\left(n^{-1/8}(\log p)^2\sqrt{s} + n^{-1/4}(\log p)^3 s + p^{-1}s\right) \le \frac{\lambda_{\min}\left\{\boldsymbol{\Sigma}_{0,g}(i\Delta_n)\right\}}{216}.$$

Also, using the fact that $c_3 n^{1/4} \ge (2a+4)\log p$ for large n, we have

$$\exp\left(-c_3 n^{1/4} + s\log 9p\right) + 5p^{-2-a} \le \exp\left(-c_3 n^{1/4} + c_3 n^{1/4}/2\right) + 5p^{-2-a} \le 6p^{-2-a}.$$

Thus, we have

$$\Pr\left\{(I) \le \frac{\lambda_{\min}\left\{\boldsymbol{\Sigma}_{0,g}\left(i\Delta_n\right)\right\}}{108}\right\} \ge 1 - 6p^{-2-a}. \tag{A.66}$$

Similarly, for the same s, we can show

$$\Pr\left\{(II) + (III) + (IV) \le \frac{\lambda_{\min}\left\{\boldsymbol{\Sigma}_{0,g}\left(i\Delta_n\right)\right\}}{108}\right\} \ge 1 - 6p^{-2-a}. \tag{A.67}$$

From (A.55), (A.66), and (A.67), we have

$$\Pr\left\{\sup_{\boldsymbol{\theta}\in\mathbb{K}_1^{s+r}} \left|\boldsymbol{\theta}^\top\left(\nabla^2\mathcal{L}_i(\boldsymbol{\theta}) - \boldsymbol{\Sigma}_{0,g}(i\Delta_n)\right)\boldsymbol{\theta}\right| \le \frac{\lambda_{\min}\left\{\boldsymbol{\Sigma}_{0,g}\left(i\Delta_n\right)\right\}}{54}\right\} \ge 1 - p^{-1-a} \tag{A.68}$$

for large n. Then, by Lemma 13 (Loh and Wainwright, 2012), we have, with the probability at least $1 - p^{-1-a}$,

$$\boldsymbol{\theta}^\top\nabla^2\mathcal{L}_i(\boldsymbol{\theta})\boldsymbol{\theta} \ge \frac{\lambda_{\min}\left\{\boldsymbol{\Sigma}_{0,g}\left(i\Delta_n\right)\right\}}{2}\|\boldsymbol{\theta}\|_2^2 - \frac{\lambda_{\min}\left\{\boldsymbol{\Sigma}_{0,g}\left(i\Delta_n\right)\right\}}{s}\|\boldsymbol{\theta}\|_1^2 \quad \text{for all } \boldsymbol{\theta}\in\mathbb{R}^{p+r}, \tag{A.69}$$

$$\boldsymbol{\theta}^\top\nabla^2\mathcal{L}_i(\boldsymbol{\theta})\boldsymbol{\theta} \le \frac{3\lambda_{\max}\left\{\boldsymbol{\Sigma}_{0,g}\left(i\Delta_n\right)\right\}}{2}\|\boldsymbol{\theta}\|_2^2 + \frac{\lambda_{\min}\left\{\boldsymbol{\Sigma}_{0,g}\left(i\Delta_n\right)\right\}}{s}\|\boldsymbol{\theta}\|_1^2 \quad \text{for all } \boldsymbol{\theta}\in\mathbb{R}^{p+r}, \tag{A.70}$$

which completes the proof. ■

Proof of Theorem 1. By Propositions 4–5, it is enough to show the statement under (A.34), (A.53), and (A.54). From the optimality of $\widehat{\boldsymbol{\theta}}_{i\Delta_n}$, we can show

$$\begin{aligned}
0 &\ge \mathcal{L}_i(\widehat{\boldsymbol{\theta}}_{i\Delta_n}) - \mathcal{L}_i(\boldsymbol{\theta}_0) + \eta\left(\|\widehat{\boldsymbol{\theta}}_{i\Delta_n}\|_1 - \|\boldsymbol{\theta}_0\|_1\right) \\
&\ge \eta\left(\|\widehat{\boldsymbol{\theta}}_{i\Delta_n}\|_1 - \|\boldsymbol{\theta}_0\|_1\right) + \langle\widehat{\boldsymbol{\theta}}_{i\Delta_n} - \boldsymbol{\theta}_0, \nabla\mathcal{L}_i(\boldsymbol{\theta}_0)\rangle + \frac{1}{2}\left(\widehat{\boldsymbol{\theta}}_{i\Delta_n} - \boldsymbol{\theta}_0\right)^\top\nabla^2\mathcal{L}_i(\boldsymbol{\theta})\left(\widehat{\boldsymbol{\theta}}_{i\Delta_n} - \boldsymbol{\theta}_0\right).
\end{aligned}$$

Note that

$$\left\|\widehat{\boldsymbol{\theta}}_{i\Delta_n}\right\|_1 - \|\boldsymbol{\theta}_0\|_1 = \left\|\left(\widehat{\boldsymbol{\theta}}_{i\Delta_n}\right)_{S_\theta^c}\right\|_1 + \left\|\left(\widehat{\boldsymbol{\theta}}_{i\Delta_n}\right)_{S_\theta}\right\|_1 - \left\|(\boldsymbol{\theta}_0)_{S_\theta}\right\|_1$$

$$\geq \left\| \left(\widehat{\boldsymbol{\theta}}_{i\Delta_n} - \boldsymbol{\theta}_0\right)_{S_\theta^c} \right\|_1 - \left\| \left(\widehat{\boldsymbol{\theta}}_{i\Delta_n} - \boldsymbol{\theta}_0\right)_{S_\theta} \right\|_1$$

and

$$\langle \widehat{\boldsymbol{\theta}}_{i\Delta_n} - \boldsymbol{\theta}_0, \nabla \mathcal{L}_i(\boldsymbol{\theta}_0) \rangle \leq \left(\left\| \left(\widehat{\boldsymbol{\theta}}_{i\Delta_n} - \boldsymbol{\theta}_0\right)_{S_\theta} \right\|_1 + \left\| \left(\widehat{\boldsymbol{\theta}}_{i\Delta_n} - \boldsymbol{\theta}_0\right)_{S_\theta^c} \right\|_1 \right) \eta/2,$$

where S_θ is the support of $\boldsymbol{\theta}_0$. Hence, we have

$$\begin{aligned} &\frac{3\eta}{2} \left\| \left(\widehat{\boldsymbol{\theta}}_{i\Delta_n} - \boldsymbol{\theta}_0\right)_{S_\theta} \right\|_1 - \frac{\eta}{2} \left\| \left(\widehat{\boldsymbol{\theta}}_{i\Delta_n} - \boldsymbol{\theta}_0\right)_{S_\theta^c} \right\|_1 \\ &\geq \frac{1}{2} \left(\widehat{\boldsymbol{\theta}}_{i\Delta_n} - \boldsymbol{\theta}_0\right)^\top \nabla^2 \mathcal{L}_i(\boldsymbol{\theta}) \left(\widehat{\boldsymbol{\theta}}_{i\Delta_n} - \boldsymbol{\theta}_0\right) \\ &\geq \frac{\alpha_1}{2} \left\| \widehat{\boldsymbol{\theta}}_{i\Delta_n} - \boldsymbol{\theta}_0 \right\|_2^2 - \frac{\kappa \left(n^{-1/4} + p^{-1}\right) (\log p)^{9/2}}{2} \left\| \widehat{\boldsymbol{\theta}}_{i\Delta_n} - \boldsymbol{\theta}_0 \right\|_1^2 . \end{aligned} \tag{A.71}$$

Also, using the fact that $\left\| \widehat{\boldsymbol{\theta}}_{i\Delta_n} - \boldsymbol{\theta}_0 \right\|_1 \leq C s_p$, we have, for large n,

$$\eta \geq 2\kappa \left(n^{-1/4} + p^{-1}\right) (\log p)^{9/2} \left\| \widehat{\boldsymbol{\theta}}_{i\Delta_n} - \boldsymbol{\theta}_0 \right\|_1 .$$

Therefore, by (A.71), we have

$$\left\| \left(\widehat{\boldsymbol{\theta}}_{i\Delta_n} - \boldsymbol{\theta}_0\right)_{S_\theta^c} \right\|_1 \leq 7 \left\| \left(\widehat{\boldsymbol{\theta}}_{i\Delta_n} - \boldsymbol{\theta}_0\right)_{S_\theta} \right\|_1 ,$$

which implies

$$\left\| \widehat{\boldsymbol{\theta}}_{i\Delta_n} - \boldsymbol{\theta}_0 \right\|_1 \leq 8\sqrt{s_p} \left\| \widehat{\boldsymbol{\theta}}_{i\Delta_n} - \boldsymbol{\theta}_0 \right\|_2 . \tag{A.72}$$

Combining (A.71) and (A.72), we have, for large n,

$$\begin{aligned} 12\sqrt{s_p}\eta \left\| \widehat{\boldsymbol{\theta}}_{i\Delta_n} - \boldsymbol{\theta}_0 \right\|_2 &\geq \frac{3\eta}{2} \left\| \widehat{\boldsymbol{\theta}}_{i\Delta_n} - \boldsymbol{\theta}_0 \right\|_1 \\ &\geq \frac{3\eta}{2} \left\| \left(\widehat{\boldsymbol{\theta}}_{i\Delta_n} - \boldsymbol{\theta}_0\right)_{S_\theta} \right\|_1 - \frac{\eta}{2} \left\| \left(\widehat{\boldsymbol{\theta}}_{i\Delta_n} - \boldsymbol{\theta}_0\right)_{S_\theta^c} \right\|_1 \\ &\geq \frac{\alpha_1}{2} \left\| \widehat{\boldsymbol{\theta}}_{i\Delta_n} - \boldsymbol{\theta}_0 \right\|_2^2 - 32\kappa s_p \left(n^{-1/4} + p^{-1}\right) (\log p)^{9/2} \left\| \widehat{\boldsymbol{\theta}}_{i\Delta_n} - \boldsymbol{\theta}_0 \right\|_2^2 \end{aligned}$$

$$\geq \frac{\alpha_1}{4}\left\|\widehat{\boldsymbol{\theta}}_{i\Delta_n}-\boldsymbol{\theta}_0\right\|_2^2. \tag{A.73}$$

Thus, we obtain

$$\left\|\widehat{\boldsymbol{\theta}}_{i\Delta_n}-\boldsymbol{\theta}_0\right\|_2 \leq C\sqrt{s_p}\eta. \tag{A.74}$$

Then, by (A.72), we have

$$\left\|\widehat{\boldsymbol{\theta}}_{i\Delta_n}-\boldsymbol{\theta}_0\right\|_1 \leq C s_p \eta, \tag{A.75}$$

which completes the proof. ■

A.8 Proof of Theorem 2

Proof of Theorem 2. To obtain the upper bound for $\|\widehat{I\beta}-I\beta_0\|_{\max}$, we first investigate $\widehat{\mathbf{\Omega}}_{i\Delta_n}$. We have

$$\begin{aligned}
&\sup_{0\leq i\leq n-k_2}\left\|\left(\frac{n}{\phi k_1 k_2}\widehat{\mathbf{U}}_i^\top\widehat{\mathbf{U}}_i-\frac{n\zeta}{\phi k_1}\widehat{\mathbf{V}}^X\right)\mathbf{\Omega}_0\left(i\Delta_n\right)-\mathbf{I}\right\|_{\max}\\
&\leq \sup_{0\leq i\leq n-k_2}\left\|\frac{n}{\phi k_1 k_2}\widehat{\mathbf{U}}_i^\top\widehat{\mathbf{U}}_i-\frac{n\zeta}{\phi k_1}\widehat{\mathbf{V}}^X-\mathbf{\Sigma}_{0,u}\left(i\Delta_n\right)\right\|_{\max}\times\sup_{0\leq i\leq n-k_2}\left\|\mathbf{\Omega}_0\left(i\Delta_n\right)\right\|_1.
\end{aligned}$$

By the proofs of (A.42), we can show, with the probability at least $1-p^{-2-a}$,

$$\sup_{0\leq i\leq n-k_2}\left\|\frac{n}{\phi k_1 k_2}\widehat{\mathbf{U}}_i^\top\widehat{\mathbf{U}}_i-\frac{n\zeta}{\phi k_1}\widehat{\mathbf{V}}^X-\mathbf{\Sigma}_{0,u}\left(i\Delta_n\right)\right\|_{\max}\leq C\left\{n^{-1/8}\left(\log p\right)^2+p^{-1/2}\sqrt{\log p}\right\}.$$

Thus, we have

$$\Pr\left\{\sup_{0\leq i\leq n-k_2}\|\widehat{\mathbf{\Omega}}_{i\Delta_n}\|_1\leq C\right\}\geq 1-p^{-2-a}. \tag{A.76}$$

Also, we have, with the probability at least $1-p^{-2-a}$,

$$\sup_{0\leq i\leq n-k_2}\left\|\mathbf{\Sigma}_{0,u}\left(i\Delta_n\right)\widehat{\mathbf{\Omega}}_{i\Delta_n}-\mathbf{I}\right\|_{\max} \leq \sup_{0\leq i\leq n-k_2}\left\|\left(\mathbf{\Sigma}_{0,u}\left(i\Delta_n\right)-\frac{n}{\phi k_1 k_2}\widehat{\mathbf{U}}_i^\top\widehat{\mathbf{U}}_i+\frac{n\zeta}{\phi k_1}\widehat{\mathbf{V}}^X\right)\widehat{\mathbf{\Omega}}_{i\Delta_n}\right\|_{\max}$$

$$
\begin{aligned}
& + \sup_{0 \le i \le n-k_2} \left\| \left(\frac{n}{\phi k_1 k_2} \widehat{\mathbf{U}}_i^\top \widehat{\mathbf{U}}_i - \frac{n\zeta}{\phi k_1} \widehat{\mathbf{V}}^X \right) \widehat{\mathbf{\Omega}}_{i\Delta_n} - \mathbf{I} \right\|_{\max} \\
\le \quad & C \left\{ n^{-1/8} (\log p)^2 + p^{-1/2} \sqrt{\log p} \right\}, \qquad \text{(A.77)}
\end{aligned}
$$

which implies

$$
\begin{aligned}
& \sup_{0 \le i \le n-k_2} \| \widehat{\mathbf{\Omega}}_{i\Delta_n} - \mathbf{\Omega}_0 (i\Delta_n) \|_{\max} \\
& \le \sup_{0 \le i \le n-k_2} \| \mathbf{\Omega}_0 (i\Delta_n) \|_\infty \times \sup_{0 \le i \le n-k_2} \| \mathbf{\Sigma}_{0,u} (i\Delta_n) \widehat{\mathbf{\Omega}}_{i\Delta_n} - \mathbf{I} \|_{\max} \\
& \le C \left\{ n^{-1/8} (\log p)^2 + p^{-1/2} \sqrt{\log p} \right\}.
\end{aligned}
$$

Then, similar to the proofs of Theorem 1 (Kim et al., 2024), we can show

$$
\Pr \left\{ \sup_{0 \le i \le n-k_2} \| \widehat{\mathbf{\Omega}}_{i\Delta_n} - \mathbf{\Omega}_0 (i\Delta_n) \|_1 \le C s_{\omega,p} \tau^{1-q} \right\} \ge 1 - p^{-2-a}. \qquad \text{(A.78)}
$$

Let

$$
\begin{aligned}
\widetilde{\boldsymbol{\beta}}^{(2)}_{i\Delta_n} &= \widehat{\boldsymbol{\beta}}_{i\Delta_n} + \frac{n}{\phi k_1 k_2} \widehat{\mathbf{\Omega}}^\top_{i\Delta_n} \Big[\left(\mathcal{X}_i^c + \mathcal{E}_i^X - \widehat{\mathbf{F}}_i \widehat{\mathbf{B}}^\top_{i\Delta_n} \right)^\top (\mathcal{Y}_i^c + \mathcal{E}_i^Y) \\
& \qquad - \left(\left(\mathcal{X}_i^c + \mathcal{E}_i^X - \widehat{\mathbf{F}}_i \widehat{\mathbf{B}}^\top_{i\Delta_n} \right)^\top (\mathcal{X}_i^c + \mathcal{E}_i^X) - k_2 \zeta \widehat{\mathbf{V}}^X \right) \widehat{\boldsymbol{\beta}}_{i\Delta_n} \Big], \\
\widetilde{\boldsymbol{\beta}}^{(3)}_{i\Delta_n} &= \boldsymbol{\beta}_0 (i\Delta_n) + \frac{n}{\phi k_1 k_2} \widehat{\mathbf{\Omega}}^\top_{i\Delta_n} \Big[\left(\mathbf{U}_i + \mathcal{E}_i^X + \widetilde{\mathbf{F}}_i + \mathbf{F}_i \mathbf{B}_0^\top (i\Delta_n) - \widehat{\mathbf{F}}_i \widehat{\mathbf{B}}^\top_{i\Delta_n} \right)^\top \left(\mathcal{Z}_i + \widetilde{\mathcal{X}}_i + \mathcal{E}_i^Y \right) \\
& \qquad - \left(\left(\mathbf{U}_i + \mathcal{E}_i^X + \widetilde{\mathbf{F}}_i + \mathbf{F}_i \mathbf{B}_0^\top (i\Delta_n) - \widehat{\mathbf{F}}_i \widehat{\mathbf{B}}^\top_{i\Delta_n} \right)^\top \mathcal{E}_i^X - k_2 \zeta \widehat{\mathbf{V}}^X \right) \widehat{\boldsymbol{\beta}}_{i\Delta_n} \Big], \\
\widetilde{\boldsymbol{\beta}}^{(4)}_{i\Delta_n} &= \boldsymbol{\beta}_0 (i\Delta_n) + \frac{n}{\phi k_1 k_2} \mathbf{\Omega}_0 (i\Delta_n) \Big[\left(\mathbf{U}_i + \mathcal{E}_i^X + \widetilde{\mathbf{F}}_i + \mathbf{F}_i \mathbf{B}_0^\top (i\Delta_n) - \widehat{\mathbf{F}}_i \widehat{\mathbf{B}}^\top_{i\Delta_n} \right)^\top \\
& \times \left(\mathcal{Z}_i + \widetilde{\mathcal{X}}_i + \mathcal{E}_i^Y \right) - \left(\left(\mathbf{U}_i + \mathcal{E}_i^X + \widetilde{\mathbf{F}}_i + \mathbf{F}_i \mathbf{B}_0^\top (i\Delta_n) - \widehat{\mathbf{F}}_i \widehat{\mathbf{B}}^\top_{i\Delta_n} \right)^\top \mathcal{E}_i^X - k_2 \zeta \mathbf{V}^X \right) \boldsymbol{\beta}_0 (i\Delta_n) \Big].
\end{aligned}
$$

Then, we have

$$
\| \widehat{I\beta} - I\beta_0 \|_{\max} \le \left\| \sum_{i=0}^{[1/(k_2\Delta_n)]-1} \left(\widetilde{\boldsymbol{\beta}}_{ik_2\Delta_n} - \widetilde{\boldsymbol{\beta}}^{(2)}_{ik_2\Delta_n} \right) k_2 \Delta_n \right\|_\infty
$$

$$
\begin{aligned}
&+\left\|\sum_{i=0}^{[1/(k_2\Delta_n)]-1}\left(\widetilde{\boldsymbol{\beta}}^{(2)}_{ik_2\Delta_n}-\widetilde{\boldsymbol{\beta}}^{(3)}_{ik_2\Delta_n}\right)k_2\Delta_n\right\|_\infty\\
&+\left\|\sum_{i=0}^{[1/(k_2\Delta_n)]-1}\left(\widetilde{\boldsymbol{\beta}}^{(3)}_{ik_2\Delta_n}-\widetilde{\boldsymbol{\beta}}^{(4)}_{ik_2\Delta_n}\right)k_2\Delta_n\right\|_\infty\\
&+\left\|\sum_{i=0}^{[1/(k_2\Delta_n)]-1}\left(\widetilde{\boldsymbol{\beta}}^{(4)}_{ik_2\Delta_n}-\boldsymbol{\beta}_0(ik_2\Delta_n)\right)k_2\Delta_n\right\|_\infty\\
&+\left\|\sum_{i=0}^{[1/(k_2\Delta_n)]-1}\int_{ik_2\Delta_n}^{(i+1)k_2\Delta_n}\left(\boldsymbol{\beta}_0(ik_2\Delta_n)-\boldsymbol{\beta}_0(t)\right)dt\right\|_\infty\\
&+\left\|\int_{[1/(k_2\Delta_n)]k_2\Delta_n}^{1}\boldsymbol{\beta}_0(t)dt\right\|_\infty\\
=\;&(I)+(II)+(III)+(IV)+(V)+(VI). \qquad \text{(A.79)}
\end{aligned}
$$

Consider (I). By (3.12), (A.13), Proposition 3, and (A.76), we have

$$
\Pr\left\{(I)\le Cs_pn^{-1/4}(\log p)^2\right\}\ge 1-p^{-1-a}. \qquad \text{(A.80)}
$$

Consider (II). We have

$$
\begin{aligned}
&\sup_{0\le i\le n-k_2}\left\|\frac{n}{\phi k_1k_2}\left(\mathbf{U}_i+\mathcal{E}_i^X+\widetilde{\mathbf{F}}_i+\mathbf{F}_i\mathbf{B}_0^\top(i\Delta_n)-\widehat{\mathbf{F}}_i\widehat{\mathbf{B}}_{i\Delta_n}^\top\right)^\top\right.\\
&\qquad\qquad\left.\times\left(\mathbf{U}_i+\mathbf{F}_i\mathbf{B}_0^\top(i\Delta_n)+\widetilde{\mathbf{F}}_i\right)-\boldsymbol{\Sigma}_{0,u}(i\Delta_n)\right\|_{\max}\\
\le&\sup_{0\le i\le n-k_2}\left\|\frac{n}{\phi k_1k_2}\mathbf{U}_i^\top\mathbf{U}_i-\boldsymbol{\Sigma}_{0,u}(i\Delta_n)\right\|_{\max}\\
&+\sup_{0\le i\le n-k_2}\left\|\frac{n}{\phi k_1k_2}\left(\mathbf{U}_i^\top\mathbf{F}_i\mathbf{B}_0^\top(i\Delta_n)+\left(\mathcal{E}_i^X\right)^\top\mathbf{U}_i+\left(\mathcal{E}_i^X\right)^\top\mathbf{F}_i\mathbf{B}_0^\top(i\Delta_n)\right)\right\|_{\max}\\
&+\sup_{0\le i\le n-k_2}\left\|\frac{n}{\phi k_1k_2}\left(2\mathbf{U}_i^\top\widetilde{\mathbf{F}}_i+\left(\mathcal{E}_i^X\right)^\top\widetilde{\mathbf{F}}_i+\widetilde{\mathbf{F}}_i^\top\mathbf{F}_i\mathbf{B}_0^\top(i\Delta_n)+\widetilde{\mathbf{F}}_i^\top\widetilde{\mathbf{F}}_i\right)\right\|_{\max}\\
&+\sup_{0\le i\le n-k_2}\left\|\frac{n}{\phi k_1k_2}\left(\mathbf{F}_i\mathbf{B}_0^\top(i\Delta_n)-\widehat{\mathbf{F}}_i\widehat{\mathbf{B}}_{i\Delta_n}^\top\right)^\top\left(\mathbf{U}_i+\mathbf{F}_i\mathbf{B}_0^\top(i\Delta_n)+\widetilde{\mathbf{F}}_i\right)\right\|_{\max}\\
=&(II)^{(1)}+(II)^{(2)}+(II)^{(3)}+(II)^{(4)}.
\end{aligned}
$$

By Proposition 2, we have

$$\Pr\left\{(II)^{(1)} \leq Cn^{-1/8}\sqrt{\log p}\right\} \geq 1 - p^{-2-a}.$$

For $(II)^{(2)}$, note that the elements of $\mathbf{U}_i$, $\mathbf{F}_i$, and $\mathcal{E}_i^X$ have sub-Gaussian tails. Thus, by the Bernstein's inequality for martingales, we have

$$\Pr\left\{(II)^{(2)} \leq Cn^{-1/8}\sqrt{\log p}\right\} \geq 1 - p^{-2-a}.$$

Also, by (A.13), we have

$$\Pr\left\{(II)^{(3)} \leq Cn^{-1/8}\left(\log p\right)^{3/2}\right\} \geq 1 - p^{-2-a}.$$

For $(II)^{(4)}$, let $\mathbf{F}_i\mathbf{B}_0^\top\left(i\Delta_n\right) - \widehat{\mathbf{F}}_i\widehat{\mathbf{B}}_{i\Delta_n}^\top = \left(T_{jl}^{(i)}\right)_{1\leq j\leq k_2-k_1+1, 1\leq l\leq p}$. By the proofs of (A.37), we have

$$\Pr\left(\sup_{0\leq i\leq n-k_2}\max_{1\leq l\leq p}\sum_{j=1}^{k_2-k_1+1}\left(T_{jl}^{(i)}\right)^2 \leq C\left\{\left(\log p\right)^3 + p^{-1}n^{1/4}\right\}\right) \geq 1 - p^{-2-a}.$$

Then, from the Cauchy-Schwarz inequality and (A.13), we have, with the probability at least $1 - 2p^{-2-a}$,

$$\begin{aligned}(II)^{(4)} &\leq \frac{Cn}{\phi k_1 k_2}\sqrt{\sup_{0\leq i\leq n-k_2}\max_{1\leq l\leq p}\sum_{j=1}^{k_2-k_1+1}\left(T_{jl}^{(i)}\right)^2 \times \left(k_2 - k_1 + 1\right)n^{-1/2}\log p} \\ &\leq C\left\{n^{-1/8}\left(\log p\right)^2 + p^{-1/2}\sqrt{\log p}\right\}.\end{aligned}$$

Thus, we have, with the probability at least $1 - 5p^{-2-a}$,

$$\sup_{0\leq i\leq n-k_2}\left\|\frac{n}{\phi k_1 k_2}\left(\mathbf{U}_i + \mathcal{E}_i^X + \widetilde{\mathbf{F}}_i + \mathbf{F}_i\mathbf{B}_0^\top\left(i\Delta_n\right) - \widehat{\mathbf{F}}_i\widehat{\mathbf{B}}_{i\Delta_n}^\top\right)^\top\right.$$

$$\times \left(\mathbf{U}_i + \mathbf{F}_i \mathbf{B}_0^\top (i\Delta_n) + \widetilde{\mathbf{F}}_i\right) - \boldsymbol{\Sigma}_{0,u}(i\Delta_n)\Big\|_{\max}$$
$$\leq C\left\{n^{-1/8}(\log p)^2 + p^{-1/2}\sqrt{\log p}\right\}. \tag{A.81}$$

Combining (A.76), (A.77), and (A.81), we have, with the probability at least $1 - 6p^{-2-a}$,

$$\sup_{0\leq i\leq n-k_2}\left\|\frac{n}{\phi k_1 k_2}\widehat{\boldsymbol{\Omega}}_{i\Delta_n}^\top\left(\mathbf{U}_i + \mathcal{E}_i^X + \widetilde{\mathbf{F}}_i + \mathbf{F}_i\mathbf{B}_0^\top(i\Delta_n) - \widehat{\mathbf{F}}_i\widehat{\mathbf{B}}_{i\Delta_n}^\top\right)^\top \right.$$
$$\left.\times\left(\mathbf{U}_i + \mathbf{F}_i\mathbf{B}_0^\top(i\Delta_n) + \widetilde{\mathbf{F}}_i\right) - \mathbf{I}\right\|_{\max}$$
$$\leq C\left\{n^{-1/8}(\log p)^2 + p^{-1/2}\sqrt{\log p}\right\}.$$

Then, from (3.12), we have, with the probability at least $1 - p^{-1-a}$,

$$\begin{aligned}(II) \quad \leq \quad & \sup_{0\leq i\leq n-k_2}\left\|\frac{n}{\phi k_1 k_2}\widehat{\boldsymbol{\Omega}}_{i\Delta_n}^\top\left(\mathbf{U}_i + \mathcal{E}_i^X + \widetilde{\mathbf{F}}_i + \mathbf{F}_i\mathbf{B}_0^\top(i\Delta_n) - \widehat{\mathbf{F}}_i\widehat{\mathbf{B}}_{i\Delta_n}^\top\right)^\top\right. \\ & \left.\times\left(\mathbf{U}_i + \mathbf{F}_i\mathbf{B}_0^\top(i\Delta_n) + \widetilde{\mathbf{F}}_i\right) - \mathbf{I}\right\|_{\max} \\ & \times \sup_{0\leq i\leq n-k_2}\left\|\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0(i\Delta_n)\right\|_1 \\ \leq \quad & C\left\{s_p^2 n^{-1/4}(\log p)^4 + p^{-1}s_p^2\log p\right\}.\end{aligned} \tag{A.82}$$

Consider (III). We have

$$\begin{aligned}&(III) \\ &\leq \sup_{0\leq i\leq n-k_2}\left\|\widehat{\boldsymbol{\Omega}}_{i\Delta_n}^\top\right\|_\infty \times \sup_{0\leq i\leq n-k_2}\left\|\frac{n\zeta}{\phi k_1}\left(\widehat{\mathbf{V}}^X - \mathbf{V}^X\right)\right\|_{\max} \times \sup_{0\leq i\leq n-k_2}\left\|\widehat{\boldsymbol{\beta}}_{i\Delta_n}\right\|_1 \\ &\quad + \sup_{0\leq i\leq n-k_2}\left\|\widehat{\boldsymbol{\Omega}}_{i\Delta_n}^\top - \boldsymbol{\Omega}_0(i\Delta_n)\right\|_\infty \\ &\quad \times \sup_{0\leq i\leq n-k_2}\left\|\frac{n}{\phi k_1 k_2}\left(\mathbf{U}_i + \mathcal{E}_i^X + \widetilde{\mathbf{F}}_i + \mathbf{F}_i\mathbf{B}_0^\top(i\Delta_n) - \widehat{\mathbf{F}}_i\widehat{\mathbf{B}}_{i\Delta_n}^\top\right)^\top\left(\mathcal{Z}_i + \widetilde{\mathcal{X}}_i + \mathcal{E}_i^Y\right)\right\|_{\max} \\ &\quad + \sup_{0\leq i\leq n-k_2}\left\|\widehat{\boldsymbol{\Omega}}_{i\Delta_n}^\top - \boldsymbol{\Omega}_0(i\Delta_n)\right\|_\infty \\ &\quad \times \sup_{0\leq i\leq n-k_2}\left\|\frac{n}{\phi k_1 k_2}\left(\mathbf{U}_i + \mathcal{E}_i^X + \widetilde{\mathbf{F}}_i + \mathbf{F}_i\mathbf{B}_0^\top(i\Delta_n) - \widehat{\mathbf{F}}_i\widehat{\mathbf{B}}_{i\Delta_n}^\top\right)^\top\mathcal{E}_i^X - \frac{n\zeta}{\phi k_1}\mathbf{V}^X\right\|_{\max} \\ &\quad \times \sup_{0\leq i\leq n-k_2}\left\|\widehat{\boldsymbol{\beta}}_{i\Delta_n}\right\|_1\end{aligned}$$

$$
\begin{aligned}
&+ \sup_{0\le i\le n-k_2} \left\|\boldsymbol{\Omega}_0\left(i\Delta_n\right)\right\|_\infty \\
&\quad\times \sup_{0\le i\le n-k_2}\left\|\frac{n}{\phi k_1 k_2}\left(\mathbf{U}_i+\mathcal{E}_i^X+\widetilde{\mathbf{F}}_i+\mathbf{F}_i\mathbf{B}_0^\top\left(i\Delta_n\right)-\widehat{\mathbf{F}}_i\widehat{\mathbf{B}}_{i\Delta_n}^\top\right)^\top\mathcal{E}_i^X-\frac{n\zeta}{\phi k_1}\mathbf{V}^X\right\|_{\max} \\
&\quad\times \sup_{0\le i\le n-k_2}\left\|\widehat{\boldsymbol{\beta}}_{i\Delta_n}-\boldsymbol{\beta}_0\left(i\Delta_n\right)\right\|_1 .
\end{aligned}
$$

By Proposition 2, we have

$$
\Pr\left\{\sup_{0\le i\le n-k_2}\left\|\frac{n\zeta}{\phi k_1}\left(\widehat{\mathbf{V}}^X-\mathbf{V}^X\right)\right\|_{\max}\le Cn^{-1/2}\sqrt{\log p}\right\}\ge 1-p^{-2-a}.
$$

Similar to the proofs of (A.81), we can show, with the probability at least $1-p^{-2-a}$,

$$
\begin{aligned}
&\sup_{0\le i\le n-k_2}\left\|\frac{n}{\phi k_1 k_2}\left(\mathbf{U}_i+\mathcal{E}_i^X+\widetilde{\mathbf{F}}_i+\mathbf{F}_i\mathbf{B}_0^\top\left(i\Delta_n\right)-\widehat{\mathbf{F}}_i\widehat{\mathbf{B}}_{i\Delta_n}^\top\right)^\top\left(\mathcal{Z}_i+\widetilde{\mathcal{X}}_i+\mathcal{E}_i^Y\right)\right\|_{\max} \\
&\le C\left\{s_p n^{-1/8}\left(\log p\right)^{3/2}+n^{-1/8}\left(\log p\right)^2+p^{-1/2}\sqrt{\log p}\right\}\quad \text{and} \\
&\sup_{0\le i\le n-k_2}\left\|\frac{n}{\phi k_1 k_2}\left(\mathbf{U}_i+\mathcal{E}_i^X+\widetilde{\mathbf{F}}_i+\mathbf{F}_i\mathbf{B}_0^\top\left(i\Delta_n\right)-\widehat{\mathbf{F}}_i\widehat{\mathbf{B}}_{i\Delta_n}^\top\right)^\top\mathcal{E}_i^X-\frac{n\zeta}{\phi k_1}\mathbf{V}^X\right\|_{\max} \\
&\le C\left\{n^{-1/8}\left(\log p\right)^2+p^{-1/2}\sqrt{\log p}\right\}.
\end{aligned}
$$

Then, from (3.12), (A.76), and (A.78), we have

$$
\begin{aligned}
&\Pr\Big((III)\le C\left\{s_p^2 n^{-1/4}\left(\log p\right)^4+s_p s_{\omega,p}n^{(-2+q)/8}\left(\log p\right)^{4-2q}+p^{-1}s_p^2\log p\right. \\
&\qquad\left.+p^{(-2+q)/2}s_p s_{\omega,p}(\log p)^{(2-q)/2}+p^{-1/2}s_p s_{\omega,p}n^{(-1+q)/8}(\log p)^{(5-4q)/2}\right\}\Big) \\
&\ge 1-p^{-1-a}. \qquad\qquad (\text{A.83})
\end{aligned}
$$

Consider (IV). We have

$$
\begin{aligned}
(IV)\ \le\ Cn^{-1/2}\Bigg\|\sum_{i=0}^{[1/(k_2\Delta_n)]-1}&\boldsymbol{\Omega}_0\left(ik_2\Delta_n\right)\Big[\left(\mathbf{U}_{ik_2}+\mathcal{E}_{ik_2}^X\right)^\top\left(\mathcal{Z}_{ik_2}+\mathcal{E}_{ik_2}^Y\right) \\
&-\left(\left(\mathbf{U}_{ik_2}+\mathcal{E}_{ik_2}^X\right)^\top\mathcal{E}_{ik_2}^X-k_2\zeta\mathbf{V}^X\right)\boldsymbol{\beta}_0\left(ik_2\Delta_n\right)\Big]\Bigg\|_\infty
\end{aligned}
$$

$$
\begin{aligned}
&+Cn^{-1/2}\left\|\sum_{i=0}^{[1/(k_2\Delta_n)]-1}\mathbf{\Omega}_0\left(ik_2\Delta_n\right)\Big[\left(\mathbf{U}_{ik_2}+\mathcal{E}^X_{ik_2}\right)^\top\widetilde{\mathcal{X}}_{ik_2}\right.\\
&\qquad\qquad\left.+\widetilde{\mathbf{F}}^\top_{ik_2}\left(\mathcal{Z}_{ik_2}+\mathcal{E}^Y_{ik_2}+\mathcal{E}^X_{ik_2}\boldsymbol{\beta}_0\left(ik_2\Delta_n\right)\right)\Big]\right\|_\infty\\
&+Cn^{-1/2}\left\|\sum_{i=0}^{[1/(k_2\Delta_n)]-1}\mathbf{\Omega}_0\left(ik_2\Delta_n\right)\Big[\left(\widetilde{\mathbf{F}}_{ik_2}+\mathbf{F}_{ik_2}\mathbf{B}_0^\top\left(ik_2\Delta_n\right)-\widehat{\mathbf{F}}_{ik_2}\widehat{\mathbf{B}}^\top_{ik_2\Delta_n}\right)^\top\widetilde{\mathcal{X}}_{ik_2}\Big]\right\|_\infty\\
&+Cn^{-1/2}\left\|\sum_{i=0}^{[1/(k_2\Delta_n)]-1}\mathbf{\Omega}_0\left(ik_2\Delta_n\right)\Big[\left(\mathbf{F}_{ik_2}\mathbf{B}_0^\top\left(ik_2\Delta_n\right)-\widehat{\mathbf{F}}_{ik_2}\widehat{\mathbf{B}}^\top_{ik_2\Delta_n}\right)^\top\left(\mathcal{Z}_{ik_2}+\mathcal{E}^Y_{ik_2}\right)\right.\\
&\qquad\qquad\left.-\left(\mathbf{F}_{ik_2}\mathbf{B}_0^\top\left(ik_2\Delta_n\right)-\widehat{\mathbf{F}}_{ik_2}\widehat{\mathbf{B}}^\top_{ik_2\Delta_n}\right)^\top\mathcal{E}^X_{ik_2}\boldsymbol{\beta}_0\left(ik_2\Delta_n\right)\Big]\right\|_\infty\\
=\;&(IV)^{(1)}+(IV)^{(2)}+(IV)^{(3)}+(IV)^{(4)}.
\end{aligned}
$$

For $(IV)^{(1)}$, note that each element of $\mathbf{U}_i$, $\mathcal{E}^X_i$, $\mathcal{Z}_i$, and $\mathcal{E}^Y_i$ has sub-Gaussian tail. Thus, by Bernstein's inequality for martingales, we have, with the probability at least $1-p^{-2-a}$,

$$
(IV)^{(1)}\le Cs_pn^{-1/4}\sqrt{\log p}.
$$

Consider $(IV)^{(2)}$. Let

$$
\chi_1=\left\{\sup_{0\le i\le n-k_2}\sup_{t\in[i\Delta_n,(i+k_2)\Delta_n]}\sum_{j=1}^p\left|\beta_{0,j}(t)-\beta_{0,j}(i\Delta_n)\right|\le Cs_pn^{-1/8}\sqrt{\log p}\right\}.
$$

By the proofs of (A.13), we have

$$
\Pr\left(\chi_1\right)\ge 1-p^{-2-a}.
$$

Under the event χ_1, each element of $\mathbf{U}_i+\mathcal{E}^X_i$ and $\widetilde{\mathcal{X}}_i$ has sub-Gaussian tail with the order of $n^{-1/4}$ and $s_pn^{-3/8}\sqrt{\log p}$, respectively. Thus, by Bernstein's inequality for martingales, we can show, with the probability at least $1-2p^{-2-a}$,

$$
\left\|\sum_{i=0}^{[1/(k_2\Delta_n)]-1}\mathbf{\Omega}_0\left(ik_2\Delta_n\right)\left(\mathbf{U}_{ik_2}+\mathcal{E}^X_{ik_2}\right)^\top\widetilde{\mathcal{X}}_{ik_2}\right\|_\infty\le Cs_pn^{1/4}\left(\log p\right)^{3/2}.
$$

Similarly, let

$$\chi_2 = \left\{ \sup_{0\le i\le n-k_2} \sup_{t\in[i\Delta_n,(i+k_2)\Delta_n]} \|\mathbf{B}_0(t) - \mathbf{B}_0(i\Delta_n)\|_\infty \le Cn^{-1/8}\sqrt{\log p} \right\}.$$

We have

$$\Pr\left(\chi_2\right) \ge 1 - p^{-2-a}.$$

Then, we can show, with the probability at least $1 - 2p^{-2-a}$,

$$\left\| \sum_{i=0}^{[1/(k_2\Delta_n)]-1} \mathbf{\Omega}_0\left(ik_2\Delta_n\right) \widetilde{\mathbf{F}}_{ik_2}^\top \left(\mathcal{Z}_{ik_2} + \mathcal{E}_{ik_2}^Y + \mathcal{E}_{ik_2}^X \boldsymbol{\beta}_0\left(ik_2\Delta_n\right)\right) \right\|_\infty \le Cn^{1/4}\left(\log p\right)^{3/2},$$

which implies

$$\Pr\left\{ (IV)^{(2)} \le Cs_p n^{-1/4}\left(\log p\right)^{3/2} \right\} \ge 1 - 4p^{-2-a}.$$

Consider $(IV)^{(3)}$. Recall that

$$\Pr\left(\sup_{0\le i\le n-k_2} \max_{1\le l\le p} \sum_{j=1}^{k_2-k_1+1} \left(T_{jl}^{(i)}\right)^2 \le C\left\{\left(\log p\right)^3 + p^{-1}n^{1/4}\right\} \right) \ge 1 - p^{-2-a},$$

where $\mathbf{F}_i\mathbf{B}_0^\top\left(i\Delta_n\right) - \widehat{\mathbf{F}}_i\widehat{\mathbf{B}}_{i\Delta_n}^\top = \left(T_{jl}^{(i)}\right)_{1\le j\le k_2-k_1+1, 1\le l\le p}$. Thus, from the Cauchy-Schwarz inequality and (A.13), we have, with the probability at least $1 - 2p^{-2-a}$,

$$\begin{aligned} (IV)^{(3)} &\le Cn^{-1/4}\sqrt{\left\{\left(\log p\right)^3 + p^{-1}n^{1/4}\right\} \times \left(k_2 - k_1 + 1\right)\left(s_p n^{-3/8}\log p\right)^2} \\ &\le C\left\{ s_p n^{-1/4}\left(\log p\right)^{5/2} + p^{-1/2}s_p n^{-1/8}\log p \right\}. \end{aligned}$$

For $(IV)^{(4)}$, we first consider

$$\left\| \sum_{i=0}^{[1/(k_2\Delta_n)]-1} \mathbf{\Omega}_0\left(ik_2\Delta_n\right) \left(\mathbf{F}_{ik_2}\mathbf{B}_0^\top\left(ik_2\Delta_n\right) - \widehat{\mathbf{F}}_{ik_2}\widehat{\mathbf{B}}_{ik_2\Delta_n}^\top\right)^\top \mathcal{E}_{ik_2}^X \boldsymbol{\beta}_0\left(ik_2\Delta_n\right) \right\|_\infty.$$

By Proposition 3, we have, with the probability at least $1 - p^{-2-a}$,

$$
\begin{aligned}
& \mathbf{F}_i \left(\mathbf{H}_i^\top\right)^{-1} - \widehat{\mathbf{F}}_i \\
&= \mathbf{F}_i \left(\mathbf{H}_i^\top\right)^{-1} - p^{-1}\left(\mathbf{F}_i \mathbf{B}_0^\top (i\Delta_n) + \mathbf{U}_i + \widetilde{\mathbf{F}}_i + \mathcal{E}_i^X\right) \widehat{\mathbf{B}}_{i\Delta_n} - p^{-1}\left(\mathcal{X}_i - \mathcal{X}_i^c - \mathcal{E}_i^X\right) \widehat{\mathbf{B}}_{i\Delta_n} \\
&= \mathbf{F}_i \left(\mathbf{H}_i^\top\right)^{-1}\left(\mathbf{I}_r - p^{-1}\mathbf{H}_i^\top \mathbf{B}_0^\top (i\Delta_n) \widehat{\mathbf{B}}_{i\Delta_n}\right) - p^{-1}\left(\mathbf{U}_i + \widetilde{\mathbf{F}}_i + \mathcal{E}_i^X\right) \widehat{\mathbf{B}}_{i\Delta_n} \\
&\quad - p^{-1}\left(\mathcal{X}_i - \mathcal{X}_i^c - \mathcal{E}_i^X\right) \widehat{\mathbf{B}}_{i\Delta_n} \\
&= p^{-1}\mathbf{F}_i \left(\mathbf{H}_i^\top\right)^{-1}\left(\widehat{\mathbf{B}}_{i\Delta_n} - \mathbf{B}_0 (i\Delta_n) \mathbf{H}_i\right)^\top \widehat{\mathbf{B}}_{i\Delta_n} - p^{-1}\left(\mathbf{U}_i + \widetilde{\mathbf{F}}_i + \mathcal{E}_i^X\right) \widehat{\mathbf{B}}_{i\Delta_n} \\
&\quad - p^{-1}\left(\mathcal{X}_i - \mathcal{X}_i^c - \mathcal{E}_i^X\right) \widehat{\mathbf{B}}_{i\Delta_n}
\end{aligned}
$$

for all $0 \le i \le n - k_2$. Thus, we have, with the probability at least $1 - p^{-2-a}$,

$$
\begin{aligned}
& \left\| \sum_{i=0}^{[1/(k_2\Delta_n)]-1} \boldsymbol{\Omega}_0 (ik_2\Delta_n) \left(\mathbf{F}_{ik_2}\mathbf{B}_0^\top (ik_2\Delta_n) - \widehat{\mathbf{F}}_{ik_2}\widehat{\mathbf{B}}_{ik_2\Delta_n}^\top\right)^\top \mathcal{E}_{ik_2}^X \boldsymbol{\beta}_0 (ik_2\Delta_n) \right\|_\infty \\
&= \left\| \sum_{i=0}^{[1/(k_2\Delta_n)]-1} \boldsymbol{\Omega}_0 (ik_2\Delta_n) \left(\mathbf{F}_{ik_2}\left(\mathbf{H}_{ik_2}^\top\right)^{-1}\left[\left(\mathbf{B}_0 (ik_2\Delta_n) \mathbf{H}_{ik_2}\right)^\top - \widehat{\mathbf{B}}_{ik_2\Delta_n}^\top\right] \right.\right. \\
&\qquad \left.\left. + \left[\mathbf{F}_{ik_2}\left(\mathbf{H}_{ik_2}^\top\right)^{-1} - \widehat{\mathbf{F}}_{ik_2}\right]\widehat{\mathbf{B}}_{ik_2\Delta_n}^\top\right)^\top \mathcal{E}_{ik_2}^X \boldsymbol{\beta}_0 (ik_2\Delta_n) \right\|_\infty \\
&\le \left\| \sum_{i=0}^{[1/(k_2\Delta_n)]-1} \boldsymbol{\Omega}_0 (ik_2\Delta_n) \left(\widehat{\mathbf{B}}_{ik_2\Delta_n} - \mathbf{B}_0 (ik_2\Delta_n) \mathbf{H}_{ik_2}\right) \mathbf{H}_{ik_2}^{-1}\mathbf{F}_{ik_2}^\top \mathcal{E}_{ik_2}^X \boldsymbol{\beta}_0 (ik_2\Delta_n) \right\|_\infty \\
&\quad + p^{-1}\left\| \sum_{i=0}^{[1/(k_2\Delta_n)]-1} \boldsymbol{\Omega}_0 (ik_2\Delta_n) \widehat{\mathbf{B}}_{ik_2\Delta_n}\widehat{\mathbf{B}}_{ik_2\Delta_n}^\top \left(\widehat{\mathbf{B}}_{ik_2\Delta_n} - \mathbf{B}_0 (ik_2\Delta_n) \mathbf{H}_{ik_2}\right) \right. \\
&\qquad \left. \times \mathbf{H}_{ik_2}^{-1}\mathbf{F}_{ik_2}^\top \mathcal{E}_{ik_2}^X \boldsymbol{\beta}_0 (ik_2\Delta_n) \right\|_\infty \\
&\quad + p^{-1}\left\| \sum_{i=0}^{[1/(k_2\Delta_n)]-1} \boldsymbol{\Omega}_0 (ik_2\Delta_n) \left(\widehat{\mathbf{B}}_{ik_2\Delta_n}\widehat{\mathbf{B}}_{ik_2\Delta_n}^\top - \mathbf{B}_0 (ik_2\Delta_n) \mathbf{B}_0^\top (ik_2\Delta_n)\right) \right. \\
&\qquad \left. \times \left(\mathbf{U}_{ik_2} + \widetilde{\mathbf{F}}_{ik_2} + \mathcal{E}_{ik_2}^X\right)^\top \mathcal{E}_{ik_2}^X \boldsymbol{\beta}_0 (ik_2\Delta_n) \right\|_\infty \\
&\quad + p^{-1}\left\| \sum_{i=0}^{[1/(k_2\Delta_n)]-1} \boldsymbol{\Omega}_0 (ik_2\Delta_n) \mathbf{B}_0 (ik_2\Delta_n) \mathbf{B}_0^\top (ik_2\Delta_n) \right. \\
&\qquad \left. \times \left[\left(\mathbf{U}_{ik_2} + \widetilde{\mathbf{F}}_{ik_2} + \mathcal{E}_{ik_2}^X\right)^\top \mathcal{E}_{ik_2}^X - k_2\zeta\mathbf{V}^X\right] \boldsymbol{\beta}_0 (ik_2\Delta_n) \right\|_\infty
\end{aligned}
$$

$$
\begin{aligned}
&+p^{-1}k_2\zeta\left\|\sum_{i=0}^{[1/(k_2\Delta_n)]-1}\mathbf{\Omega}_0\left(ik_2\Delta_n\right)\mathbf{B}_0\left(ik_2\Delta_n\right)\mathbf{B}_0^\top\left(ik_2\Delta_n\right)\mathbf{V}^X\boldsymbol{\beta}_0\left(ik_2\Delta_n\right)\right\|_\infty \\
&+p^{-1}\left\|\sum_{i=0}^{[1/(k_2\Delta_n)]-1}\mathbf{\Omega}_0\left(ik_2\Delta_n\right)\widehat{\mathbf{B}}_{ik_2\Delta_n}\widehat{\mathbf{B}}_{ik_2\Delta_n}^\top\left(\mathcal{X}_{ik_2}-\mathcal{X}_{ik_2}^c-\mathcal{E}_{ik_2}^X\right)^\top\mathcal{E}_{ik_2}^X\boldsymbol{\beta}_0\left(ik_2\Delta_n\right)\right\|_\infty \\
&=(A)+(B)+(C)+(D)+(E)+(F).
\end{aligned}
$$

Consider (A)–(C). By the proofs of Proposition 2, we have, with the probability at least $1-p^{-2-a}$,

$$
\begin{aligned}
&\max_i\left\|\mathbf{F}_i^\top\mathcal{E}_i^X\boldsymbol{\beta}_0\left(ik_2\Delta_n\right)\right\|_\infty\le Cs_pn^{1/8}\sqrt{\log p}\quad\text{and} \\
&\max_i\left\|\left(\mathbf{U}_i+\widetilde{\mathbf{F}}_i+\mathcal{E}_i^X\right)^\top\mathcal{E}_i^X\boldsymbol{\beta}_0\left(ik_2\Delta_n\right)\right\|_1\le C\left\{ps_pn^{1/8}\sqrt{\log p}+s_pn^{1/4}\right\}.
\end{aligned}
$$

Then, by Proposition 3, we can show, with the probability at least $1-2p^{-2-a}$,

$$
(A)+(B)+(C)\le C\left\{s_pn^{1/4}\left(\log p\right)^2+p^{-1/2}s_pn^{3/8}\sqrt{\log p}+p^{-3/2}s_pn^{1/2}\right\}.
$$

Consider (D). Under the event χ_2, each element of $\widetilde{\mathcal{F}}_i$ has sub-Gaussian tail with the order of $n^{-3/8}\sqrt{\log p}$. Also, each element of $\mathbf{U}_i$ and $\mathcal{E}_i^X$ has sub-Gaussian tail with the order of $n^{-1/4}$. Thus, from Bernstein's inequality for martingales, we can show, with the probability at least $1-2p^{-2-a}$,

$$
(D)\le Cs_pn^{1/4}\sqrt{\log p}.
$$

For (E), by Assumption 1(g), we have

$$
(E)\le Cs_pn^{1/4}\text{ a.s.}
$$

For (F), by (A.13) and Proposition 3, we have, with the probability at least $1-2p^{-2-a}$,

$$
(F)\le Cn^{1/4}\left(\log p\right)^2.
$$

Thus, we have, with the probability at least $1 - 7p^{-2-a}$,

$$\left\| \sum_{i=0}^{[1/(k_2\Delta_n)]-1} \boldsymbol{\Omega}_0 (ik_2\Delta_n) \left(\mathbf{F}_{ik_2} \mathbf{B}_0^\top (ik_2\Delta_n) - \widehat{\mathbf{F}}_{ik_2} \widehat{\mathbf{B}}_{ik_2\Delta_n}^\top \right)^\top \mathcal{E}_{ik_2}^X \boldsymbol{\beta}_0 (ik_2\Delta_n) \right\|_\infty$$
$$\leq C \left\{ s_p n^{1/4} (\log p)^2 + p^{-1/2} s_p n^{3/8} \sqrt{\log p} + p^{-3/2} s_p n^{1/2} \right\}.$$

Similarly, we can show, with the probability at least $1 - 7p^{-2-a}$,

$$\left\| \sum_{i=0}^{[1/(k_2\Delta_n)]-1} \boldsymbol{\Omega}_0 (ik_2\Delta_n) \left(\mathbf{F}_{ik_2} \mathbf{B}_0^\top (ik_2\Delta_n) - \widehat{\mathbf{F}}_{ik_2} \widehat{\mathbf{B}}_{ik_2\Delta_n}^\top \right)^\top \left(\mathcal{Z}_{ik_2} + \mathcal{E}_{ik_2}^Y \right) \right\|_\infty$$
$$\leq C \left\{ n^{1/4} (\log p)^2 + p^{-1/2} n^{3/8} \sqrt{\log p} \right\},$$

which implies

$$\Pr \left((IV)^{(4)} \leq C \left\{ s_p n^{-1/4} (\log p)^2 + p^{-1/2} s_p n^{-1/8} \sqrt{\log p} + p^{-3/2} s_p \right\} \right) \geq 1 - 14p^{-2-a}$$

and

$$\Pr \left((IV) \leq C \left\{ s_p n^{-1/4} (\log p)^{5/2} + p^{-1/2} s_p n^{-1/8} \log p + p^{-3/2} s_p \right\} \right) \geq 1 - p^{-1-a}. \tag{A.84}$$

Consider (V). Since the process $\boldsymbol{\beta}_0(t)$ has the sub-Gaussian tail, we can show

$$\Pr \left\{ (V) \leq C n^{-1/4} \sqrt{\log p} \right\} \geq 1 - p^{-1-a}. \tag{A.85}$$

For (VI), by Assumption 1(a), we have

$$(VI) \leq C n^{-1/4} \text{ a.s.} \tag{A.86}$$

Combining (A.79)–(A.80) and (A.82)–(A.86), we have, with the probability at least $1 - p^{-a}$,

$$
\begin{aligned}
&\|\widehat{I\beta} - I\beta_0\|_{\max} \\
&\le C \left\{ s_p^2 n^{-1/4} (\log p)^4 + s_p s_{\omega,p} n^{(-2+q)/8} (\log p)^{4-2q} + p^{-1} s_p^2 \log p \right. \\
&\qquad \left. + p^{(-2+q)/2} s_p s_{\omega,p} (\log p)^{(2-q)/2} + p^{-1/2} s_p s_{\omega,p} n^{(-1+q)/8} (\log p)^{(5-4q)/2} \right\}. \qquad \text{(A.87)}
\end{aligned}
$$

■

A.9 Proof of Theorem 3

Proof of Theorem 3. By (3.13), there exists a constant C_h such that

$$
\Pr\left\{ \|\widehat{I\beta} - I\beta_0\|_{\max} \le h_n/2 \right\} \ge 1 - p^{-a}.
$$

Thus, it suffices to show the statement under $\{\|\widehat{I\beta} - I\beta_0\|_{\max} \le h_n/2\}$. Similar to the proofs of Theorem 1 (Kim et al., 2024), we can obtain

$$
\|\widetilde{I\beta} - I\beta_0\|_1 \le C s_p h_n.
$$

Also, (3.15) is obtained by (3.11). ■

A.10 Proof of Proposition 1

Proof of Proposition 1. By Proposition 5, we can show Proposition 1 similar to the proofs of Theorem 2 (Agarwal et al., 2012). ■

A.11 Miscellaneous materials

Algorithm 1 FATEN-LASSO estimation procedure

Step 1 Estimate the factor loading matrix and smoothed latent factor variable:

$$(\widehat{\mathbf{B}}_{i\Delta_n}, \widehat{\mathbf{F}}_i) = \arg \min_{\mathbf{B}\in\mathbb{R}^{p\times r}, \mathbf{F}\in\mathbb{R}^{(k_2-k_1+1)\times r}} \|\mathcal{X}_i - \mathbf{F}\mathbf{B}^\top\|_F^2,$$

subject to

$$p^{-1}\mathbf{B}^\top\mathbf{B} = \mathbf{I}_r \quad \text{and} \quad \mathbf{F}^\top\mathbf{F} \text{ is an } r\times r \text{ diagonal matrix.}$$

Step 2 Estimate the smoothed idiosyncratic variable:

$$\widehat{\mathbf{U}}_i = \mathcal{X}_i - \widehat{\mathbf{F}}_i\widehat{\mathbf{B}}_{i\Delta_n}^\top.$$

Step 3 Obtain the noise covariance matrix estimator:

$$\widehat{\mathbf{V}}^X = \frac{1}{2n}\sum_{i=1}^n \Delta_i^n\mathbf{X}^{\text{trunc}}\left(\Delta_i^n\mathbf{X}^{\text{trunc}}\right)^\top.$$

Step 4 Estimate the instantaneous coefficient:

$$\widehat{\boldsymbol{\beta}}_{i\Delta_n} = \left(\widehat{\theta}_{i\Delta_n,j}\right)_{j=1,\dots,p},$$

where

$$\widehat{\boldsymbol{\theta}}_{i\Delta_n} = \arg\min_{\|\boldsymbol{\theta}\|_1\le\rho} \frac{n}{2\phi k_1 k_2}\left\|\mathcal{Y}_i - \widehat{\mathbf{G}}_i\boldsymbol{\theta}\right\|_2^2 - \frac{n\zeta}{2\phi k_1}\boldsymbol{\theta}^\top\widehat{\mathbf{V}}\boldsymbol{\theta} + \eta\|\boldsymbol{\theta}\|_1, \quad \widehat{\mathbf{V}} = \left(\begin{array}{c|c} \widehat{\mathbf{V}}^X & \mathbf{0}_{p\times r} \\ \hline \mathbf{0}_{r\times p} & \mathbf{0}_{r\times r}\end{array}\right),$$

$\rho = C_\rho s_p$, $k_1 = c_{k_1}n^{1/2}$, $k_2 = c_{k_2}n^{3/4}$, and $\eta = C_\eta\left\{s_p n^{-1/8}(\log p)^2 + p^{-1/2}s_p\sqrt{\log p}\right\}$ for some constants C_ρ, c_{k_1}, c_{k_2}, and C_η.

Step 5 Estimate the inverse instantaneous idiosyncratic volatility matrix:

$$\widehat{\boldsymbol{\Omega}}_{i\Delta_n} = \arg\min\|\boldsymbol{\Omega}\|_1 \quad \text{s.t.} \quad \left\|\left(\frac{n}{\phi k_1 k_2}\widehat{\mathbf{U}}_i^\top\widehat{\mathbf{U}}_i - \frac{n\zeta}{\phi k_1}\widehat{\mathbf{V}}^X\right)\boldsymbol{\Omega} - \mathbf{I}\right\|_{\max} \le \tau,$$

where $\tau = C_\tau\left\{n^{-1/8}(\log p)^2 + p^{-1/2}\sqrt{\log p}\right\}$ for some large constant C_τ.

Step 6 Obtain the debiased instantaneous coefficient estimator:

$$\widetilde{\boldsymbol{\beta}}_{i\Delta_n} = \widehat{\boldsymbol{\beta}}_{i\Delta_n} + \frac{n}{\phi k_1 k_2}\widehat{\boldsymbol{\Omega}}_{i\Delta_n}^\top\left\{\widehat{\mathbf{U}}_i^\top\mathcal{Y}_i - \left(\widehat{\mathbf{U}}_i^\top\mathcal{X}_i - k_2\zeta\widehat{\mathbf{V}}^X\right)\widehat{\boldsymbol{\beta}}_{i\Delta_n}\right\}.$$

Step 7 Obtain the debiased integrated coefficient estimator:

$$\widehat{I\beta} = \sum_{i=0}^{[1/(k_2\Delta_n)]-1}\widetilde{\boldsymbol{\beta}}_{ik_2\Delta_n}k_2\Delta_n.$$

Step 8 Threshold the debiased integrated coefficient estimator:

$$\widetilde{I\beta}_j = s(\widehat{I\beta}_j)\mathbf{1}\left(|\widehat{I\beta}_j| \ge h_n\right) \quad \text{and} \quad \widetilde{I\beta} = \left(\widetilde{I\beta}_j\right)_{j=1,\dots,p},$$

where $s(x)$ satisfies $|s(x)-x| \le h_n$ and $h_n = C_h\Big[s_p^2 n^{-1/4}(\log p)^4 + s_p s_{\omega,p}n^{(-2+q)/8}(\log p)^{4-2q} + p^{-1}s_p^2\log p$ $+p^{(-2+q)/2}s_p s_{\omega,p}(\log p)^{(2-q)/2} + p^{-1/2}s_p s_{\omega,p}n^{(-1+q)/8}(\log p)^{(5-4q)/2}\Big]$ for some large constant C_h.
