Robust High-Dimensional Time-Varying Coefficient Estimation

Minseok Shin and Donggyu Kim*
Korea Advanced Institute of Science and Technology (KAIST)

September 27, 2023

Abstract

In this paper, we develop a novel high-dimensional coefficient estimation procedure based on high-frequency data. Unlike usual high-dimensional regression procedure such as LASSO, we additionally handle the heavy-tailedness of high-frequency observations as well as time variations of coefficient processes. Specifically, we employ Huber loss and truncation scheme to handle heavy-tailed observations, while ℓ_1-regularization is adopted to overcome the curse of dimensionality. To account for the time-varying coefficient, we estimate local coefficients which are biased due to the ℓ_1-regularization. Thus, when estimating integrated coefficients, we propose a debiasing scheme to enjoy the law of large number property and employ a thresholding scheme to further accommodate the sparsity of the coefficients. We call this Robust thrEsholding Debiased LASSO (RED-LASSO) estimator. We show that the RED-LASSO estimator can achieve a near-optimal convergence rate. In the empirical study, we apply the RED-LASSO procedure to the high-dimensional integrated coefficient estimation using high-frequency trading data.

Keywords: Debias, diffusion process, LASSO, factor model, sparsity, Huber loss, heavy-tail.

*Corresponding author. Address: College of Business, KAIST, Seoul 02455, South Korea. E-mail: donggyukim@kaist.ac.kr.

1   Introduction

With the wide availability of high-frequency financial data, researchers have developed financial models that can incorporate high-frequency data, and the empirical studies have shown that these models better account for market dynamics. For example, auto-regressive-type models have been introduced based on high-frequency-based measures, such as realized volatility and realized beta estimators (Andersen et al., 2006; Corsi, 2009; Engle and Gallo, 2006; Hansen et al., 2012; Kim and Wang, 2016; Kim and Fan, 2019; Shephard and Sheppard, 2010; Song et al., 2021). Empirical studies have demonstrated that capturing the auto-regressive structures of high-frequency measures helps explain financial market dynamics. On the other hand, we often employ the realized volatility estimators when analyzing regression models, such as the Capital Asset Pricing Model (CAPM) (Lintner, 1965; Sharpe, 1964) and multi-factor models (Fama and French, 1992). For example, market beta can be estimated by a ratio of the realized covariance between assets and systematic factors to the realized variance of the systematic factors (Barndorff-Nielsen and Shephard, 2004). See Andersen et al. (2006); Mykland and Zhang (2009); Reiß et al. (2015) for the related literatures. Li et al. (2017) derived the asymptotic efficiency bound for betas in a linear continuous-time regression model. Furthermore, to handle the time-varying feature of beta process (Ferson and Harvey, 1999; Kalnina, 2022; Reiß et al., 2015), Aït-Sahalia et al. (2020) employed time-localized regressions for the multi-factor models. Chen (2018) introduced the general nonparametric inference for nonlinear volatility functionals of general multivariate Itô semimartingales. These models and estimation methods have shown that incorporating high-frequency data helps better account for the beta dynamics in the finite dimensional set-up.

In modern financial studies and practices, researchers have found a large number of factor candidates (Bali et al., 2011; Campbell et al., 2008; Cochrane, 2011; Harvey et al., 2016; Hou et al., 2020; McLean and Pontiff, 2016). Thus, we often encounter the curse of dimensionality, and

the beta estimation methods designed for the finite dimension are neither efficient nor effective. To handle the high-dimensionality, we often employ LASSO (Tibshirani, 1996), SCAD (Fan and Li, 2001), and the Dantzig selector (Candes and Tao, 2007) under the sparsity condition of model parameters. However, direct application of these methods cannot handle the time-varying feature of beta processes. Recently, Kim and Shin (2022) developed a Thresholded dEbiased Dantzig (TED) estimator that can handle the high-dimensionality and time variation of beta processes. Specifically, they employed the Dantzig selector (Candes and Tao, 2007) for each time window and estimated the integrated beta with the debiasing and truncation schemes. They established the asymptotic properties of the TED estimator under the sub-Gaussianity assumption on the high-frequency log-return data. However, we often observe that the high-frequency financial data exhibit heavy tails (Cont, 2001; Fan and Kim, 2018; Mao and Zhang, 2018; Shin et al., 2023). Under the heavy-tailedness assumption, the existing estimation methods, including the TED estimator (Kim and Shin, 2022), cannot consistently estimate the time-varying betas. These facts lead to the demand for developing methodologies that can simultaneously handle heavy-tailed observations, the curse of dimensionality, and time-varying beta processes.

In this paper, we develop a robust integrated beta estimator based on high-dimensional regression jump-diffusion processes. To handle the high-dimensionality and time-varying beta, we assume that the beta processes are sparse and follow a continuous diffusion process. To account for the heavy-tailedness of financial data, we assume that the residual process and jump size processes satisfy the finite γth moment condition for $\gamma > 2$. That is, we assume that the sources of the heavy-tailedness are the residual process and jump. We first estimate the instantaneous betas as follows. We employ the ℓ_1-penalty, Huber loss, and truncation method to manage the curse of dimensionality, heavy-tailedness of the residual process, and jumps, respectively. We show that the proposed instantaneous beta estimator has the desirable convergence rate. However, the instantaneous beta estimator has non-negligible biases coming from the Huber loss and ℓ_1-penalty.

Thus, to estimate the integrated beta using the instantaneous beta estimators, we need to mitigate the biases. Since the biases are heavy-tailed, the existing debiasing scheme cannot efficiently adjust the biases. To tackle this problem, we propose a novel debiasing scheme and obtain an integrated beta estimator. We show that the debiased integrated beta estimator has a near-optimal convergence rate and outperforms the simple integration of the instantaneous beta estimators without a debiasing scheme. However, due to the bias adjustment, the debiased integrated beta estimator is not sparse; thus, we further regularize it to accommodate the sparsity. We call this the Robust thrEsholding Debiased LASSO (RED-LASSO) estimator. We also show that the RED-LASSO estimator has a near-optimal convergence rate.

The rest of paper is organized as follows. Section 2 introduces the high-dimensional regression jump-diffusion process. Section 3 proposes the RED-LASSO estimator and establishes its asymptotic properties. In Section 4, we conduct a simulation study to check the finite sample performance of the proposed estimation method. In Section 5, we apply the proposed estimation procedure to high-frequency financial data. The conclusion is presented in Section 6, and all of the proofs are collected in the Appendix.

2   The model set-up

We first fix some notations. For any given p_1 by p_2 matrix $\mathbf{A} = (A_{ij})$, let

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

The Frobenius norm of $\mathbf{A}$ is denoted by $\|\mathbf{A}\|_F = \sqrt{\operatorname{tr}(\mathbf{A}^\top \mathbf{A})}$ and the matrix spectral norm $\|\mathbf{A}\|_2$ is the square root of the largest eigenvalue of $\mathbf{A}\mathbf{A}^\top$. We will use C's to denote generic constants whose values are free of n and p and may change from appearance to appearance.

Let $Y(t)$ and $\mathbf{X}(t) = (X_1(t), \ldots, X_p(t))^\top$ be the dependent process and p-dimensional multivariate covariate process, respectively. We employ the following non-parametric time-series regression jump-diffusion model:

$$dY(t) = dY^c(t) + dY^J(t),$$

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

where $Y^c(t)$ and $\mathbf{X}^c(t) = (X_1^c(t), \ldots, X_p^c(t))^\top$ are the continuous parts of $Y(t)$ and $\mathbf{X}(t)$, respectively, $Y^J(t)$ is the jump part of $Y(t)$, $J^y(t)$ is a jump size, $\Lambda^y(t)$ is a Poisson process with a bounded intensity process, $\boldsymbol{\beta}(t) = (\beta_1(t), \ldots, \beta_p(t))^\top$ is a coefficient process, and $Z^c(t)$ is a residual process. We note that the subscript c represents the continuous part of the process. The covariate process $\mathbf{X}(t)$ and residual process $Z^c(t)$ satisfy

$$d\mathbf{X}(t) = d\mathbf{X}^c(t) + d\mathbf{X}^J(t), \quad d\mathbf{X}^c(t) = \boldsymbol{\mu}(t)dt + \boldsymbol{\sigma}(t)d\mathbf{B}(t),$$

$$d\mathbf{X}^J(t) = \mathbf{J}(t)d\boldsymbol{\Lambda}(t), \quad \text{and} \quad dZ^c(t) = \nu(t)dW(t), \tag{2.2}$$

where $\mathbf{X}^J(t)$ is the jump part of $\mathbf{X}(t)$, $\mathbf{J}(t) = (J_1(t), \ldots, J_p(t))^\top$ is a jump size process, $\boldsymbol{\Lambda}(t)$ is a p-dimensional Poisson process with bounded intensity processes, $\boldsymbol{\sigma}(t)$ is a p by q matrix, and $\mathbf{B}(t)$ and $W(t)$ are q-dimensional and one-dimensional independent Brownian motions, respectively. The stochastic processes $\boldsymbol{\mu}(t)$, $\boldsymbol{\beta}(t)$, $\boldsymbol{\sigma}(t)$, and $\nu(t)$ are defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t, t \in [0, 1]\}, P)$ with filtration $\mathcal{F}_t$ satisfying the usual conditions, such as adapted and càdlàg process. We assume that the coefficient $\boldsymbol{\beta}(t)$ satisfies the following diffusion model:

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

where $\boldsymbol{\nu}_\beta(t)$ is a p by r matrix, $\mathbf{W}_\beta(t)$ is a r-dimensional independent Brownian motion, and $\boldsymbol{\mu}_\beta(t)$

5

and $\boldsymbol{\nu}_\beta(t)$ are predictable. The main interest of this paper is to investigate the latent regression diffusion process. In this point of view, the jump part can be considered as noises, and we discuss how to overcome this in the following section. The parameter of interest is the integrated beta:

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

The integrated beta can be considered as the average of spot betas. That is, the integrated beta presents the average effect of the increment of the covariate process. When the beta process is constant, the integrated beta is the same as the usual beta in the regression model.

In the regression-based financial models, there are 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 account for this, we allow the dimension p can be large; thus, we need to handle the curse of dimensionality. To do this, we assume that the coefficient beta process $\boldsymbol{\beta}(t) = (\beta_1(t),\ldots,\beta_p(t))^\top$ satisfies the following sparsity condition:

$$\sup_{0\leq t\leq 1} \sum_{i=1}^p |\beta_i(t)|^\delta \leq s_p \quad \text{and} \quad \sum_{i=1}^p |I\beta_i|^\delta \leq s_p \text{ a.s.,} \tag{2.3}$$

where $\delta \in [0,1)$, s_p is diverging slowly in p, and 0^0 is defined as 0. This general sparsity condition includes the exact sparsity condition, i.e., $\delta = 0$. We note that the exact sparsity condition implies that only several factors are significant, while most factors do not affect the dependent process. Thus, intuitively, we assume that the relatively small number of factors are significant. We note that since the beta process is an Itô diffusion process, in general, the boundedness in the sparsity condition (2.3) is satisfied with high probability. However, for simplicity, we assume the almost sure boundedness.

3 Robust high-dimensional high-frequency regression

3.1 Integrated beta estimation procedure

In this section, we propose a robust integrated beta estimation procedure for the high-dimensional regression diffusion model defined in (2.1)–(2.2). Recently, with the sub-Gaussian assumption, Kim and Shin (2022) proposed the integrated beta estimator that can handle the curse of dimensionality and time-varying beta. However, empirical studies have demonstrated that the stock log-return data often exhibit heavy-tails (Cont, 2001; Fan and Kim, 2018; Mao and Zhang, 2018; Shin et al., 2023). To account for this, we impose the finite moment condition for the residual process, $Z^c(t)$, and jump sizes, $J^y(t)$ and $\mathbf{J}(t)$ (see Assumption 1). Then, we propose a robust estimation procedure. We first estimate the instantaneous betas. To do this, we employ the local regression as follows. For any process $g(t)$ and $\Delta_n = 1/n$, let $\Delta_i^n g = g(i\Delta_n) - g((i-1)\Delta_n)$ for $1 \le i \le 1/\Delta_n$. Define

$$
\mathcal{Y}_i = \begin{pmatrix} \Delta_{i+1}^n Y \\ \Delta_{i+2}^n Y \\ \vdots \\ \Delta_{i+k_n}^n Y \end{pmatrix}, \quad
\mathcal{Z}_i = \begin{pmatrix} \Delta_{i+1}^n Z^c \\ \Delta_{i+2}^n Z^c \\ \vdots \\ \Delta_{i+k_n}^n Z^c \end{pmatrix},
$$

$$
\mathcal{X}_i = \begin{pmatrix} \Delta_{i+1}^n \widehat{\mathbf{X}}^{c\top} \\ \Delta_{i+2}^n \widehat{\mathbf{X}}^{c\top} \\ \vdots \\ \Delta_{i+k_n}^n \widehat{\mathbf{X}}^{c\top} \end{pmatrix}, \quad \text{and} \quad
\Delta_i^n \widehat{\mathbf{X}}^c = \begin{pmatrix} \Delta_i^n X_1 \, \mathbf{1}_{\{|\Delta_i^n X_1| \le v_{1,n}\}} \\ \Delta_i^n X_2 \, \mathbf{1}_{\{|\Delta_i^n X_2| \le v_{2,n}\}} \\ \vdots \\ \Delta_i^n X_p \, \mathbf{1}_{\{|\Delta_i^n X_p| \le v_{p,n}\}} \end{pmatrix},
$$

where k_n is the number of observations for each local regression, $\mathbf{1}_{\{\cdot\}}$ is an indicator function, and $v_{j,n}$, $j = 1, \ldots, p$, are the threshold levels. We use $v_{j,n} = C_{j,v}\sqrt{\log p}\, n^{-1/2}$ for some large constants

$C_{j,v}$, $j = 1, \ldots, p$. In the numerical study, we choose

$$v_{j,n} = 3\sqrt{BV_j}n^{-1/2}, \qquad (3.1)$$

where the bipower variation $BV_j = \frac{\pi}{2}\sum_{i=2}^{n}|\Delta_{i-1}^n X_j| \cdot |\Delta_i^n X_j|$. This choice of $v_{j,n}$ is similar to the usual choice in the literatures (Aït-Sahalia et al., 2020; Aït-Sahalia and Xiu, 2019). We note that the thresholding can detect the jumps in the covariate process $X(t)$ and mitigate their impact on beta estimators. On the other hand, the thresholding is not used for the dependent process $Y(t)$ since the robustification method outlined in (3.3) and (3.5) can handle both heavy-tailedness of the residual process $Z^c(t)$ and jumps in the dependent process $Y(t)$. Meanwhile, when calculating local regressions, we need to handle the curse of dimensionality and heavy-tailedness. To overcome high-dimensionality, we often employ the penalized regression procedure under the sparsity assumption. For example, we often use the LASSO (Tibshirani, 1996) and Dantzig (Candes and Tao, 2007) estimators with the sub-Gaussian conditions. However, these estimators cannot handle the heavy-tailed observations, and furthermore, they are not consistent. To tackle this issue, we use the following Huber loss l_τ (Huber, 1964):

$$l_\tau(x) = \begin{cases} x^2/2 & \text{if } |x| \le \tau \\ \tau|x| - \tau^2/2 & \text{if } |x| > \tau, \end{cases}$$

where $\tau > 0$ is the robustification parameter. We denote $l_\tau(\mathbf{x}) = (l_\tau(x_1), \ldots, l_\tau(x_{p_1}))^\top$ for any vector $\mathbf{x} = (x_1, \ldots, x_{p_1})^\top \in \mathbb{R}^{p_1}$. The Huber loss l_τ mitigates the effect of outliers coming from the heavy-tailedness of the residual process $Z^c(t)$ and jump size process $J^y(t)$. Thus, by employing the truncation, Huber loss, and ℓ_1-regularization, we can simultaneously deal with the three issues of the jumps, heavy-tailedness, and curse of dimensionality. Specifically, we propose the following

instantaneous beta estimator at time $i\Delta_n$:

$$\widehat{\boldsymbol{\beta}}_{i\Delta_n} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \mathcal{L}_{\tau,i}(\boldsymbol{\beta}) + \eta\,\|\boldsymbol{\beta}\|_1\,, \tag{3.2}$$

where $\eta > 0$ is the regularization parameter, and the empirical loss function is

$$\mathcal{L}_{\tau,i}(\boldsymbol{\beta}) = \left\|l_\tau\left(\mathcal{Y}_i - \mathcal{X}_i\boldsymbol{\beta}\right)/k_n\right\|_1. \tag{3.3}$$

In Theorem 1, we show that the proposed instantaneous beta estimator $\widehat{\boldsymbol{\beta}}_{i\Delta_n}$ is consistent with appropriate τ and η. Then, we can estimate the integrated beta using the integration of $\widehat{\boldsymbol{\beta}}_{i\Delta_n}$'s. However, their integration cannot enjoy the law of large number properties since each $\widehat{\boldsymbol{\beta}}_{i\Delta_n}$ is biased due to the regularization term. That is, the error of their integration is dominated by the bias terms, which leads to the same convergence rate as that of $\widehat{\boldsymbol{\beta}}_{i\Delta_n}$. Thus, to reduce the effect of the bias and obtain faster convergence rate, we propose a debiasing scheme as follows. First, we estimate the inverse instantaneous volatility matrix at time $i\Delta_n$, $\boldsymbol{\Omega}(i\Delta_n) = \boldsymbol{\Sigma}^{-1}(i\Delta_n)$, where $\boldsymbol{\Sigma}(t) = \boldsymbol{\sigma}(t)\boldsymbol{\sigma}^\top(t)$. Specifically, we use the following constrained ℓ_1-minimization for inverse matrix estimation (CLIME) (Cai et al., 2011):

$$\widehat{\boldsymbol{\Omega}}_{i\Delta_n} = \arg\min \|\boldsymbol{\Omega}\|_1 \quad \text{s.t.} \quad \|\frac{1}{k_n\Delta_n}\mathcal{X}_i^\top\mathcal{X}_i\boldsymbol{\Omega} - \mathbf{I}\|_{\max} \le \lambda, \tag{3.4}$$

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

$$\widetilde{\boldsymbol{\beta}}'_{i\Delta_n} = \widehat{\boldsymbol{\beta}}_{i\Delta_n} + \frac{1}{k_n\Delta_n}\widehat{\boldsymbol{\Omega}}_{i\Delta_n}^\top\mathcal{X}_i^\top\left(\mathcal{Y}_i - \mathcal{X}_i\widehat{\boldsymbol{\beta}}_{i\Delta_n}\right).$$

9

This debiasing scheme performs well under the sub-Gaussian assumption (Javanmard and Montanari, 2014, 2018; Kim and Shin, 2022; Van de Geer et al., 2014). However, $\Delta_i^n Z^c$ has only finite γth moment for $\gamma > 2$; thus, the debiased instantaneous beta estimator has the heavy-tails. To handle this issue, we employ the Winsorization method as follows. Define the truncation (Winsorization) function

$$\psi_\varpi(x) = \begin{cases} x & \text{if } |x| \leq \varpi \\ \text{sign}(x)\varpi & \text{if } |x| > \varpi, \end{cases}$$

where $\varpi > 0$ is a truncation parameter and denote $\psi_\varpi(\mathbf{x}) = (\psi_\varpi(x_1), \ldots, \psi_\varpi(x_{p_1}))^\top$ for any vector $\mathbf{x} = (x_1, \ldots, x_{p_1})^\top \in \mathbb{R}^{p_1}$. Using this truncation function, we adjust $\widehat{\boldsymbol{\beta}}_{i\Delta_n}$ as

$$\widetilde{\boldsymbol{\beta}}_{i\Delta_n} = \widehat{\boldsymbol{\beta}}_{i\Delta_n} + \psi_\varpi\left(\frac{1}{k_n\Delta_n}\widehat{\boldsymbol{\Omega}}_{i\Delta_n}^\top \mathcal{X}_{(i+k_n)}^\top \left(\mathcal{Y}_{(i+k_n)} - \mathcal{X}_{(i+k_n)}\widehat{\boldsymbol{\beta}}_{i\Delta_n}\right)\right), \tag{3.5}$$

where the truncation parameter ϖ will be specified in Theorem 2. We note that for the debiasing step, we use the non-overlapping window for $\mathcal{X}$ and $\mathcal{Y}$, which helps enjoy the martingale property. Specifically, since $\boldsymbol{\beta}_0((i+k_n)\Delta_n) - \widehat{\boldsymbol{\beta}}_{i\Delta_n}$ is measurable at time $(i+k_n)\Delta_n$, we can handle the noises from $\mathcal{X}_{(i+k_n)}$ and $\mathcal{Y}_{(i+k_n)}$ using the martingale convergence theorem. We also note that the purpose of the debiasing is to enjoy the law of large number property when obtaining the integrated beta estimator. Usually, the debiasing scheme is employed to obtain the asymptotic normality, which enables the hypothesis test or confidence interval construction (Javanmard and Montanari, 2014, 2018; Van de Geer et al., 2014; Zhang and Zhang, 2014). However, in this paper, we do not focus on this issue and mainly focus on the integrated beta estimation. Then, the integrated beta estimator is defined as follows:

$$\widehat{I\beta} = \sum_{i=0}^{[1/(k_n\Delta_n)]-2} \widetilde{\boldsymbol{\beta}}_{ik_n\Delta_n} k_n\Delta_n. \tag{3.6}$$

The debiased LASSO integrated beta estimator $\widehat{I\beta}$ can achieve a faster convergence rate than the simple integration of the instantaneous beta estimators. However, due to the bias adjustment term, it cannot account for the sparsity structure of the integrated beta. To accommodate the sparsity, we employ the following thresholding scheme:

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

where the thresholding function $s(\cdot)$ satisfies $|s(x)-x| \leq h_n$ and h_n is a thresholding level, which will be specified in Theorem 3. For example, we can employ the hard thresholding function $s(x) = x$ or soft thresholding function $s(x) = x - \text{sign}(x)h_n$. In the empirical study, we used the hard thresholding function $s(x) = x$. We call this the Robust thrEsholding Debiased LASSO (RED-LASSO) estimator. We describe the RED-LASSO estimation procedure in Algorithm 1.

Algorithm 1 RED-LASSO estimation procedure

Step 1 Obtain the instantaneous beta estimator:

$$\widehat{\boldsymbol{\beta}}_{i\Delta_n} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \|l_\tau\left(\mathcal{Y}_i - \mathcal{X}_i\boldsymbol{\beta}\right)/k_n\|_1 + \eta\,\|\boldsymbol{\beta}\|_1,$$

where $\tau = C_\tau n^{-1/4}(\log p)^{-3/4}$, $\eta = C_\eta\left[s_p n^{-5/4}\sqrt{\log p} + n^{-5/4}(\log p)^{3/4}\right]$, and $k_n = c_k n^{1/2}$ for some large constants C_τ, C_η, and c_k.

Step 2 Obtain the inverse instantaneous volatility matrix estimator:

$$\widehat{\boldsymbol{\Omega}}_{i\Delta_n} = \arg\min \|\boldsymbol{\Omega}\|_1 \quad \text{s.t.} \quad \|\frac{1}{k_n\Delta_n}\mathcal{X}_i^\top\mathcal{X}_i\boldsymbol{\Omega} - \mathbf{I}\|_{\max} \le \lambda,$$

where $\lambda = C_\lambda n^{-1/4}\sqrt{\log p}$ for some large constant C_λ.

Step 3 Debias the instantaneous beta estimator:

$$\widetilde{\boldsymbol{\beta}}_{i\Delta_n} = \widehat{\boldsymbol{\beta}}_{i\Delta_n} + \psi_\varpi\left(\frac{1}{k_n\Delta_n}\widehat{\boldsymbol{\Omega}}_{i\Delta_n}^\top\mathcal{X}_{(i+k_n)}^\top\left(\mathcal{Y}_{(i+k_n)} - \mathcal{X}_{(i+k_n)}\widehat{\boldsymbol{\beta}}_{i\Delta_n}\right)\right),$$

where $\varpi = C_\varpi s_p^{2-\delta} n^{\delta/4}(\log p)^{(1-3\delta)/4}$ for some large constant C_ϖ.

Step 4 Obtain the integrated beta estimator:

$$\widehat{I\beta} = \sum_{i=0}^{[1/(k_n\Delta_n)]-2} \widetilde{\boldsymbol{\beta}}_{ik_n\Delta_n} k_n\Delta_n.$$

Step 5 Threshold the integrated beta estimator:

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

where $s(\cdot)$ satisfies $|s(x) - x| \le h_n$, $h_n = C_h b_n$ for some large constant C_h, and b_n is defined in Theorem 2.

3.2   Theoretical results

In this section, we investigate asymptotic properties of the proposed RED-LASSO estimation procedure. To investigate the theoretical properties, we make the following assumptions.

Assumption 1.

(a) The residual process $Z^c(t)$ and jump size processes, $J^y(t)$ and $\mathbf{J}(t) = (J_1(t),\dots,J_p(t))^\top$,

satisfy, for some $\gamma > 2$,

$$\max_{1 \leq i \leq n} \mathbb{E}\left\{ |\Delta_i^n Z^c|^\gamma \middle| \mathcal{F}_{(i-1)\Delta_n} \right\} \leq Cn^{-\gamma/2},$$

$$\sup_{0 \leq t \leq 1} \mathbb{E}\{|J^y(t)|^\gamma\} \leq C, \quad \text{and} \quad \sup_{0 \leq t \leq 1} \max_{1 \leq i \leq p} \mathbb{E}\{|J_i(t)|^\gamma\} \leq C \quad a.s.$$

(b) The processes $\boldsymbol{\mu}(t)$, $\boldsymbol{\mu}_\beta(t)$, $\boldsymbol{\beta}(t)$, $\boldsymbol{\Sigma}(t)$, and $\boldsymbol{\Sigma}_\beta(t) = \boldsymbol{\nu}_\beta(t)\boldsymbol{\nu}_\beta^\top(t)$ are almost surely entry-wise bounded, and $\|\boldsymbol{\Sigma}^{-1}(t)\|_1 \leq C$ a.s.

(c) The processes $\boldsymbol{\mu}_\beta(t) = (\mu_{\beta,1}(t), \ldots, \mu_{\beta,p}(t))^\top$ and $\boldsymbol{\Sigma}_\beta(t) = (\Sigma_{\beta,ij}(t))_{i,j=1,\ldots,p}$ satisfy the following sparsity condition for $\delta \in [0, 1)$:

$$\sup_{0 \leq t \leq 1} \sum_{i=1}^p |\mu_{\beta,i}(t)|^\delta \leq s_p \quad \text{and} \quad \sup_{0 \leq t \leq 1} \sum_{i=1}^p |\Sigma_{\beta,ii}(t)|^{\delta/2} \leq s_p \quad a.s.$$

(d) $n^{c_1} \leq p \leq c_2 \exp(n^{c_3})$ for some positive constants c_1, c_2, and $c_3 < 1/6$, and $s_p^2 \log p \Delta_n k_n \to 0$ as $n, p \to \infty$.

(e) Define $\mathcal{W}_t = \left\{ \mathbf{w} \in \mathbb{R}^p : \|\mathbf{w}_{S_t^c}\|_1 \leq 3\|\mathbf{w}_{S_t}\|_1 + 4\|(\boldsymbol{\beta}_0(t))_{S_t^c}\|_1 \right\}$, where $\mathbf{w}_{S_t^c}$ is the subvector obtained by stacking $\{\mathbf{w}_j : j \in S_t^c\}$, $\mathbf{w}_{S_t}$ is the subvector obtained by stacking $\{\mathbf{w}_j : j \in S_t\}$, $(\boldsymbol{\beta}_0(t))_{S_t^c}$ is the subvector obtained by stacking $\{(\boldsymbol{\beta}_0(t))_j : j \in S_t^c\}$, and $S_t = \{j : j\text{th element}$ of $|\boldsymbol{\beta}_0(t)| > n\eta\}$. Then, there exists a positive constant κ such that the following inequality holds for some $D = (8 + 48/\kappa)s_p(n\eta)^{1-\delta}$ and $0 \leq i \leq n - k_n$, where the specific value of η is given in Theorem 1:

$$\inf\{\mathbf{w}^\top \nabla^2 \mathcal{L}_{\tau,i}(\boldsymbol{\beta})\mathbf{w} : \mathbf{w} \in \mathcal{W}_{i\Delta_n}, \ \|\mathbf{w}\|_2 = 1, \ \|\boldsymbol{\beta} - \boldsymbol{\beta}_0(i\Delta_n)\|_1 \leq D\} \geq \kappa/n.$$

(f) The volatility process $\mathbf{\Sigma}(t) = (\Sigma_{ij}(t))_{i,j=1,\ldots,p}$ satisfies the following condition:

$$|\Sigma_{ij}(t) - \Sigma_{ij}(s)| \leq C\sqrt{|t-s|\log p} \quad a.s.$$

Remark 1. Assumption 1(a) is the finite moment condition, which implies that the dependent process $Y(t)$, covariate process $\mathbf{X}(t)$, and residual process $Z^c(t)$ have heavy-tails. We note that the moment condition for $Z^c(t)$ is satisfied when $\Delta_i^n Z^c$ is an independent random variable and $\mathbb{E}\{|\Delta_i^n Z^c|^\gamma\} \leq Cn^{-\gamma/2}$, or $\sup_{0 \leq t \leq 1} \sup_{t \leq s \leq 1} \mathbb{E}\left\{|\nu(s)|^\gamma \big| \mathcal{F}_t\right\} \leq C$ a.s. The latter condition can be satisfied when $\nu(t)$ consists of the bounded continuous process and independent jump process. The boundedness condition Assumption 1(b) implies the sub-Gaussianity for the continuous part of the covariate process, $\mathbf{X}^c(t)$, and target parameter, $\boldsymbol{\beta}(t)$, which are often required to investigate high-dimensional inferences. However, the boundedness condition can be relaxed to the locally boundedness condition by Lemma 4.4.9 in Jacod and Protter (2011). Specifically, if the asymptotic result, such as stable convergence in law or convergence in probability, is satisfied under the boundedness condition, it is also satisfied under the locally boundedness condition. On the other hand, for the continuous-time regression model, we usually assume that the smallest eigenvalue of $\mathbf{\Sigma}(t)$ is bounded from below, which implies that the largest eigenvalue of $\mathbf{\Sigma}^{-1}(t)$ is bounded. In this point of view, the condition $\|\mathbf{\Sigma}^{-1}(t)\|_1 \leq C$ a.s. is not restrictive. Even if this condition is replaced by the sparsity condition $\sup_{0 \leq t \leq 1} \max_{1 \leq i \leq p} \sum_{j=1}^p |\omega_{ij}(t)|^q \leq s_{\omega,p}$ a.s., where $\mathbf{\Sigma}^{-1}(t) = (\omega_{ij}(t))_{i,j=1,\ldots,p}$, and $q \in [0,1)$ and $s_{\omega,p}$ are the sparsity related variables, the difference in theoretical results is up to $s_{\omega,p}$ order. Assumption 1(c) is the sparsity condition for the beta process, which is required to investigate the discretization error when estimating instantaneous betas. Assumption 1(e) is the eigenvalue condition for the Hessian matrix $\nabla^2 \mathcal{L}_{\tau,i}(\boldsymbol{\beta})$, which is called the localized restricted eigenvalue (LRE) condition (Fan et al., 2018; Sun et al., 2020). This implies strictly positive restricted eigenvalues over a local neighborhood. We note that $n\eta$ converges to

zero for the choice of η in Theorems 1–2. When the coefficient process $\boldsymbol{\beta}(t)$ satisfies the exact sparsity condition, i.e., $\delta = 0$, $\mathcal{W}_t$ is replaced by a ℓ_1-cone $\left\{ \mathbf{w} \in \mathbb{R}^p : \ \left\| \mathbf{w}_{S_t^c} \right\|_1 \leq 3 \left\| \mathbf{w}_{S_t} \right\|_1 \right\}$, where $S_t = \{ j : \ j\text{th element of } \boldsymbol{\beta}_0(t) \neq 0 \}$. Finally, we need the continuity condition Assumption 1(f) to investigate asymptotic behaviors of the CLIME estimator. We note that this condition is obtained with high probability when $\boldsymbol{\Sigma}(t)$ follows a continuous Itô diffusion process with bounded drift and instantaneous volatility processes.

The following theorem derives the asymptotic properties of instantaneous beta estimator $\widehat{\boldsymbol{\beta}}_{i\Delta_n}$. Note that the subscript 0 represents the true parameters.

Theorem 1. *Under Assumption 1(a)–(e), let $k_n = c_k n^c$ for some constants c_k and $c \in [3/8, 3/4]$. For any given positive constant a, choose $\tau = C_{\tau,a} \sqrt{k_n \Delta_n} (\log p)^{-3/4}$ and $\eta = C_{\eta,a} \Big[s_p n^{-3/2} \sqrt{k_n \log p}$ $+ n^{-1} k_n^{-1/2} (\log p)^{3/4} \Big]$ for some large constants $C_{\tau,a}$ and $C_{\eta,a}$. Then, we have, for large n,*

$$\max_i \| \widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0(i\Delta_n) \|_1 \leq C s_p (n\eta)^{1-\delta} \quad and \quad \max_i \| \widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0(i\Delta_n) \|_2 \leq C \sqrt{s_p} (n\eta)^{1-\delta/2}, \quad (3.7)$$

with probability greater than $1 - p^{-a}$.

Remark 2. Theorem 1 shows the ℓ_1 and ℓ_2 norm error bounds of the instantaneous beta estimator. We note that as k_n increases, the statistical estimation error decreases and time variation approximation error increases. To achieve the optimality, we choose $c = 1/2$, which implies that these two errors have the same convergence rate. Then, the instantaneous beta estimator has the ℓ_1 convergence rate of $n^{-(1-\delta)/4}$ and ℓ_2 convergence rate of $n^{-(2-\delta)/8}$ with the log order and sparsity level terms.

To estimate the integrated beta, we can use the integration of the instantaneous beta estimators. However, as discussed in Section 3.1, it cannot enjoy the law of large number property due to the heavy-tailed biases. To tackle this problem, we employ the robust debiasing method (3.5)

and obtain the debiased LASSO integrated beta estimator $\widehat{I\beta}$ in (3.6). The following theorem establishes the asymptotic behaviors of $\widehat{I\beta}$.

Theorem 2. *Under the assumptions in Theorem 1 and Assumption 1(f), choose $k_n = c_k n^{1/2}$ for some constant c_k. For any given positive constant a, let $\lambda = C_{\lambda,a} n^{-1/4}\sqrt{\log p}$ and $\varpi = C_\varpi s_p^{2-\delta} n^{\delta/4}(\log p)^{(1-3\delta)/4}$ for some constants $C_{\lambda,a}$ and C_ϖ. Then, we have, with probability greater than $1 - p^{-a}$,*

$$\|\widehat{I\beta} - I\beta_0\|_{\max} \leq Cb_n, \tag{3.8}$$

where $b_n = s_p^{2-\delta} n^{(-2+\delta)/4}(\log p)^{(5-3\delta)/4} + s_p n^{-1/2}(\log p)^{3/2}$.

Remark 3. Theorem 2 shows the max norm error bound of the debiased LASSO integrated beta estimator. When the beta process satisfies the exact sparsity condition, i.e., $\delta = 0$, the debiased LASSO integrated beta estimator has the convergence rate of $s_p^2 n^{-1/2}(\log p)^{5/4} + s_p n^{-1/2}(\log p)^{3/2}$, while we have a slower convergence rate of $s_p^2 n^{-1/4}\sqrt{\log p} + s_p n^{-1/4}(\log p)^{3/4}$ without a debiasing scheme. The $n^{1/2}$ term is the optimal convergence rate of estimating model parameters given n observations. For the log order term, the usual optimal rate is $\sqrt{\log p}$ in high dimensional inferences. However, we have $(\log p)^{3/2}$ term since the additional $\log p$ term comes from bounding the time-varying processes, such as the target process $\boldsymbol{\beta}(t)$. In sum, the debiased LASSO integrated beta estimator has the optimal converence rate with up to $\log p$ and s_p orders.

Theorem 2 reveals that the debiased LASSO integrated beta estimator performs better than the integration of the instantaneous beta estimators. Finally, to account for the sparsity structure, we threshold the debiased LASSO integrated beta estimator and obtain the RED-LASSO estimator. Theorem 3 establishes the ℓ_1 convergence rate of the RED-LASSO estimator.

Theorem 3. *Under the assumptions in Theorem 2, for any given positive constant a, choose*

16

$h_n = C_{h,a} b_n$ for some constant $C_{h,a}$. Then, we have, with probability greater than $1 - p^{-a}$,

$$\|\widehat{I\beta} - I\beta_0\|_1 \leq C s_p b_n^{1-\delta}. \tag{3.9}$$

Theorem 3 shows that the proposed RED-LASSO estimator is consistent in terms of the ℓ_1 norm. We note that under the sub-Gaussian assumption on the log-return data, Kim and Shin (2022) proposed the integrated beta estimator that has the ℓ_1 convergence rate of $s_p a_n^{1-\delta}$, where $a_n = s_p^{2-\delta} n^{(-2+\delta)/4} (\log p)^{(2-\delta)/2} + s_p s_{\omega,p} n^{(-2+q)/4} (\log p)^{(2-q)/2} + s_p n^{-1/2} (\log p)^{3/2}$, and $s_{\omega,p}$ and q are the sparsity related terms for the inverse volatility matrix. Thus, the cost of handling the heavy-tailedness is at most $\log p$ order.

3.3 Discussion on the tuning parameter selection

In this section, we discuss how to choose the tuning parameters to implement the RED-LASSO estimation procedure. We first obtain the variables $\Delta_i^n \widehat{X}_j^c$, $j = 1, \ldots, p$, based on the threshold level (3.1). Then, to handle the scale problem, we standardize the variables $\Delta_i^n Y$ and $\Delta_i^n \widehat{X}_j^c$, $j = 1, \ldots, p$, to have a zero mean and unit variance. The re-scaling is employed after obtaining the RED-LASSO estimator. In the local regression stage (3.2), we select $k_n = [n^{1/2}]$. Also, we choose

$$\tau = c_\tau n^{-1/4} (\log p)^{-3/4}, \quad \eta = c_\eta n^{-5/4} (\log p)^{3/4},$$

$$\lambda = c_\lambda n^{-1/4} \sqrt{\log p}, \quad \varpi = c_\varpi (\log p)^{1/4}, \quad \text{and} \quad h_n = c_h n^{-1/2} (\log p)^{3/2}, \tag{3.10}$$

where c_τ, c_η, c_λ, c_ϖ, and c_h are tuning parameters. For the simulation and empirical studies, we choose c_τ, c_ϖ, and c_h that minimize the corresponding mean squared prediction error (MSPE). The results are $c_\tau = 16$, $c_\varpi = 1/64$, and $c_h = 1/4$. Details can be found in Section 5. Also, we select $c_\eta \in [0.1, 10]$, which minimizes the corresponding Bayesian information criterion (BIC). Finally, we

choose $c_\lambda \in [0.1, 10]$ that minimizes the following loss function:

$$\mathrm{tr}\left[\left(\frac{1}{k_n\Delta_n}\mathcal{X}_i^\top\mathcal{X}_i\widehat{\boldsymbol{\Omega}}_{i\Delta_n} - \mathbf{I}_p\right)^2\right],$$

where $\mathbf{I}_p$ is the p-dimensional identity matrix.

4   A simulation study

To check the finite sample performance of the proposed RED-LASSO estimator, we conducted simulations. Based on the models (2.1)–(2.2), we generated the data using the heavy-tail and sub-Gaussian processes with frequency $1/n^{all}$. Specifically, we employed the following time-series regression jump-diffusion model:

$$dY(t) = \boldsymbol{\beta}^\top(t)d\mathbf{X}^c(t) + dZ^c(t) + J^y(t)d\Lambda^y(t), \quad d\mathbf{X}(t) = d\mathbf{X}^c(t) + d\mathbf{X}^J(t),$$

$$d\mathbf{X}^c(t) = \boldsymbol{\sigma}(t)d\mathbf{B}(t), \quad d\mathbf{X}^J(t) = \mathbf{J}(t)d\boldsymbol{\Lambda}(t), \quad \text{and} \quad dZ^c(t) = \nu(t)dW(t),$$

where the jump sizes $J_i(t)$ and $J^y(t)$ were obtained from 0.1 times i.i.d. t-distribution with degrees of freedom df, and $\boldsymbol{\Lambda}(t) = (\Lambda_1(t), \ldots, \Lambda_p(t))^\top$ and $\Lambda^y(t)$ were generated by Poisson processes with the intensities $(20, \ldots, 20)^\top$ and 10, respectively. We chose df as 2 and ∞ for the heavy-tailed and sub-Gaussian processes, respectively. The initial values of $X(t)$ and $Y(t)$ were set as zero, and we generated $\nu(t)$ as follows:

$$\nu(t_l) = (1 + 0.5\,|t_{df,l}|)\,\nu'(t_l),$$

where $t_{df,l}$, $l = 1, \ldots, n^{all}$, are the i.i.d. t-distributions with degrees of freedom df, and $\nu'(t_l)$, $l = 1, \ldots, n^{all}$, were generated from the following Ornstein-Uhlenbeck process:

$$d\nu'(t) = 3\,(0.4 - \nu'(t))\,dt + 0.12d\mathbf{W}^\nu(t),$$

where $\nu'(0) = 0.5$ and $\mathbf{W}^\nu(t)$ is an independent Brownian motion. We note that the process $\nu(t)$ is not realistic. However, to investigate the effect of the heavy-tailedness of the return process, the structure of $\nu(t)$ is imposed. To generate the volatility process $\boldsymbol{\sigma}(t)$, we first generated the Ornstein-Uhlenbeck process $u(t)$ as follows:

$$du(t) = 5\left(0.45 - u(t)\right)dt + 0.2d\mathbf{W}^u(t),$$

where $u(0) = 1$ and $\mathbf{W}^u(t)$ is an independent Brownian motion. Then, we took $\boldsymbol{\sigma}(t)$ as a Cholesky decomposition of $\boldsymbol{\Sigma}(t) = (\Sigma_{ij}(t))_{1 \leq i,j \leq p}$, where $\Sigma_{ij}(t) = u(t)0.8^{|i-j|}$. To generate the coefficient $\boldsymbol{\beta}(t)$, we considered the exact sparse process, i.e., $\beta_i(t) = 0$ for $[s_p] + 1 \leq i \leq p$. Specifically, we generated $\boldsymbol{\beta}(t)$ as follows:

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

where $\boldsymbol{\mu}_\beta(t) = (\mu_{1,\beta}(t), \ldots, \mu_{p,\beta}(t))^\top$, $\boldsymbol{\nu}_\beta(t) = (\nu_{i,j,\beta}(t))_{1 \leq i,j \leq p}$, and $\mathbf{W}_\beta(t)$ is a p-dimensional independent Brownian motion. For $1 \leq i \leq [s_p]$, the initial value $\beta_i(0) = 1$ and $\mu_{i,\beta}(t) = 0.1$ for $0 \leq t \leq 1$. The process $(\nu_{i,j,\beta}(t))_{1 \leq i,j \leq [s_p]}$ was taken to be $\xi(t)\mathbf{I}_{[s_p]}$, where $\mathbf{I}_{[s_p]}$ is the $[s_p]$-dimensional identity matrix and $\xi(t)$ follows the Ornstein-Uhlenbeck process:

$$d\xi(t) = 3\left(0.3 - \xi(t)\right)dt + 0.1d\mathbf{W}^\xi(t),$$

where $\xi(0) = 0.15$ and $\mathbf{W}^\xi(t)$ is an independent Brownian motion. We chose $p = 100$, $s_p = \log p$, $n^{all} = 4000$, and we varied n from 1000 to 4000. When implementing the RED-LASSO estimation procedure, the tuning parameters were selected as discussed in Section 3.3.

To investigate the effect of the robustification of the RED-LASSO estimator, we employed a thrEsholding Debiased LASSO (ED-LASSO) estimator. The ED-LASSO estimator uses the same estimation procedure as the RED-LASSO estimator with $\tau = \varpi = \infty$. Since the ED-LASSO

estimator does not employ the Huber loss and Winsorization method, the jump adjustment for the dependent process $Y(t)$ is needed. Thus, we used $\mathcal{Y}_i'$ instead of $\mathcal{Y}_i$ for the ED-LASSO estimator, where

$$\mathcal{Y}_i' = \begin{pmatrix} \Delta_{i+1}^n \widehat{Y}^c \\ \Delta_{i+2}^n \widehat{Y}^c \\ \vdots \\ \Delta_{i+k_n}^n \widehat{Y}^c \end{pmatrix} \qquad \text{and} \qquad \Delta_i^n \widehat{Y}^c = \Delta_i^n Y \, \mathbf{1}_{\{|\Delta_i^n Y| \leq u_n\}}. \tag{4.1}$$

In the simulation and empirical studies, we choose $u_n = 3\sqrt{BV^Y}n^{-1/2}$, where the bipower variation $BV^Y = \dfrac{\pi}{2}\sum_{i=2}^n |\Delta_{i-1}^n Y| \cdot |\Delta_i^n Y|$. We note that the ED-LASSO estimator can enjoy the same theoretical properties as the RED-LASSO estimator under the sub-Gaussian process, but it cannot explain the heavy-tailed process. As a benchmark, we also considered the LASSO estimator (Tibshirani, 1996), which cannot account for any of the heavy-tailed distribution or the time-varying beta process. Specifically, we employed the LASSO estimator as follows:

$$\widetilde{I\beta}^{\text{LASSO}} = \text{argmin}_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^n \left(\Delta_i^n \widehat{Y}^c - \Delta_i^n \widehat{\mathbf{X}}^{c\top} \boldsymbol{\beta} \right)^2 + \eta^{\text{LASSO}} \|\boldsymbol{\beta}\|_1 \right\}, \tag{4.2}$$

where the regularization parameter $\eta^{\text{LASSO}} \in [0.1, 10]$ was selected by minimizing the corresponding Bayesian information criterion (BIC). The average estimation errors under the max norm, ℓ_1 norm, and ℓ_2 norm were computed by 1000 simulations.

Figure 1 plots the log max, ℓ_1, and ℓ_2 norm errors of the RED-LASSO, ED-LASSO, and LASSO estimators with $n = 1000, 2000, 4000$ for the heavy-tail and sub-Gaussian processes. From Figure 1, we can find that the estimation errors of the RED-LASSO estimator decrease as the sample size n increases. As expected, the RED-LASSO estimator performed the best for the heavy-tail process. This may be because the RED-LASSO estimator can explain the heavy-tailedness while other
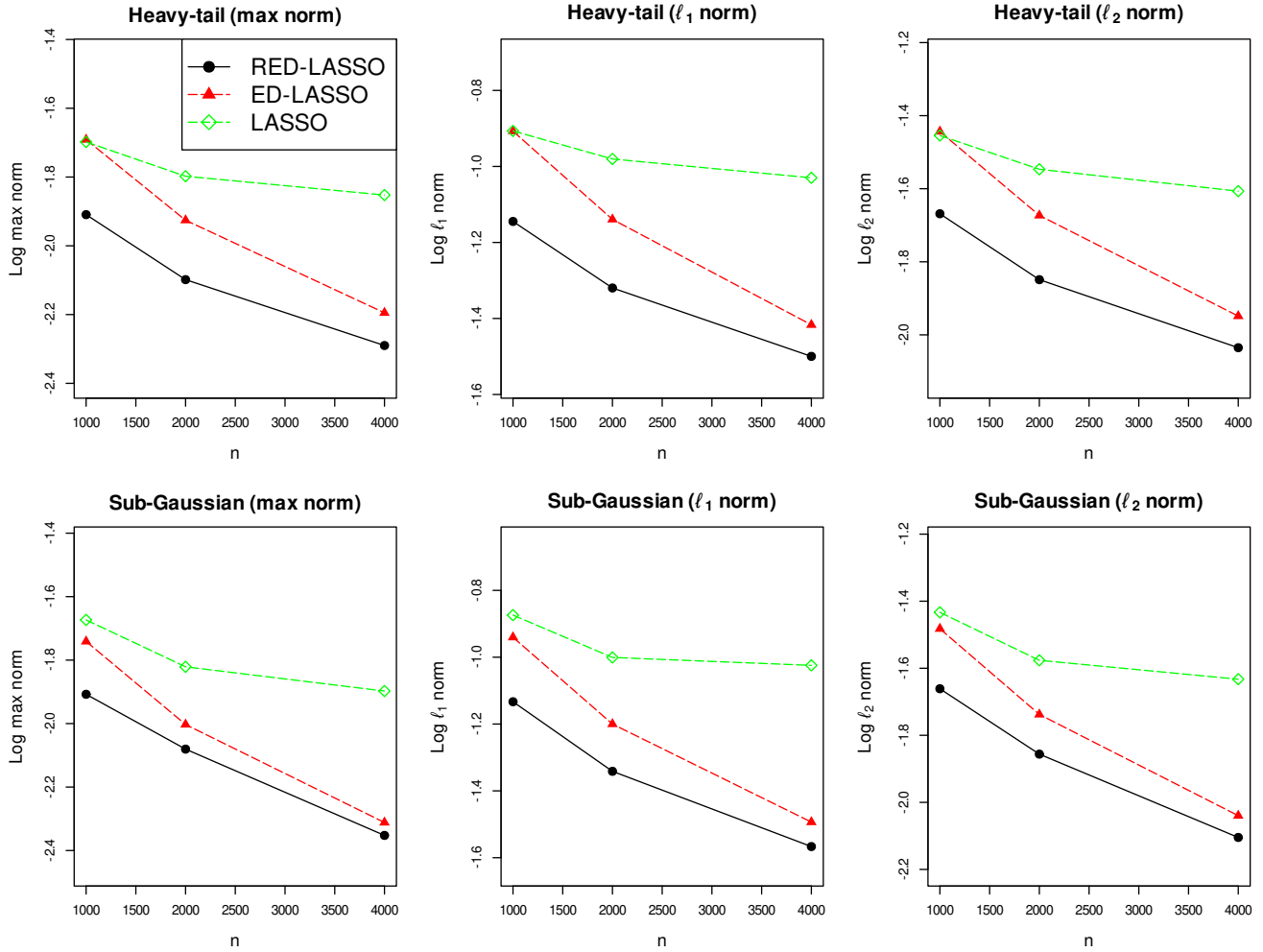
Figure 1: The log max, ℓ_1, and ℓ_2 norm error plots (corresponding to three columns) of the RED-LASSO (black dot), ED-LASSO (red triangle), and LASSO (green diamond) estimators for $p = 100$ and $n = 1000, 2000, 4000$.

estimators cannot. For the sub-Gaussian process, the RED-LASSO and ED-LASSO estimators showed better performance than the LASSO estimator. This is because the LASSO estimator cannot account for the time variation of the beta process. We note that, even for the sub-Gaussian process, the RED-LASSO estimator showed better performance than the ED-LASSO estimator. One possible explanation for this is that the true return process can have some extreme values over time even if the sub-Gaussian random variables are used. From this result, we can conjecture that the RED-LASSO estimator is robust to the heavy-tailedness of the log-return process.

5 An empirical study

In this section, we applied the proposed RED-LASSO estimator to high-frequency trading data from January 2013 to December 2019. We took stock price data, futures price data, and firm fundamentals from the End of Day website, FirstRate Data website, and Center for Research in Security Prices (CRSP)/Compustat Merged Database, respectively. We obtained 5-min log-price data with the previous tick scheme (Wang and Zou, 2010; Zhang, 2011) and processed the data similar to the procedure in Kim and Shin (2022). The days with half trading hours were not included. For the dependent process, we collected the log-price data of the following five assets: Apple Inc. (AAPL), Berkshire Hathaway Inc. (BRK.B), General Motors Company (GM), Alphabet Inc. (GOOG), and Exxon Mobil Corporation (XOM). These firms have the top market values in their global industry classification standard (GICS) sectors. For the covariate process, we first obtained the log-prices of 54 futures, which are often used as the market macro variables. For example, we selected 20 commodity data, 10 currency data, 10 interest rate data, and 14 stock market index data. The specific list is presented in Table 4 in the Appendix. Then, we constructed Fama-French five factors (Fama and French, 2015) and the momentum factor (Carhart, 1997) with the assets listed on NYSE, NASDAQ, and AMEX, which are widely used in the stock market analysis. We note that the MKT, HML, SMB, RMW, CMA, and MOM represent the market, value, size, profitability, investment, and momentum factors, respectively. First, we calculated MKT as the return of a value-weighted portfolio of whole assets. Then, we obtained other factors 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 represent the small and big market equities, respectively, while we classified high (H), medium (M), and low (L) portfolios according to their ratio of book equity to market equity. On the other hand, robust (R), neutral (N), and weak (W) portfolios were classified by their profitability, while we obtained conservative (C), neutral (N), and aggressive (A) portfolios using their investment data. Also, up (U), flat (F), and down (D) portfolios were classified by their momentum of the return. The portfolio constituents were updated monthly, and, with 5-min frequency, we obtained the portfolio return as follows:

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

where $WRet_{d,i}$ is the portfolio return for the dth day and ith time interval, N_d is the number of portfolio components on the dth day, the superscript j is used to represent the jth stock of the portfolio, and $w_{d,i}^j$ is calculated by

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

where w_d^j is the market capitalization of the jth stock at the market close time on the day $d-1$, and $Ret_{d,0}^j$ represents the overnight return from the day $(d-1)$ to day d. To sum up, the five assets and 60 factors were used for the dependent and covariate processes, respectively. The details of the data processing can be found in Aït-Sahalia et al. (2020) and Kim and Shin (2022).

To determine the tuning parameters c_τ, c_ϖ, and c_h, we employed the mean squared prediction

error (MSPE) with the data in 2013. For the choice of c_τ, we defined

$$\Lambda(c_\tau) = \frac{1}{60} \sum_{s=1}^{5} \sum_{j=1}^{12} \sum_{i=0}^{[1/(k_n\Delta_n)]-2} \left\| \widehat{\boldsymbol{\beta}}_{i\Delta_n}^{j,s}(c_\tau) - \widehat{\boldsymbol{\beta}}_{(i+1)\Delta_n}^{j,s}(\infty) \right\|_2^2,$$

where $\widehat{\boldsymbol{\beta}}_{i\Delta_n}^{j,s}(c_\tau)$ is the instantaneous beta estimator at time $i\Delta_n$ with the tuning parameter c_τ for the jth month in 2013 and sth stock. Then, we selected c_τ by minimizing $\Lambda(c_\tau)$ over $c_\tau \in \{2^l \mid 0 \leq l \leq 10, l \in \mathbb{Z}\}$. Based on the selected c_τ, we defined

$$\Lambda(c_\varpi) = \frac{1}{60} \sum_{s=1}^{5} \sum_{j=1}^{12} \sum_{i=0}^{[1/(k_n\Delta_n)]-2} \left\| \widetilde{\boldsymbol{\beta}}_{i\Delta_n}^{j,s}(c_\varpi) - \widehat{\boldsymbol{\beta}}_{(i+1)\Delta_n}^{j,s} \right\|_2^2,$$

where $\widetilde{\boldsymbol{\beta}}_{i\Delta_n}^{j,s}(c_\varpi)$ is the debiased instantaneous beta estimator at time $i\Delta_n$ with the tuning parameter c_ϖ for the jth month in 2013 and sth stock. We chose c_ϖ which minimizes $\Lambda(c_\varpi)$ over $c_\varpi \in \{2^l \mid -10 \leq l \leq 0, l \in \mathbb{Z}\}$. Finally, with the selected c_τ and c_ϖ, we defined

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

where $\widetilde{I\beta}^{j,s}(c_h)$ is the RED-LASSO estimator with the tuning parameter c_h and $\widehat{I\beta}^{j,s}$ is the debiased integrated beta estimator for the jth month in 2013 and sth stock. Then, we selected c_h by minimizing $\Lambda(c_h)$ over $c_h \in \{2^l \mid -5 \leq l \leq 5, l \in \mathbb{Z}\}$. The results are $c_\tau = 16$, $c_\varpi = 1/64$, and $c_h = 1/4$. We note that the stationarity assumption for the beta process is reasonable, which motivates and justifies the above tuning parameter selection procedure. Then, using the RED-LASSO, ED-LASSO, and LASSO estimation procedures, we obtained the monthly integrated betas for each of the five assets. The tuning parameters were selected based on Section 3.3 and Section 4. For the non-trading period, we set the beta estimates as zero.

We first compare the performances of the RED-LASSO, ED-LASSO, and LASSO estimators.

Table 1: The average in-sample and out-of-sample R^2 of the RED-LASSO, ED-LASSO, and LASSO estimators over the five assets.

| | In-sample R^2 | | |
| | Estimator | | |
	RED-LASSO	ED-LASSO	LASSO
whole period	0.261	0.196	0.220
2013	0.254	0.151	0.202
2014	0.233	0.201	0.187
2015	0.282	0.272	0.257
2016	0.267	0.085	0.214
2017	0.206	0.137	0.158
2018	0.339	0.335	0.315
2019	0.247	0.191	0.208
	Out-of-sample R^2		
	Estimator		
	RED-LASSO	ED-LASSO	LASSO
whole period	0.248	0.167	0.216
2014	0.214	0.136	0.181
2015	0.270	0.234	0.245
2016	0.248	0.082	0.210
2017	0.194	0.094	0.152
2018	0.329	0.281	0.304
2019	0.231	0.173	0.202

To do this, we calculated the monthly in-sample and out-of-sample R^2 with the monthly integrated beta estimates. The out-of-sample R^2 was calculated using the integrated betas from the previous month, and it was obtained excluding the year 2013 since the tuning parameters were chosen based on the data in 2013. For each year, we calculated the average R^2 across the five assets and twelve months. Table 1 shows the average in-sample and out-of-sample R^2 of the RED-LASSO, ED-LASSO, and LASSO estimators. As seen in Table 1, the RED-LASSO estimator shows the best performance for all periods. This may be because only the RED-LASSO estimator can handle both the heavy-tailed distribution of the return process and time-varying property of the beta process.

Table 2 shows the non-zero frequency of the RED-LASSO, ED-LASSO, and LASSO estimators for the five assets and 60 factors over 84 months. Table 3 shows the monthly average of non-zero frequency over factors and time for the RED-LASSO, ED-LASSO, and LASSO estimators for the five assets. As seen in Tables 2 and 3, the RED-LASSO estimator can better account for the sparsity of the integrated betas than the ED-LASSO and LASSO estimators. From this result, we

Table 2: The number of non-zero monthly integrated beta estimates from the RED-LASSO (RED), ED-LASSO (ED), and LASSO estimators for the five assets and 60 factors over 84 months.

Type	Symbol	AAPL			BRK.B			GM			GOOG			XOM		
		RED	ED	LASSO	RED	ED	LASSO	RED	ED	LASSO	RED	ED	LASSO	RED	ED	LASSO
Commodity	CA	0	20	0	0	22	0	0	27	0	1	28	0	0	29	0
	CL	0	15	0	0	15	2	1	21	1	1	15	0	11	34	48
	GC	1	17	0	2	11	0	4	25	0	2	15	0	0	25	1
	HG	1	16	0	1	23	0	1	19	1	4	18	0	3	16	4
	HO	0	12	0	0	20	0	3	12	1	0	7	0	2	14	42
	ML	0	20	0	0	15	0	0	22	0	1	17	0	0	12	1
	NG	2	8	0	0	9	0	1	10	0	0	3	0	0	6	1
	OJ	1	11	0	0	9	0	0	15	0	0	23	0	1	15	0
	PA	0	9	0	1	7	0	1	11	0	1	13	0	0	11	1
	PL	2	7	0	1	14	0	0	22	0	2	14	0	0	15	1
	RB	1	9	0	2	15	0	2	14	1	0	17	0	2	12	36
	RM	0	15	0	0	14	0	0	15	0	0	12	0	0	10	0
	RS	0	9	0	0	10	0	0	7	0	0	7	0	0	6	0
	SI	0	18	0	1	16	0	3	13	0	2	18	0	0	17	1
	ZC	0	21	0	1	19	0	3	26	0	0	16	0	0	16	0
	ZL	1	19	0	1	15	0	0	20	0	2	15	0	0	19	1
	ZM	1	13	0	1	17	0	1	19	0	1	19	0	2	14	0
	ZO	0	10	0	0	16	0	1	16	0	1	19	0	1	16	0
	ZR	0	12	0	0	14	0	0	12	0	0	16	0	0	17	0
	ZW	1	15	0	0	13	0	0	16	0	3	23	0	0	12	0
Currency	A6	1	15	0	2	23	1	1	24	2	1	17	0	1	11	6
	AD	0	20	0	2	13	0	2	14	2	3	11	0	4	18	13
	B6	0	19	0	0	17	0	4	23	0	1	20	0	1	21	0
	BR	0	6	0	0	14	0	0	10	0	0	13	0	1	11	1
	DX	3	25	1	0	15	0	1	26	0	0	16	0	1	16	1
	E1	1	15	0	2	17	0	2	22	0	1	18	0	0	14	0
	E6	1	16	0	0	26	0	0	25	0	0	15	0	1	17	0
	J1	2	18	2	2	18	9	3	23	5	1	20	1	3	23	3
	RP	0	15	0	0	11	0	1	24	0	0	14	0	0	21	0
	RU	0	7	0	0	9	0	0	9	0	0	11	0	1	8	1
Interest rate	BTP	0	39	0	1	33	0	2	44	0	0	30	0	0	30	0
	ED	0	2	0	0	4	0	0	10	0	0	3	0	0	7	0
	G	0	46	0	2	47	0	2	39	1	1	41	0	1	44	0
	GG	0	27	0	0	19	2	2	27	1	3	20	0	0	27	0
	HR	0	9	0	2	14	0	1	14	0	3	16	0	0	17	0
	US	1	15	1	0	9	5	3	21	1	1	14	0	3	15	1
	ZF	1	14	0	2	12	4	0	19	1	0	10	0	3	14	0
	ZN	0	10	1	0	13	4	1	13	2	1	13	0	2	15	0
	ZQ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	ZT	1	14	1	0	9	0	0	8	2	1	10	0	0	3	0
Stock market index	DY	3	34	18	0	38	26	3	40	18	3	30	10	3	43	25
	ES	53	45	81	71	71	84	14	36	67	64	54	81	30	35	76
	EW	8	42	18	20	35	56	28	49	58	6	46	13	6	43	41
	FX	1	25	9	2	35	19	3	27	20	1	26	5	0	21	27
	MME	18	25	28	8	24	33	12	29	26	26	34	23	8	25	51
	MX	1	35	17	1	24	32	3	30	24	4	25	7	1	27	38
	NQ	84	84	84	14	40	37	8	40	36	84	84	84	13	51	21
	RTY	9	26	27	4	28	27	14	34	48	5	39	24	3	29	22
	VX	14	34	29	14	24	37	5	20	21	21	30	23	5	15	29
	X	3	26	11	6	26	32	7	30	26	4	27	5	12	43	58
	XAE	0	19	0	1	17	1	2	22	5	0	20	0	66	71	57
	XAF	1	17	3	39	50	35	4	27	12	2	16	0	1	18	11
	XAI	2	23	2	2	24	4	1	18	5	0	15	0	2	17	6
	YM	43	60	62	67	60	84	13	31	56	6	27	50	55	63	84
Six factors	HML	9	23	0	28	37	12	26	50	9	22	38	1	23	45	42
	SMB	8	30	0	65	68	19	10	29	1	5	22	0	57	63	19
	RMW	2	13	0	25	44	7	37	51	4	8	24	0	59	65	49
	CMA	2	18	2	5	24	0	11	35	4	18	32	5	45	56	39
	MOM	6	30	2	29	46	11	40	54	23	22	37	5	63	72	39
	MKT	13	51	32	84	83	84	82	82	83	19	35	32	84	84	84

can conjecture that the proposed RED-LASSO provides more sparse beta estimates, which is the important property in practice. Furthermore, as discussed above, the RED-LASSO estimator shows

Table 3: The monthly average of non-zero frequency over 60 factors and 84 months for the RED-LASSO, ED-LASSO, and LASSO estimators for the five assets.

| | AAPL | | | BRK.B | | | GM | | | GOOG | | | XOM | | |
	RED	ED	LASSO	RED	ED	LASSO	RED	ED	LASSO	RED	ED	LASSO	RED	ED	LASSO
Non-zero frequency	3.595	15.095	5.130	6.083	16.845	7.940	4.392	17.511	6.750	4.261	15.333	4.392	6.904	18.261	11.678

the best performance in terms of R^2 in Table 1. That is, the RED-LASSO estimator can explain the market dynamics well with a simpler model. We note that for the RED-LASSO estimates, the stock market index futures factors had non-zero integrated betas more often than the other futures factors. This result is consistent with the multi-factor models (Asness et al., 2013; Carhart, 1997; Fama and French, 1992, 2015) since the market factors can be partially explained by the stock market index futures factors.

Now, we investigate the result of the RED-LASSO estimator. Figure 2 shows the monthly integrated betas from the RED-LASSO estimator for the five assets and 60 factors. Figure 3 depicts the non-zero frequency of the RED-LASSO estimator for the five groups, consisting of the commodity futures group, currency futures group, interest rate futures group, stock market index futures group, and market factor group. From Figures 2 and 3, we see that integrated betas change over time, and only a small number of factors had non-zero integrated betas in most periods. To investigate time-series of the significant betas, we plotted the integrated beta estimates for the three factors that most frequently had non-zero integrated betas in Figure 4. The AAPL has NQ (E-mini Nasdaq 100), ES (E-mini S&P 500), and YM (E-mini Dow); BRK.B has MKT, ES, and YM; GM has MKT, MOM, and RMW; GOOG has NQ, ES, and MME (MSCI Emerging Markets Index); and XOM has MKT, XAE (E-mini Energy Select Sector), and MOM. In sum, either the NQ factor or MKT factor most frequently had non-zero integrated betas, while the other factors had non-zero integrated betas only for some time periods.

When modeling regression-based financial models, we often employ the six factors, MKT, HML, SMB, RMW, CMA, and MOM (Asness et al., 2013; Carhart, 1997; Fama and French, 2015, 2016). To investigate their beta behaviors in more detail, we plotted the integrated betas with the RED-
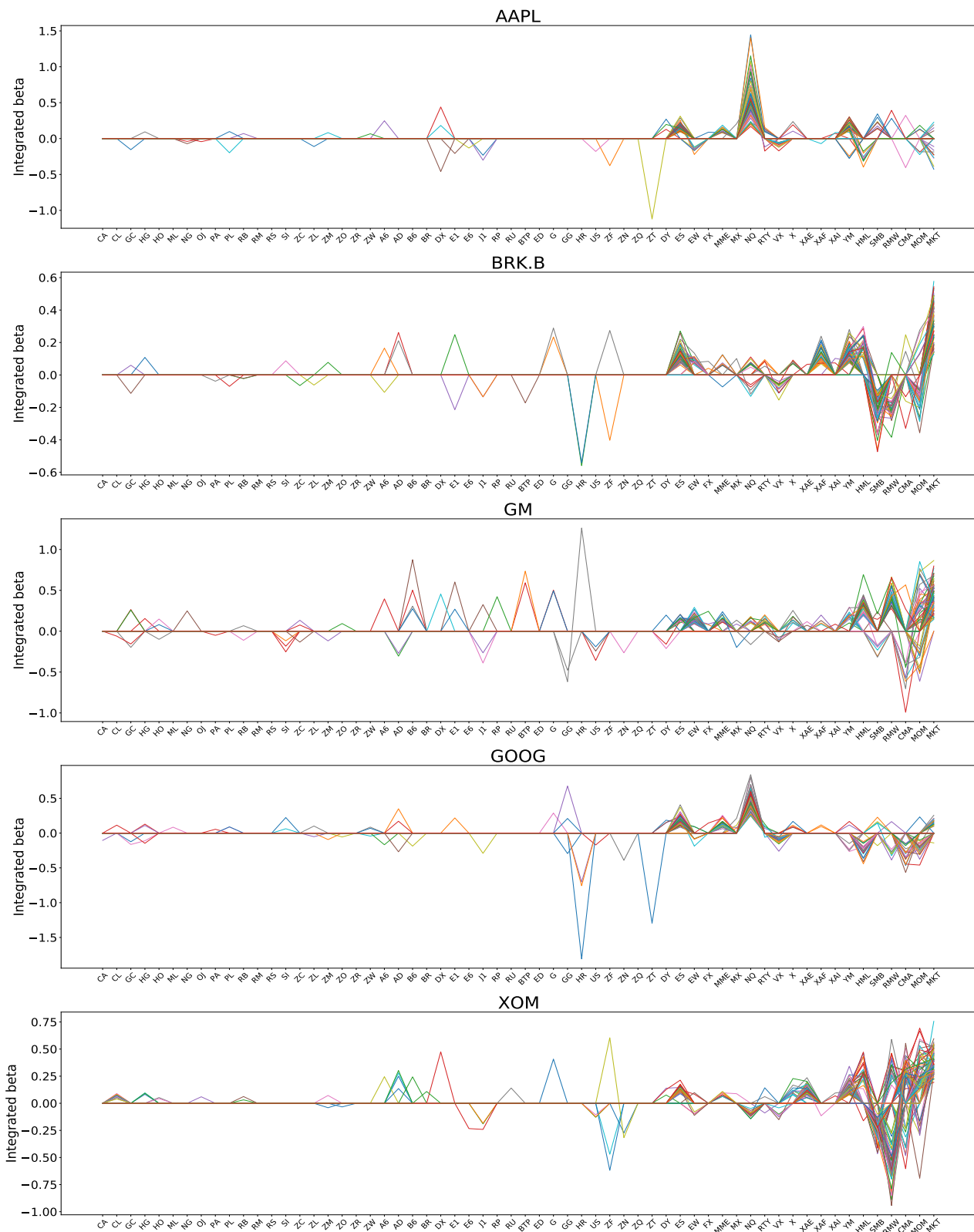
Figure 2: The monthly integrated betas from the RED-LASSO estimator for the five assets and 60 factors. Each line connects the 60 integrated beta estimates for each month.
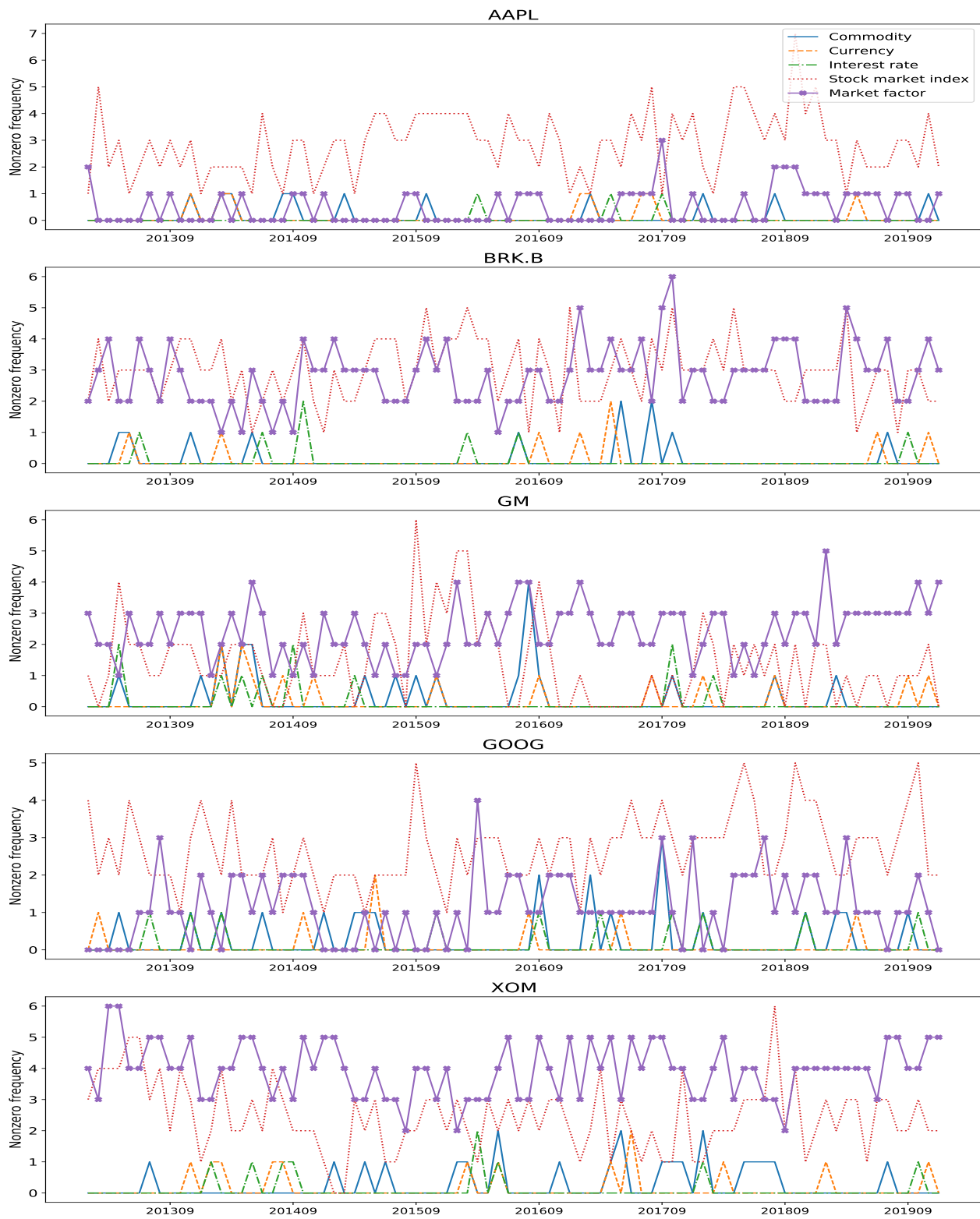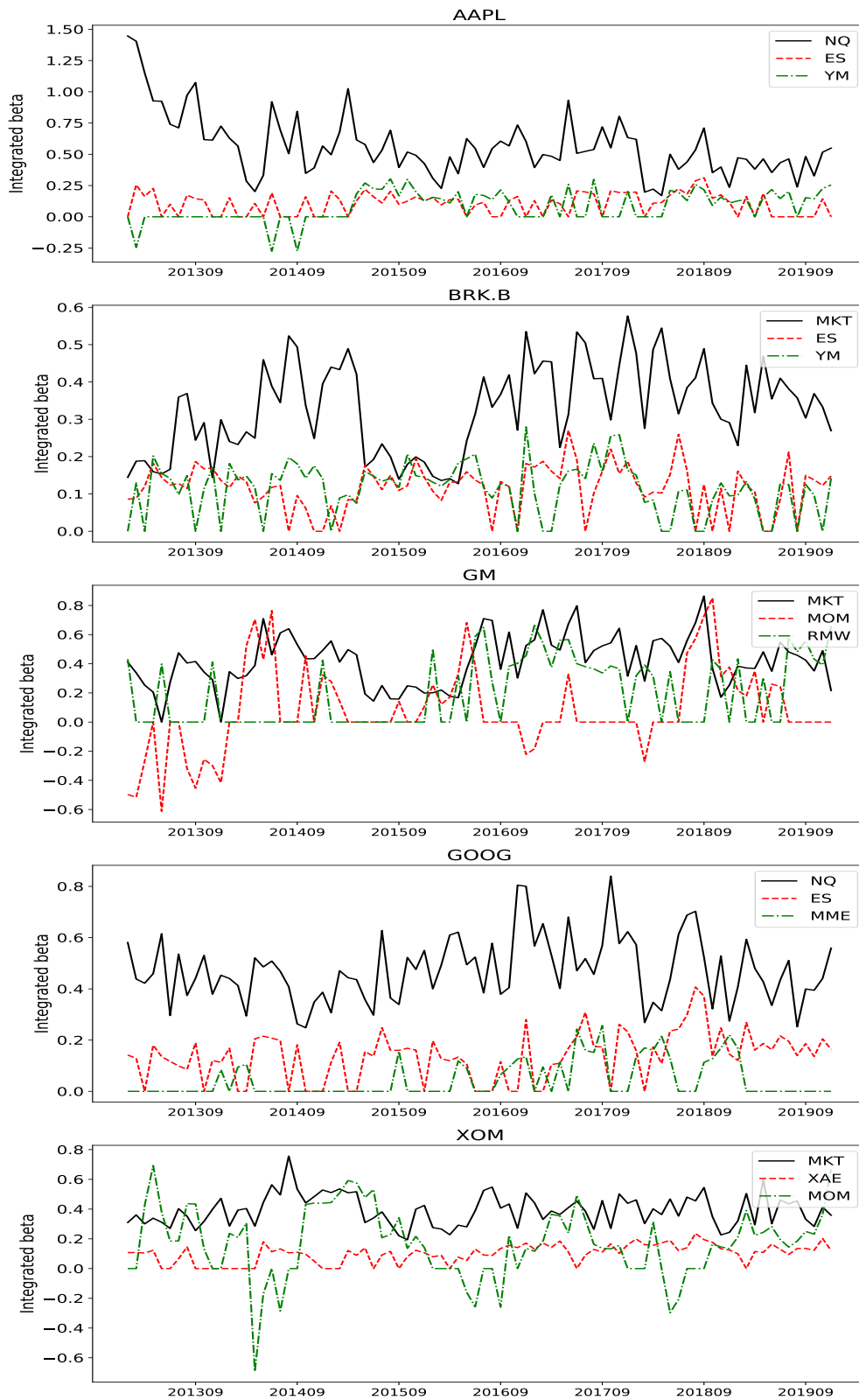
Figure 3: The nonzero frequency of the monthly integrated betas from the RED-LASSO estimator for the five assets and five groups. The five groups consist of the commodity futures group, currency futures group, interest rate futures group, stock market index futures group, and market factor group.

Figure 4: The monthly integrated betas from the RED-LASSO estimator for the three factors that most frequently had non-zero integrated betas among the 60 factors for each of the five assets.
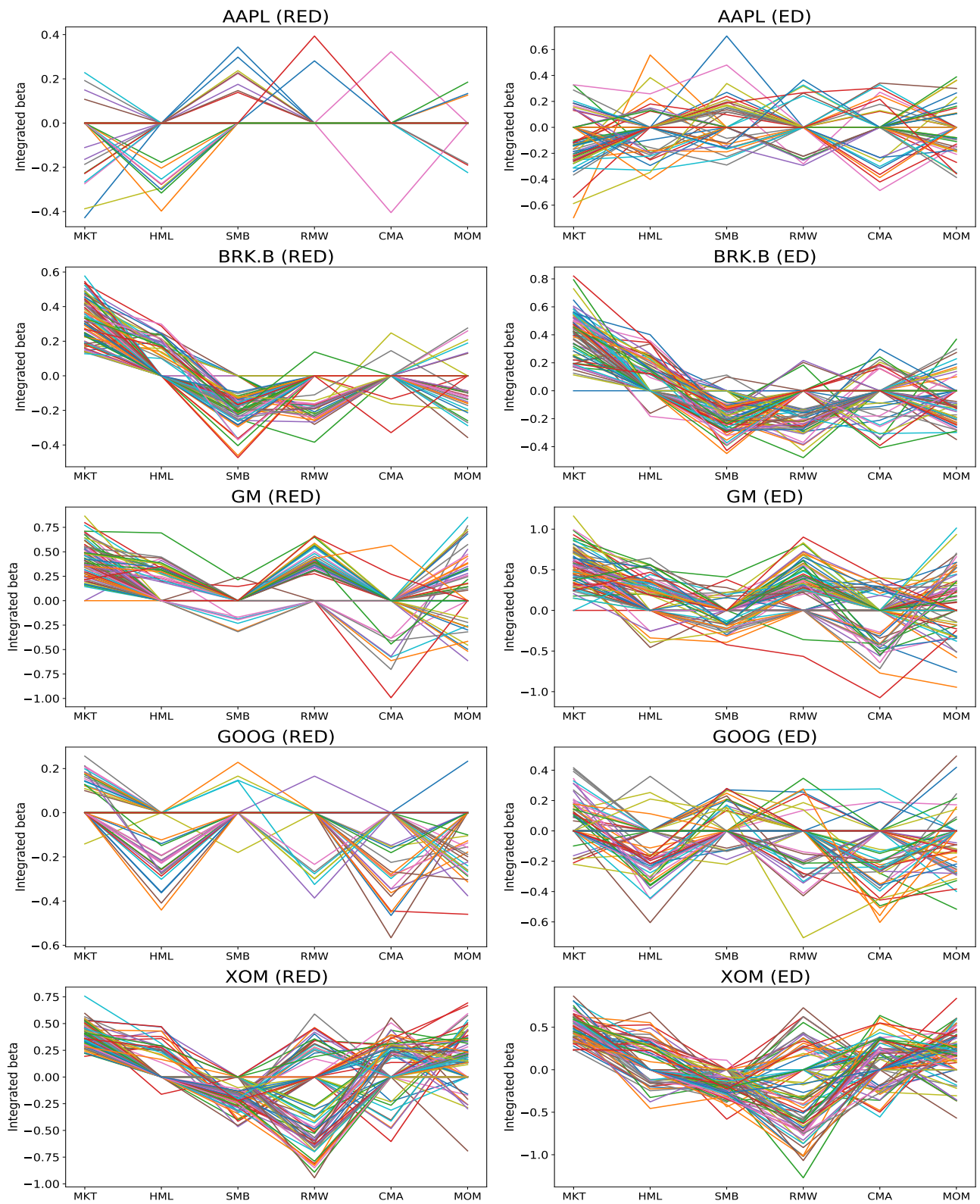
Figure 5: The monthly integrated betas from the RED-LASSO and ED-LASSO estimators for MKT, HML, SMB, RMW, CMA, and MOM over 84 months. Each line connects the six integrated beta estimates for each month.

LASSO and ED-LASSO estimators for these six factors in Figure 5. As expected, the MKT factor played a significant role for BRK.B, GM, and XOM; however, the six factors had zero integrated betas in most periods for the AAPL and GOOG. This may be because technology companies, such as AAPL and GOOG, have recently shown outstanding performance in the U.S. market; thus, the NQ (E-mini Nasdaq-100) factor can explain their movements well, as shown in Figure 4. We note that the results of the two other estimators are similar, but the RED-LASSO estimator has the more stable result. Thus, we can conjecture that considering both heavy-tailed distribution and time variation of beta process helps better explain the beta dynamics.

6   Conclusion

In this paper, we developed a novel RED-LASSO estimation procedure that can handle the heavy-tailedness of financial data and account for the time variation and sparsity of the high-dimensional beta process. To estimate the instantaneous beta, we propose a robust estimator that employs the Huber loss, truncation method, and ℓ_1-penalty. We demonstrated that the proposed instantaneous beta estimator can handle the heavy-tailedness and the curse of dimensionality with a desirable convergence rate. To handle the heavy-tailed bias coming from the Huber loss and ℓ_1-penalty, we developed a robust debiasing scheme and propose an integrated beta estimator. We showed that the proposed debiasing method sufficiently mitigates the effect of the bias, and the integrated beta estimator can enjoy the law of large number property. Then, the debiased integrated beta estimator is further regularized to account for the sparsity of the integrated beta. We demonstrated that the proposed RED-LASSO estimator can achieve the near-optimal convergence rate.

In the empirical study, the RED-LASSO estimation procedure shows the best performance in terms of R^2 and the sparsity of the beta estimates. It suggests that when estimating integrated beta in the high-dimensional high-frequency set-up, the RED-LASSO estimation method helps

account for the features of the time-varying beta process and heavy-tailed distributions of observed log-returns. On the other hand, we did not consider microstructure noises. The microstructure noise could be another source of the heavy tails and accommodating them leads to an application for higher frequency observations. However, if we impose the microstructure noise structure on the regression diffusion model, we have an unbalanced order relationship between the noise and regression variables, which ruins the usual regression structure. Hence, it is difficult to apply the existing estimation methods. It would be interesting and important to develop a robust estimation method that can handle microstructure noises. We leave this issue for a future study.

Funding

References

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

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Wu, G. (2006). *Realized beta: Persistence and predictability*. Emerald Group Publishing Limited.

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

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.

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.

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.

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

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

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Engle, R. F. and Gallo, G. M. (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics*, 131(1):3–27.

Fama, E. F. and French, K. R. (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.

Fama, E. F. and French, K. R. (2016). Dissecting anomalies with a five-factor model. *The Review of Financial Studies*, 29(1):69–103.

Fan, J. and Kim, D. (2018). Robust high-dimensional volatility matrix estimation for high-frequency factor model. *Journal of the American Statistical Association*, 113(523):1268–1283.

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., Liu, H., Sun, Q., and Zhang, T. (2018). I-LAMM for sparse learning: Simultaneous control of algorithmic complexity and statistical error. *Annals of statistics*, 46(2):814.

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.

Freedman, D. A. (1975). On tail probabilities for martingales. *The Annals of Probability*, 3(1):100–118.

Hansen, P. R., Huang, Z., and Shek, H. H. (2012). Realized garch: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, 27(6):877–906.

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.

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

Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101.

Jacod, J. and Protter, P. E. (2011). *Discretization of processes*, volume 67. Springer Science & Business Media.

Javanmard, A. and Montanari, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *The Journal of Machine Learning Research*, 15(1):2869–2909.

Javanmard, A. and Montanari, A. (2018). Debiasing the lasso: Optimal sample size for gaussian designs. *The Annals of Statistics*, 46(6A):2593–2622.

Kalnina, I. (2022). Inference for nonparametric high-frequency estimators with an application to time variation in betas. *Journal of Business & Economic Statistics*, pages 1–12.

Kim, D. and Fan, J. (2019). Factor garch-itô models for high-frequency data with application to large volatility matrix prediction. *Journal of econometrics*, 208(2):395–417.

Kim, D. and Shin, M. (2022). High-dimensional high-frequency regression. *arXiv preprint arXiv:2202.08419*.

Kim, D. and Wang, Y. (2016). Unified discrete-time and continuous-time models and statistical inferences for merged low-frequency and high-frequency financial data. *Journal of Econometrics*, 194:220–230.

Li, J., Todorov, V., and Tauchen, G. (2017). Adaptive estimation of continuous-time regression models using high-frequency data. *Journal of Econometrics*, 200(1):36–47.

Lintner, J. (1965). Security prices, risk, and maximal gains from diversification. *The journal of finance*, 20(4):587–615.

Mao, G. and Zhang, Z. (2018). Stochastic tail index model for high frequency financial data with bayesian analysis. *Journal of Econometrics*, 205(2):470–487.

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

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

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.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3):425–442.

Shephard, N. and Sheppard, K. (2010). Realising the future: forecasting with high-frequency-based volatility (heavy) models. *Journal of Applied Econometrics*, 25(2):197–231.

Shin, M., Kim, D., and Fan, J. (2023). Adaptive robust large volatility matrix estimation based on high-frequency financial data. *Journal of Econometrics*, 237(1).

Song, X., Kim, D., Yuan, H., Cui, X., Lu, Z., Zhou, Y., and Wang, Y. (2021). Volatility analysis with realized garch-itô models. *Journal of Econometrics*, 222(1):393–410.

Sun, Q., Zhou, W.-X., and Fan, J. (2020). Adaptive huber regression. *Journal of the American Statistical Association*, 115(529):254–265.

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

Van de Geer, S., Bühlmann, P., Ritov, Y., and Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*, 42(3):1166–1202.

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

Zhang, C.-H. and Zhang, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 76(1):217–242.

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

A   Appendix

A.1   Proof of Theorem 1

Without loss of generality, it is enough to show the statement for fixed i. For simplicity, we denote $\boldsymbol{\beta}_0(i\Delta_n)$ by $\boldsymbol{\beta}_0 = (\beta_{10}, \ldots, \beta_{p0})^\top$.

Proposition 1. *Under the assumptions in Theorem 1, we have*

$$\|\nabla \mathcal{L}_{\tau,i}(\boldsymbol{\beta}_0)\|_{\max} \leq \eta/2, \tag{A.1}$$

with probability greater than $1 - p^{-a}$ for any given positive constant a.

Proof of Proposition 1. Define

$$\mathcal{Y}_i^c = \begin{pmatrix} \Delta_{i+1}^n Y^c \\ \Delta_{i+2}^n Y^c \\ \vdots \\ \Delta_{i+k_n}^n Y^c \end{pmatrix}, \quad \mathcal{X}_i^c = \begin{pmatrix} \Delta_{i+1}^n \mathbf{X}^{c\top} \\ \Delta_{i+2}^n \mathbf{X}^{c\top} \\ \vdots \\ \Delta_{i+k_n}^n \mathbf{X}^{c\top} \end{pmatrix}, \quad \text{and} \quad \widetilde{\mathcal{X}}_i = \begin{pmatrix} \int_{i\Delta_n}^{(i+1)\Delta_n} (\boldsymbol{\beta}(t) - \boldsymbol{\beta}_0)^\top d\mathbf{X}^c(t) \\ \int_{(i+1)\Delta_n}^{(i+2)\Delta_n} (\boldsymbol{\beta}(t) - \boldsymbol{\beta}_0)^\top d\mathbf{X}^c(t) \\ \vdots \\ \int_{(i+k_n-1)\Delta_n}^{(i+k_n)\Delta_n} (\boldsymbol{\beta}(t) - \boldsymbol{\beta}_0)^\top d\mathbf{X}^c(t) \end{pmatrix}.$$

We have

$$\begin{aligned} (\mathcal{Y}_i^c)_k &= \int_{(i+k-1)\Delta_n}^{(i+k)\Delta_n} \boldsymbol{\beta}^\top(t) d\mathbf{X}^c(t) + \int_{(i+k-1)\Delta_n}^{(i+k)\Delta_n} dZ^c(t) \\ &= \boldsymbol{\beta}_0^\top \Delta_{i+k}^n \mathbf{X}^c + \Delta_{i+k}^n Z^c + \int_{(i+k-1)\Delta_n}^{(i+k)\Delta_n} (\boldsymbol{\beta}(t) - \boldsymbol{\beta}_0)^\top d\mathbf{X}^c(t) \\ &= \left(\mathcal{X}_i^c \boldsymbol{\beta}_0 + \mathcal{Z}_i + \widetilde{\mathcal{X}}_i \right)_k. \end{aligned}$$

Thus, for $1 \leq j \leq p$, we have

$$|\nabla_j \mathcal{L}_{\tau,i}(\boldsymbol{\beta}_0)| = \left| \frac{\partial \mathcal{L}_{\tau,i}(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}_j} \right| \leq (I)_j + (II)_j, \tag{A.2}$$

where

$$(I)_j = \frac{1}{k_n} \left| \sum_{h=1}^{k_n} \psi_\tau \left(\mathcal{Z}_{ih} + \widetilde{\mathcal{X}}_{ih} \right) \Delta_{i+h}^n X_j^c \right|,$$

$$(II)_j = \frac{1}{k_n} \left| \sum_{h=1}^{k_n} \left[\psi_\tau \left(\Delta_{i+h}^n Y - \langle \Delta_{i+h}^n \widehat{\mathbf{X}}^c, \boldsymbol{\beta}_0 \rangle \right) \Delta_{i+h}^n \widehat{X}_j^c \right. \right.$$
$$\left. \left. - \psi_\tau \left(\Delta_{i+h}^n Y^c - \langle \Delta_{i+h}^n \mathbf{X}^c, \boldsymbol{\beta}_0 \rangle \right) \Delta_{i+h}^n X_j^c \right] \right|.$$

First, we consider $(I)_j$. By the boundedness condition Assumption 1(b), we can show, with probability at least $1 - p^{-2-a}$,

$$\sup_{1 \le h \le k_n} \left| \Delta_{i+h}^n X_j^c \right| \le C_X \sqrt{\log p} \, n^{-1/2} \tag{A.3}$$

for some positive constant C_X. Then, we have

$$
\begin{aligned}
(I)_j \le{}& \frac{1}{k_n} \sum_{h=1}^{k_n} \left| \psi_\tau \left(\mathcal{Z}_{ih} + \widetilde{\mathcal{X}}_{ih} \right) \mathbf{1} \left(\left| \Delta_{i+h}^n X_j^c \right| > C_X \sqrt{\log p} \, n^{-1/2} \right) \right| \\
&+ \frac{1}{k_n} \sum_{h=1}^{k_n} \left| \mathbb{E} \left\{ \psi_\tau \left(\mathcal{Z}_{ih} + \widetilde{\mathcal{X}}_{ih} \right) \Delta_{i+h}^n X_j^c \mathbf{1} \left(\left| \Delta_{i+h}^n X_j^c \right| \le C_X \sqrt{\log p} \, n^{-1/2} \right) \middle| \mathcal{F}_{(i+h-1)\Delta_n} \right\} \right| \\
&+ \frac{1}{k_n} \left| \sum_{h=1}^{k_n} \psi_\tau \left(\mathcal{Z}_{ih} + \widetilde{\mathcal{X}}_{ih} \right) \Delta_{i+h}^n X_j^c \mathbf{1} \left(\left| \Delta_{i+h}^n X_j^c \right| \le C_X \sqrt{\log p} \, n^{-1/2} \right) \right. \\
&\qquad\qquad \left. - \mathbb{E} \left\{ \psi_\tau \left(\mathcal{Z}_{ih} + \widetilde{\mathcal{X}}_{ih} \right) \Delta_{i+h}^n X_j^c \mathbf{1} \left(\left| \Delta_{i+h}^n X_j^c \right| \le C_X \sqrt{\log p} \, n^{-1/2} \right) \middle| \mathcal{F}_{(i+h-1)\Delta_n} \right\} \right| \\
={}& (I)_j^{(1)} + (I)_j^{(2)} + (I)_j^{(3)}. \tag{A.4}
\end{aligned}
$$

For $(I)_j^{(1)}$, by (A.3), we have

$$\Pr \left\{ (I)_j^{(1)} = 0 \right\} \ge 1 - p^{-2-a}. \tag{A.5}$$

For $(I)_j^{(2)}$, let $f(h) = \psi_\tau \left(\mathcal{Z}_{ih} + \widetilde{\mathcal{X}}_{ih} \right) \Delta_{i+h}^n X_j^c \mathbf{1} \left(\left| \Delta_{i+h}^n X_j^c \right| \le C_X \sqrt{\log p} \, n^{-1/2} \right)$. Then, we have

$$
\begin{aligned}
&\left| \mathbb{E} \left\{ f(h) \middle| \mathcal{F}_{(i+h-1)\Delta_n} \right\} \right| \\
&\le \left| \mathbb{E} \left\{ \left(\mathcal{Z}_{ih} + \widetilde{\mathcal{X}}_{ih} \right) \Delta_{i+h}^n X_j^c \mathbf{1} \left(\left| \Delta_{i+h}^n X_j^c \right| \le C_X \sqrt{\log p} \, n^{-1/2} \right) \middle| \mathcal{F}_{(i+h-1)\Delta_n} \right\} \right|
\end{aligned}
$$

$$+ \left| \mathbb{E} \left\{ f(h) - \left(\mathcal{Z}_{ih} + \widetilde{\mathcal{X}}_{ih} \right) \Delta^n_{i+h} X^c_j \mathbf{1} \left(|\Delta^n_{i+h} X^c_j| \leq C_X \sqrt{\log p} n^{-1/2} \right) \Big| \mathcal{F}_{(i+h-1)\Delta_n} \right\} \right|.$$

Consider the first term. Similar to the proofs of Theorem 1 (Kim and Shin, 2022), we can show, for any constant $b \geq 1$,

$$
\Pr \left\{ \sup_{1 \leq h \leq k_n} \mathbb{E} \left\{ |\widetilde{\mathcal{X}}_{ih}|^b \Big| \mathcal{F}_{(i+h-1)\Delta_n} \right\} \leq \left(C s_p \Delta_n \sqrt{b k_n \log p} \right)^b \right\} \geq 1 - p^{-2-a} \quad \text{and}
$$
$$
\sup_{1 \leq h \leq k_n} \mathbb{E} \left\{ |\Delta^n_{i+h} X^c_j|^b \Big| \mathcal{F}_{(i+h-1)\Delta_n} \right\} \leq (C\Delta_n b)^{b/2} \quad \text{a.s.} \tag{A.6}
$$

Then, by the Cauchy–Schwarz inequality, we have, with probability at least $1 - p^{-2-a}$,

$$
\sup_{1 \leq h \leq k_n} \left| \mathbb{E} \left\{ \widetilde{\mathcal{X}}_{ih} \Delta^n_{i+h} X^c_j \mathbf{1} \left(|\Delta^n_{i+h} X^c_j| \leq C_X \sqrt{\log p} n^{-1/2} \right) \Big| \mathcal{F}_{(i+h-1)\Delta_n} \right\} \right|
$$
$$
\leq \sup_{1 \leq h \leq k_n} \sqrt{ \mathbb{E} \left\{ |\widetilde{\mathcal{X}}_{ih}|^2 \Big| \mathcal{F}_{(i+h-1)\Delta_n} \right\} } \sup_{1 \leq h \leq k_n} \sqrt{ \mathbb{E} \left\{ |\Delta^n_{i+h} X^c_j|^2 \Big| \mathcal{F}_{(i+h-1)\Delta_n} \right\} }
$$
$$
\leq C s_p \Delta_n^{3/2} \sqrt{k_n \log p}.
$$

Also, for $1 \leq h \leq k_n$, we have

$$
\mathbb{E} \left\{ \mathcal{Z}_{ih} \Delta^n_{i+h} X^c_j \mathbf{1} \left(|\Delta^n_{i+h} X^c_j| \leq C_X \sqrt{\log p} n^{-1/2} \right) \Big| \mathcal{F}_{(i+h-1)\Delta_n} \right\} = 0.
$$

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

$$
\sup_{1 \leq h \leq k_n} \left| \mathbb{E} \left\{ \left(\mathcal{Z}_{ih} + \widetilde{\mathcal{X}}_{ih} \right) \Delta^n_{i+h} X^c_j \mathbf{1} \left(|\Delta^n_{i+h} X^c_j| \leq C_X \sqrt{\log p} n^{-1/2} \right) \Big| \mathcal{F}_{(i+h-1)\Delta_n} \right\} \right|
$$
$$
\leq C s_p \Delta_n^{3/2} \sqrt{k_n \log p}. \tag{A.7}
$$

Consider the second term. By (A.6) and Hölder's inequality, we have, with probability at least $1 - p^{-2-a}$,

$$
\sup_{1 \leq h \leq k_n} \mathbb{E} \left\{ |\mathcal{Z}_{ih} + \widetilde{\mathcal{X}}_{ih}|^\gamma \Big| \mathcal{F}_{(i+h-1)\Delta_n} \right\} \leq C n^{-\gamma/2}. \tag{A.8}
$$

Then, using the fact that

$$\left|\psi_\tau\left(\mathcal{Z}_{ih} + \widetilde{\mathcal{X}}_{ih}\right) - \left(\mathcal{Z}_{ih} + \widetilde{\mathcal{X}}_{ih}\right)\right| \leq \left|\mathcal{Z}_{ih} + \widetilde{\mathcal{X}}_{ih}\right| \mathbf{1}\left(\left|\mathcal{Z}_{ih} + \widetilde{\mathcal{X}}_{ih}\right| > \tau\right)$$

$$\leq \tau^{-1}\left|\mathcal{Z}_{ih} + \widetilde{\mathcal{X}}_{ih}\right|^2 \quad \text{a.s.,}$$

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

$$\sup_{1 \leq h \leq k_n} \left|\mathbb{E}\left\{f(h) - \left(\mathcal{Z}_{ih} + \widetilde{\mathcal{X}}_{ih}\right)\Delta_{i+h}^n X_j^c \mathbf{1}\left(|\Delta_{i+h}^n X_j^c| \leq C_X \sqrt{\log p} n^{-1/2}\right) \Big| \mathcal{F}_{(i+h-1)\Delta_n}\right\}\right|$$

$$\leq \tau^{-1} \sup_{1 \leq h \leq k_n} \mathbb{E}\left\{|\mathcal{Z}_{ih} + \widetilde{\mathcal{X}}_{ih}|^2 |\Delta_{i+h}^n X_j^c| \Big| \mathcal{F}_{(i+h-1)\Delta_n}\right\}$$

$$\leq \tau^{-1} \sup_{1 \leq h \leq k_n} \left(\mathbb{E}\left\{|\mathcal{Z}_{ih} + \widetilde{\mathcal{X}}_{ih}|^\gamma \Big| \mathcal{F}_{(i+h-1)\Delta_n}\right\}\right)^{2/\gamma} \left(\mathbb{E}\left\{|\Delta_{i+h}^n X_j^c|^{\gamma/(\gamma-2)} \Big| \mathcal{F}_{(i+h-1)\Delta_n}\right\}\right)^{(\gamma-2)/\gamma}$$

$$\leq C\tau^{-1} n^{-3/2}, \tag{A.9}$$

where the second inequality is due to the Hölder's inequality and the last inequality is from (A.6) and (A.8). By (A.7) and (A.9), we have

$$\Pr\left\{(I)_j^{(2)} \leq C\left(s_p \Delta_n^{3/2}\sqrt{k_n \log p} + \tau^{-1} n^{-3/2}\right)\right\} \geq 1 - 2p^{-2-a}. \tag{A.10}$$

For $(I)_j^{(3)}$, by (2.1) in Freedman (1975), we have

$$\Pr\left\{(I)_j^{(3)} \geq s \text{ and } \sum_{h=1}^{k_n} \mathbb{E}\left\{(f(h))^2 \Big| \mathcal{F}_{(i+h-1)\Delta_n}\right\} \leq C n^{-2} k_n\right\}$$

$$\leq 2\exp\left\{-C k_n^2 s^2 \Big/ \left(\tau\sqrt{\log p} n^{-1/2} k_n s + n^{-2} k_n\right)\right\}. \tag{A.11}$$

Also, by (A.6) and (A.8), we have, with probability at least $1 - p^{-2-a}$,

$$\sup_{1 \leq h \leq k_n} \left|\mathbb{E}\left\{(f(h))^2 \Big| \mathcal{F}_{(i+h-1)\Delta_n}\right\}\right|$$

$$\leq \sup_{1 \leq h \leq k_n} \mathbb{E}\left\{|\mathcal{Z}_{ih} + \widetilde{\mathcal{X}}_{ih}|^2 |\Delta_{i+h}^n X_j^c|^2 \Big| \mathcal{F}_{(i+h-1)\Delta_n}\right\}$$

$$\leq \left(\sup_{1 \leq h \leq k_n} \mathbb{E} \left\{ |\mathcal{Z}_{ih} + \widetilde{\mathcal{X}}_{ih}|^\gamma \Big| \mathcal{F}_{(i+h-1)\Delta_n} \right\} \right)^{2/\gamma} \left(\sup_{1 \leq h \leq k_n} \mathbb{E} \left\{ |\Delta_{i+h}^n X_j^c|^{2\gamma/(\gamma-2)} \Big| \mathcal{F}_{(i+h-1)\Delta_n} \right\} \right)^{(\gamma-2)/\gamma}$$

$$\leq Cn^{-2},$$

where the second inequality is due to the Hölder's inequality. Thus, we have

$$Pr \left\{ (I)_j^{(3)} \leq C \left(\tau n^{-1/2} k_n^{-1} (\log p)^{3/2} + n^{-1} k_n^{-1/2} \sqrt{\log p} \right) \right\} \geq 1 - 2p^{-2-a}. \tag{A.12}$$

By (A.4), (A.5), (A.10), and (A.12), we have, with probability at least $1 - 5p^{-1-a}$,

$$\max_{1 \leq j \leq p} (I)_j \leq C \left[s_p \Delta_n^{3/2} \sqrt{k_n \log p} + \tau^{-1} n^{-3/2} + \tau n^{-1/2} k_n^{-1} (\log p)^{3/2} \right]. \tag{A.13}$$

Consider $(II)_j$. For some large constant $C > 0$, define

$$Q_1 = \{\max_{i,j} |\Delta_i^n X_j^c| \leq C \sqrt{\log p} \, n^{-1/2}\},$$

$$Q_2 = \{\max_{i,j} \int_{i\Delta_n}^{(i+k_n)\Delta_n} d\Lambda_j(t) \leq C \log p\} \cap \{\max_i \int_{i\Delta_n}^{(i+k_n)\Delta_n} d\Lambda^y(t) \leq C \log p\},$$

$$Q_3 = \{\max_{i,j} \sum_{k=1}^{k_n} \mathbf{1}_{\{|\Delta_{i+k}^n X_j| > v_{j,n}\}} \leq C \log p\}.$$

By (A.3), we have

$$Pr(Q_1) \geq 1 - p^{-2-a}.$$

By the boundedness of the intensity process, we have

$$Pr(Q_2) \geq 1 - p^{-2-a}.$$

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

$$\max_{i,j} \sum_{k=1}^{k_n} \mathbf{1}_{\{|\Delta_{i+k}^n X_j| > v_{j,n}\}} \leq \max_{i,j} \int_{i\Delta_n}^{(i+k_n)\Delta_n} d\Lambda_j(t) \leq C \log p.$$

Thus, we have

$$\Pr\left(Q_1 \cap Q_2 \cap Q_3\right) \geq 1 - 2p^{-2-a}. \tag{A.14}$$

We note that, for any $x_1, x_2, y_1, y_2 \in \mathbb{R}$,

$$|x_1 y_1 - x_2 y_2| \leq |(x_1 - x_2)(y_1 - y_2)| + |(x_1 - x_2) y_2| + |x_2 (y_1 - y_2)|.$$

Hence, under the event $Q_1 \cap Q_2 \cap Q_3$, we have

$$(II)_j$$
$$\leq \frac{1}{k_n} \sum_{h=1}^{k_n} \left[\left| \psi_\tau \left(\Delta_{i+h}^n Y - \langle \Delta_{i+h}^n \widehat{\mathbf{X}}^c, \boldsymbol{\beta}_0 \rangle \right) - \psi_\tau \left(\Delta_{i+h}^n Y^c - \langle \Delta_{i+h}^n \mathbf{X}^c, \boldsymbol{\beta}_0 \rangle \right) \right| \left| \Delta_{i+h}^n \widehat{X}_j^c - \Delta_{i+h}^n X_j^c \right| \right.$$
$$+ \left| \psi_\tau \left(\Delta_{i+h}^n Y - \langle \Delta_{i+h}^n \widehat{\mathbf{X}}^c, \boldsymbol{\beta}_0 \rangle \right) - \psi_\tau \left(\Delta_{i+h}^n Y^c - \langle \Delta_{i+h}^n \mathbf{X}^c, \boldsymbol{\beta}_0 \rangle \right) \right| \left| \Delta_{i+h}^n X_j^c \right|$$
$$\left. + \left| \psi_\tau \left(\Delta_{i+h}^n Y^c - \langle \Delta_{i+h}^n \mathbf{X}^c, \boldsymbol{\beta}_0 \rangle \right) \right| \left| \Delta_{i+h}^n \widehat{X}_j^c - \Delta_{i+h}^n X_j^c \right| \right]$$
$$\leq \frac{C}{k_n} \sum_{h=1}^{k_n} \left[\tau \left| \Delta_{i+h}^n \widehat{X}_j^c - \Delta_{i+h}^n X_j^c \right| \right.$$
$$\left. + \sqrt{\log p} \, n^{-1/2} \left| \psi_\tau \left(\Delta_{i+h}^n Y - \langle \Delta_{i+h}^n \widehat{\mathbf{X}}^c, \boldsymbol{\beta}_0 \rangle \right) - \psi_\tau \left(\Delta_{i+h}^n Y^c - \langle \Delta_{i+h}^n \mathbf{X}^c, \boldsymbol{\beta}_0 \rangle \right) \right| \right]$$
$$\leq C \left\{ \tau n^{-1}(\log p)^{3/2} + \tau n^{-1}(\log p)^{3/2} + k_n^{-1}\sqrt{\log p}\, n^{-1/2} \sum_{h=1}^{k_n} \left| \langle \Delta_{i+h}^n \widehat{\mathbf{X}}^c - \Delta_{i+h} \mathbf{X}^c, \boldsymbol{\beta}_0 \rangle \right| \right\}$$
$$\leq C \left\{ \tau n^{-1}(\log p)^{3/2} + s_p n^{-3/2}(\log p)^2 \right\} \quad \text{a.s.,}$$

which implies

$$\Pr\left(\max_{1 \leq j \leq p} (II)_j \leq C \left\{ \tau n^{-1}(\log p)^{3/2} + s_p n^{-3/2}(\log p)^2 \right\} \right) \geq 1 - 2p^{-1-a}. \tag{A.15}$$

44

Combining (A.2), (A.13), and (A.15), we have, with probability greater than $1 - p^{-a}$,

$$\|\nabla \mathcal{L}_{\tau,i}(\boldsymbol{\beta}_0)\|_\infty \leq C \left[s_p \Delta_n^{3/2} \sqrt{k_n \log p} + \tau^{-1} n^{-3/2} + \tau n^{-1/2} k_n^{-1} (\log p)^{3/2} \right]. \tag{A.16}$$

∎

Proof of Theorem 1. By Proposition 1, it is enough to show the statement under (A.1). First, we investigate $\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0$. Since

$$\mathcal{L}_{\tau,i}(\widehat{\boldsymbol{\beta}}_{i\Delta_n}) + \eta \|\widehat{\boldsymbol{\beta}}_{i\Delta_n}\|_1 \leq \mathcal{L}_{\tau,i}(\boldsymbol{\beta}_0) + \eta \|\boldsymbol{\beta}_0\|_1 ,$$

we have

$$
\begin{aligned}
\eta \left(\|\boldsymbol{\beta}_0\|_1 - \|\widehat{\boldsymbol{\beta}}_{i\Delta_n}\|_1 \right) &\geq \mathcal{L}_{\tau,i}(\widehat{\boldsymbol{\beta}}_{i\Delta_n}) - \mathcal{L}_{\tau,i}(\boldsymbol{\beta}_0) \\
&\geq \langle \nabla \mathcal{L}_{\tau,i}(\boldsymbol{\beta}_0), \widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0 \rangle \\
&\geq -\eta \|\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0\|_1/2.
\end{aligned}
$$

Then, we have

$$
\begin{aligned}
&\|(\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0)_{S_{i\Delta_n}}\|_1 + \|(\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0)_{S_{i\Delta_n}^c}\|_1 \\
&= \|\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0\|_1 \\
&\geq 2 \left(\|\widehat{\boldsymbol{\beta}}_{i\Delta_n}\|_1 - \|\boldsymbol{\beta}_0\|_1 \right) \\
&= 2 \left(\|(\widehat{\boldsymbol{\beta}}_{i\Delta_n})_{S_{i\Delta_n}^c}\|_1 + \|(\widehat{\boldsymbol{\beta}}_{i\Delta_n})_{S_{i\Delta_n}}\|_1 - \|(\boldsymbol{\beta}_0)_{S_{i\Delta_n}}\|_1 - \|(\boldsymbol{\beta}_0)_{S_{i\Delta_n}^c}\|_1 \right) \\
&\geq 2 \left(\|(\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0)_{S_{i\Delta_n}^c}\|_1 - \|(\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0)_{S_{i\Delta_n}}\|_1 - 2\|(\boldsymbol{\beta}_0)_{S_{i\Delta_n}^c}\|_1 \right).
\end{aligned}
$$

Thus, we have

$$\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0 \in \mathcal{W}_{i\Delta_n}, \tag{A.17}$$

where $W_{i\Delta_n}$ is defined in Assumption 1(e).

Now, we investigate $\|\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0\|_1$ and $\|\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0\|_2$. By (2.3), we have

$$(n\eta)^\delta |S_{i\Delta_n}| \leq s_p \quad \text{and} \quad \|(\boldsymbol{\beta}_0)_{S_{i\Delta_n}^c}\|_1 \leq \Sigma_{j \in S_{i\Delta_n}^c} |(\boldsymbol{\beta}_0)_j|^\delta |(\boldsymbol{\beta}_0)_j|^{1-\delta} \leq s_p(n\eta)^{1-\delta}. \tag{A.18}$$

Thus, by (A.17)–(A.18), we have

$$\begin{aligned}
\|\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0\|_1 &\leq 4\|(\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0)_{S_{i\Delta_n}}\|_1 + 4\|(\boldsymbol{\beta}_0)_{S_{i\Delta_n}^c}\|_1 \\
&\leq 4\sqrt{s_p}(n\eta)^{-\delta/2}\|\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0\|_2 + 4s_p(n\eta)^{1-\delta},
\end{aligned} \tag{A.19}$$

where the second inequality is due to the Cauchy–Schwarz inequality. Suppose that

$$\|\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0\|_2 > (1 + 12/\kappa)\sqrt{s_p}(n\eta)^{1-\delta/2}. \tag{A.20}$$

Then, we have

$$\|\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0\|_1 < \frac{8\kappa + 48}{\kappa + 12}\sqrt{s_p}(n\eta)^{-\delta/2}\|\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0\|_2. \tag{A.21}$$

From the optimality of $\widehat{\boldsymbol{\beta}}_{i\Delta_n}$ and the integral form of the Taylor expansion, we have

$$\begin{aligned}
0 &\geq \mathcal{L}_{\tau,i}(\widehat{\boldsymbol{\beta}}_{i\Delta_n}) - \mathcal{L}_{\tau,i}(\boldsymbol{\beta}_0) + \eta\left(\|\widehat{\boldsymbol{\beta}}_{i\Delta_n}\|_1 - \|\boldsymbol{\beta}_0\|_1\right) \\
&= \eta\left(\|\widehat{\boldsymbol{\beta}}_{i\Delta_n}\|_1 - \|\boldsymbol{\beta}_0\|_1\right) + \langle \nabla\mathcal{L}_{\tau,i}(\boldsymbol{\beta}_0), \widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0 \rangle \\
&\quad + \int_0^1 (1-t)(\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0)^\top \nabla^2\mathcal{L}_{\tau,i}(\boldsymbol{\beta}_0 + t(\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0))(\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0)dt.
\end{aligned} \tag{A.22}$$

For the first and second terms, we have

$$\begin{aligned}
&\eta\left(\|\widehat{\boldsymbol{\beta}}_{i\Delta_n}\|_1 - \|\boldsymbol{\beta}_0\|_1\right) + \langle \nabla\mathcal{L}_{\tau,i}(\boldsymbol{\beta}_0), \widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0 \rangle \\
&\geq -\eta\|\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0\|_1 - \|\nabla\mathcal{L}_{\tau,i}(\boldsymbol{\beta}_0)\|_{\max}\|\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0\|_1
\end{aligned}$$

$$\geq -\frac{12\kappa + 72}{\kappa + 12}\sqrt{s_p}n^{-\delta/2}\eta^{1-\delta/2}\|\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0\|_2, \tag{A.23}$$

where the second inequality is due to (A.21). For the last term, let

$$z = \frac{(\kappa + 12)(n\eta)^{\delta/2}D}{(8\kappa + 48)\sqrt{s_p}\|\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0\|_2} < 1.$$

Then, for any $0 \leq t \leq z$, we have

$$\|\boldsymbol{\beta}_0 + t(\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0) - \boldsymbol{\beta}_0\|_1 \leq z\|\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0\|_1 \leq D,$$

where the last inequality is due to (A.21). Thus, by Assumption 1(e), we have

$$\begin{aligned}
&\int_0^1 (1-t)\,(\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0)^\top \nabla^2 \mathcal{L}_{\tau,i}(\boldsymbol{\beta}_0 + t(\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0))(\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0)dt \\
&\geq \int_0^z (1-t)\kappa n^{-1}\|\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0\|_2^2 dt \\
&= (\kappa + 12)\sqrt{s_p}n^{-\delta/2}\eta^{1-\delta/2}\|\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0\|_2 - \frac{(\kappa + 12)^2}{2\kappa}s_p n^{1-\delta}\eta^{2-\delta}. \tag{A.24}
\end{aligned}$$

Combining (A.22)–(A.24), we have

$$\frac{(\kappa + 12)^2}{2\kappa}s_p n^{1-\delta}\eta^{2-\delta} \geq \frac{\kappa^2 + 12\kappa + 72}{\kappa + 12}\sqrt{s_p}n^{-\delta/2}\eta^{1-\delta/2}\|\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0\|_2,$$

which implies

$$\begin{aligned}
\|\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0\|_2 &\leq \frac{(\kappa + 12)^2}{2(\kappa^2 + 12\kappa + 72)}\frac{\kappa + 12}{\kappa}\sqrt{s_p}(n\eta)^{1-\delta/2} \\
&\leq (1 + 12/\kappa)\sqrt{s_p}(n\eta)^{1-\delta/2}.
\end{aligned}$$

This contradicts to (A.20), thus, we obtain the ℓ_2 norm error bound. Then, by (A.19), we can

show the ℓ_1 norm error bound. ∎

A.2 Proof of Theorem 2

Proof of Theorem 2. We first investigate $\widehat{\boldsymbol{\beta}}_{i\Delta_n}$ and $\widehat{\boldsymbol{\Omega}}_{i\Delta_n}$. By (2.3), (3.7), and Assumption 1(c), we can show, with probability at least $1 - p^{-2-a}$,

$$\sup_{0 \le i \le n-k_n} \|\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0((i+k_n)\Delta_n)\|_1 \le C s_p \left(s_p n^{-1/4} \left(\log p\right)^{3/4} \right)^{1-\delta} \quad \text{and}$$

$$\sup_{0 \le i \le n-k_n} \|\widehat{\boldsymbol{\beta}}_{i\Delta_n}\|_1 \le C s_p. \tag{A.25}$$

For $\widehat{\boldsymbol{\Omega}}_{i\Delta_n}$, similar to the proofs of Theorem 1 (Kim and Shin, 2022), we can show, with probability at least $1 - p^{-2-a}$,

$$\sup_{0 \le i \le n-k_n} \|\frac{1}{k_n\Delta_n}\mathcal{X}_i^\top \mathcal{X}_i \boldsymbol{\Omega}_0(i\Delta_n) - \mathbf{I}\|_{\max} \le \lambda. \tag{A.26}$$

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

$$\sup_{0 \le i \le n-k_n} \|\widehat{\boldsymbol{\Omega}}_{i\Delta_n}\|_1 \le C. \tag{A.27}$$

Consider $\widetilde{\boldsymbol{\beta}}_{i\Delta_n}$. For each $1 \le m \le p$, there exists standard Brownian motion $W_m^*(t)$ such that

$$d\beta_m(t) = \mu_{\beta,m}(t)dt + \sqrt{\Sigma_{\beta,mm}(t)}dW_m^*(t).$$

Then, by the proofs of Theorem 1 (Kim and Shin, 2022), we have

$$\frac{1}{k_n\Delta_n}\mathcal{X}_i^{c\top}\widetilde{\mathcal{X}}_i = \frac{1}{k_n\Delta_n}\mathcal{A}_i + R_i,$$

where

$$\mathcal{A}_i = \left(\sum_{m=1}^{p} \int_{i\Delta_n}^{(i+k_n)\Delta_n} \int_{i\Delta_n}^{t} \sqrt{\Sigma_{\beta,mm}(s)} dW_m^*(s) \Sigma_{jm}(t) dt \right)_{j=1,\dots,p} \qquad \text{and}$$

$$\Pr\left\{ \sup_{0 \leq i \leq n-k_n} \|R_i\|_{\max} \leq C s_p n^{-1/2} (\log p)^{3/2} \right\} \geq 1 - p^{-2-a}.$$

Note that

$$\Pr\left\{ \sup_{0 \leq i \leq n-k_n} \left\| \widetilde{\mathcal{X}}_i \right\|_{\max} \leq C s_p n^{-3/4} \log p \right\} \geq 1 - p^{-2-a}.$$

Hence, similar to the proofs of (A.15), we can show

$$\frac{1}{k_n \Delta_n} \mathcal{X}_i^\top \widetilde{\mathcal{X}}_i = \frac{1}{k_n \Delta_n} \mathcal{A}_i + R_i',$$

where

$$\Pr\left\{ \sup_{0 \leq i \leq n-k_n} \|R_i'\|_{\max} \leq C s_p n^{-1/2} (\log p)^{3/2} \right\} \geq 1 - p^{-1-a}. \tag{A.28}$$

Let

$$\widetilde{\boldsymbol{\beta}}_{i\Delta_n}^{(2)} = \widehat{\boldsymbol{\beta}}_{i\Delta_n} + \psi_\varpi \left(\frac{1}{k_n \Delta_n} \widehat{\boldsymbol{\Omega}}_{i\Delta_n}^\top \mathcal{X}_{(i+k_n)}^\top \left(\mathcal{Y}_{(i+k_n)}^c - \mathcal{X}_{(i+k_n)} \widehat{\boldsymbol{\beta}}_{i\Delta_n} \right) \right),$$

$$\widetilde{\boldsymbol{\beta}}_{i\Delta_n}^{(3)} = \widehat{\boldsymbol{\beta}}_{i\Delta_n} + \psi_\varpi \left(\frac{1}{k_n \Delta_n} \widehat{\boldsymbol{\Omega}}_{i\Delta_n}^\top \mathcal{X}_{(i+k_n)}^\top \left[\mathcal{X}_{(i+k_n)}^c \boldsymbol{\beta}_0((i+k_n)\Delta_n) - \mathcal{X}_{(i+k_n)} \widehat{\boldsymbol{\beta}}_{i\Delta_n} \right] + \mathcal{B}_{i+k_n} \right),$$

$$\widetilde{\boldsymbol{\beta}}_{i\Delta_n}^{(4)} = \widehat{\boldsymbol{\beta}}_{i\Delta_n} + \psi_\varpi \left(\frac{1}{k_n \Delta_n} \widehat{\boldsymbol{\Omega}}_{i\Delta_n}^\top \mathcal{X}_{(i+k_n)}^{c\top} \mathcal{X}_{(i+k_n)}^c \left[\boldsymbol{\beta}_0((i+k_n)\Delta_n) - \widehat{\boldsymbol{\beta}}_{i\Delta_n} \right] + \mathcal{B}_{i+k_n} \right),$$

$$\widetilde{\boldsymbol{\beta}}_{i\Delta_n}^{(5)} = \widehat{\boldsymbol{\beta}}_{i\Delta_n} + \psi_\varpi \left(\boldsymbol{\beta}_0((i+k_n)\Delta_n) - \widehat{\boldsymbol{\beta}}_{i\Delta_n} + \mathcal{B}_{i+k_n} \right),$$

where

$$\mathcal{B}_i = \frac{1}{k_n \Delta_n} \widehat{\boldsymbol{\Omega}}_{(i-k_n)\Delta_n}^\top \left(\mathcal{X}_i^\top \mathcal{Z}_i + \mathcal{A}_i \right).$$

Then, we have

$$
\begin{aligned}
\|\widehat{I\beta} - I\beta_0\|_{\max} \;\leq\;& \left\|\sum_{i=0}^{[1/(k_n\Delta_n)]-2} \left(\widetilde{\beta}_{ik_n\Delta_n} - \widetilde{\beta}^{(2)}_{ik_n\Delta_n}\right) k_n\Delta_n\right\|_{\max} \\
&+ \left\|\sum_{i=0}^{[1/(k_n\Delta_n)]-2} \left(\widetilde{\beta}^{(2)}_{ik_n\Delta_n} - \widetilde{\beta}^{(3)}_{ik_n\Delta_n}\right) k_n\Delta_n\right\|_{\max} \\
&+ \left\|\sum_{i=0}^{[1/(k_n\Delta_n)]-2} \left(\widetilde{\beta}^{(3)}_{ik_n\Delta_n} - \widetilde{\beta}^{(4)}_{ik_n\Delta_n}\right) k_n\Delta_n\right\|_{\max} \\
&+ \left\|\sum_{i=0}^{[1/(k_n\Delta_n)]-2} \left(\widetilde{\beta}^{(4)}_{ik_n\Delta_n} - \widetilde{\beta}^{(5)}_{ik_n\Delta_n}\right) k_n\Delta_n\right\|_{\max} \\
&+ \left\|\sum_{i=0}^{[1/(k_n\Delta_n)]-2} \left(\widetilde{\beta}^{(5)}_{ik_n\Delta_n} - \beta_0((i+1)k_n\Delta_n)\right) k_n\Delta_n\right\|_{\max} \\
&+ \left\|\sum_{i=0}^{[1/(k_n\Delta_n)]-2} \int_{ik_n\Delta_n}^{(i+1)k_n\Delta_n} \left(\beta_0((i+1)k_n\Delta_n) - \beta_0(t)\right) dt\right\|_{\max} \\
&+ \left\|\int_{([1/(k_n\Delta_n)]-1)k_n\Delta_n}^{1} \beta_0(t)dt\right\|_{\max} \\
=\;& (I) + (II) + (III) + (IV) + (V) + (VI) + (VII).
\end{aligned}
\tag{A.29}
$$

Consider (I). By the boundedness of the intensity, we can show $\Pr\left\{\int_0^1 d\Lambda^y(t) \leq C\log p\right\} \geq 1 - p^{-1-a}$. Thus, we have

$$
\Pr\left\{(I) \leq Cs_p^{2-\delta} n^{(-2+\delta)/4}(\log p)^{(5-3\delta)/4}\right\} \geq 1 - p^{-1-a}.
\tag{A.30}
$$

For (II), by (A.27)–(A.28), we have, with probability at least $1 - 2p^{-1-a}$,

$$
\begin{aligned}
(II) \;\leq\;& \sup_{0\leq i\leq n-2k_n} \left\|\widehat{\Omega}^{\top}_{i\Delta_n}\left(\frac{1}{k_n\Delta_n}\mathcal{X}^{\top}_{i+k_n}\widetilde{\mathcal{X}}_{i+k_n} - \frac{1}{k_n\Delta_n}\mathcal{A}_{i+k_n}\right)\right\|_{\max} \\
\leq\;& Cs_p n^{-1/2}(\log p)^{3/2}.
\end{aligned}
\tag{A.31}
$$

Consider (III). Similar to the proofs of (A.20) in Kim and Shin (2022), we can show, with

probability at least $1 - p^{-1-a}$,

$$
\begin{aligned}
(III) \quad &\leq \quad \sum_{i=0}^{[1/(k_n\Delta_n)]-2} \left\| \widehat{\boldsymbol{\Omega}}_{i\Delta_n}^{\top} \left(\mathcal{X}_{(i+k_n)}^{\top} \mathcal{X}_{(i+k_n)}^{c} - \mathcal{X}_{(i+k_n)}^{c\top} \mathcal{X}_{(i+k_n)}^{c} \right) \boldsymbol{\beta}_0((i+k_n)\Delta_n) \right\|_{\max} \\
&\quad + \sum_{i=0}^{[1/(k_n\Delta_n)]-2} \left\| \widehat{\boldsymbol{\Omega}}_{i\Delta_n}^{\top} \left(\mathcal{X}_{(i+k_n)}^{\top} \mathcal{X}_{(i+k_n)} - \mathcal{X}_{(i+k_n)}^{c\top} \mathcal{X}_{(i+k_n)}^{c} \right) \widehat{\boldsymbol{\beta}}_{i\Delta_n} \right\|_{\max} \\
&\leq \quad C s_p n^{-3/4} \sqrt{\log p}.
\end{aligned} \tag{A.32}
$$

Consider (IV). By Assumption 1(b) and (f), we can show, with probability at least $1 - p^{-1-a}$,

$$
\begin{aligned}
&\sup_{0 \leq i \leq n-k_n} \left\| \boldsymbol{\Sigma}_0(i\Delta_n) - \frac{1}{k_n\Delta_n} \mathcal{X}_i^{c\top} \mathcal{X}_i^c \right\|_{\max} \\
&\leq \sup_{0 \leq i \leq n-k_n} \left\| \boldsymbol{\Sigma}_0(i\Delta_n) - \frac{1}{k_n\Delta_n} \int_{i\Delta_n}^{(i+k_n)\Delta_n} \boldsymbol{\Sigma}_0(t)dt \right\|_{\max} \\
&\quad + \sup_{0 \leq i \leq n-k_n} \left\| \frac{1}{k_n\Delta_n} \int_{i\Delta_n}^{(i+k_n)\Delta_n} \boldsymbol{\Sigma}_0(t)dt - \frac{1}{k_n\Delta_n} \mathcal{X}_i^{c\top} \mathcal{X}_i^c \right\|_{\max} \\
&\leq C n^{-1/4} \sqrt{\log p}.
\end{aligned}
$$

Thus, by Assumption 1(f), we can show, with probability at least $1 - p^{-1-a}$,

$$
\sup_{0 \leq i \leq n-k_n} \left\| \frac{1}{k_n\Delta_n} \left(\mathcal{X}_{(i+k_n)}^{c\top} \mathcal{X}_{(i+k_n)}^c - \mathcal{X}_i^{c\top} \mathcal{X}_i^c \right) \right\|_{\max} \leq C n^{-1/4} \sqrt{\log p}.
$$

Then, by (A.14), (A.25), and (A.27), we have, with probability at least $1 - p^{-1-a}$,

$$
\begin{aligned}
(IV) \quad &\leq \quad \sup_{0 \leq i \leq n-2k_n} \left\| \frac{1}{k_n\Delta_n} \widehat{\boldsymbol{\Omega}}_{i\Delta_n}^{\top} \mathcal{X}_{(i+k_n)}^{c\top} \mathcal{X}_{(i+k_n)}^{c} - \mathbf{I} \right\|_{\max} \times \sup_{0 \leq i \leq n-2k_n} \left\| \boldsymbol{\beta}_0((i+k_n)\Delta_n) - \widehat{\boldsymbol{\beta}}_{i\Delta_n} \right\|_1 \\
&\leq \quad \sup_{0 \leq i \leq n-2k_n} \left[\left\| \frac{1}{k_n\Delta_n} \widehat{\boldsymbol{\Omega}}_{i\Delta_n}^{\top} \left(\mathcal{X}_{(i+k_n)}^{c\top} \mathcal{X}_{(i+k_n)}^{c} - \mathcal{X}_i^{c\top} \mathcal{X}_i^c \right) \right\|_{\max} \right. \\
&\quad + \left\| \frac{1}{k_n\Delta_n} \widehat{\boldsymbol{\Omega}}_{i\Delta_n}^{\top} \left(\mathcal{X}_i^{c\top} \mathcal{X}_i^c - \mathcal{X}_i^{\top} \mathcal{X}_i \right) \right\|_{\max} + \left. \left\| \frac{1}{k_n\Delta_n} \widehat{\boldsymbol{\Omega}}_{i\Delta_n}^{\top} \mathcal{X}_i^{\top} \mathcal{X}_i - \mathbf{I} \right\|_{\max} \right] \\
&\quad \times C s_p \left(s_p n^{-1/4} (\log p)^{3/4} \right)^{1-\delta} \\
&\leq \quad C \left(n^{-1/4} \sqrt{\log p} + n^{-1/2} (\log p)^2 + \lambda \right) \times s_p \left(s_p n^{-1/4} (\log p)^{3/4} \right)^{1-\delta}
\end{aligned}
$$

51

$$\leq C s_p^{2-\delta} n^{(-2+\delta)/4} (\log p)^{(5-3\delta)/4}. \tag{A.33}$$

For (V), let $g(i) = \boldsymbol{\beta}_0((i+k_n)\Delta_n) - \widehat{\boldsymbol{\beta}}_{i\Delta_n} + \mathcal{B}_{i+k_n}$. We have

$$
\begin{aligned}
(V) \leq{}& \left\| \sum_{i=0}^{[1/(k_n\Delta_n)]-2} \left[\psi_\varpi\left(g(ik_n)\right) - \mathbb{E}\left\{ \psi_\varpi\left(g(ik_n)\right) \Big| \mathcal{F}_{(i+1)k_n\Delta_n} \right\} \right] k_n\Delta_n \right\|_{\max} \\
&+ \left\| \sum_{i=0}^{[1/(k_n\Delta_n)]-2} \left[\mathbb{E}\left\{ \psi_\varpi\left(g(ik_n)\right) \Big| \mathcal{F}_{(i+1)k_n\Delta_n} \right\} - \boldsymbol{\beta}_0((i+1)k_n\Delta_n) + \widehat{\boldsymbol{\beta}}_{ik_n\Delta_n} \right] k_n\Delta_n \right\|_{\max} \\
={}& \left\| (V)^{(1)} \right\|_{\max} + \left\| (V)^{(2)} \right\|_{\max}.
\end{aligned}
$$

For the first term, by the boundedness of the intensity process and (A.27), we can show, with probability at least $1 - p^{-2-a}$,

$$\sup_{k_n \leq i \leq n-k_n} \sup_{1 \leq j \leq p} \mathbb{E}\left\{ \mathcal{B}_{ij}^2 \Big| \mathcal{F}_{i\Delta_n} \right\} \leq C s_p^2 n^{-1/2}.$$

Thus, from (A.25), we have, with probability at least $1 - 2p^{-2-a}$,

$$\sup_{0 \leq i \leq n-2k_n} \sup_{1 \leq j \leq p} \mathbb{E}\left[|(g(i))_j|^2 \Big| \mathcal{F}_{(i+k_n)\Delta_n} \right] \leq C s_p^2 \left(s_p n^{-1/4} (\log p)^{3/4} \right)^{2-2\delta}.$$

Then, by (2.1) in Freedman (1975), we have, for $1 \leq j \leq p$,

$$
\begin{aligned}
&\Pr\left[(V)_j^{(1)} \geq s \text{ and } \sum_{i=0}^{[1/(k_n\Delta_n)]-2} \mathbb{E}\left[|(g(ik_n))_j|^2 \Big| \mathcal{F}_{(i+1)k_n\Delta_n} \right] \leq C s_p^2 n^{1/2} \left(s_p n^{-1/4} (\log p)^{3/4} \right)^{2-2\delta} \right] \\
&\leq 2 \exp\left\{ -Cns^2 \Big/ \left(n^{1/2} \varpi s + s_p^{4-2\delta} n^{\delta/2} (\log p)^{(3-3\delta)/2} \right) \right\},
\end{aligned}
$$

which implies

$$\Pr\left[\left\| (V)^{(1)} \right\|_{\max} \leq C s_p^{2-\delta} n^{(-2+\delta)/4} (\log p)^{(5-3\delta)/4} \right] \geq 1 - 3p^{-1-a}.$$

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

$$
\begin{aligned}
&\left\|(V)^{(2)}\right\|_{\max} \\
&\leq \sup_{0 \leq i \leq [1/(k_n \Delta_n)]-2} \sup_{1 \leq j \leq p} \left| \mathbb{E}\left\{ \left[\psi_\varpi\left(g(ik_n)\right)\right]_j \middle| \mathcal{F}_{(i+1)k_n \Delta_n} \right\} - \left[\boldsymbol{\beta}_0((i+1)k_n \Delta_n) - \widehat{\boldsymbol{\beta}}_{ik_n \Delta_n} \right]_j \right| \\
&= \sup_{0 \leq i \leq [1/(k_n \Delta_n)]-2} \sup_{1 \leq j \leq p} \left| \mathbb{E}\left\{ \left[\psi_\varpi\left(g(ik_n)\right) - g(ik_n)\right]_j \middle| \mathcal{F}_{(i+1)k_n \Delta_n} \right\} \right| \\
&\leq \sup_{0 \leq i \leq [1/(k_n \Delta_n)]-2} \sup_{1 \leq j \leq p} \mathbb{E}\left\{ \left| [g(ik_n)]_j \right| \mathbf{1}\left(\left| [g(ik_n)]_j \right| > \varpi \right) \middle| \mathcal{F}_{(i+1)k_n \Delta_n} \right\} \\
&\leq \sup_{0 \leq i \leq [1/(k_n \Delta_n)]-2} \sup_{1 \leq j \leq p} \mathbb{E}\left\{ \left| [g(ik_n)]_j \right|^2 / \varpi \middle| \mathcal{F}_{(i+1)k_n \Delta_n} \right\} \\
&\leq C s_p^{2-\delta} n^{(-2+\delta)/4} (\log p)^{(5-3\delta)/4}.
\end{aligned}
$$

Thus, we have

$$
\Pr\left\{ (V) \leq C s_p^{2-\delta} n^{(-2+\delta)/4} (\log p)^{(5-3\delta)/4} \right\} \geq 1 - 4p^{-1-a}. \tag{A.34}
$$

Consider (VI). By the sub-Gaussianity of the beta process, we can show, with probability at least $1 - p^{-1-a}$,

$$
(VI) \leq C\sqrt{\log p/n}. \tag{A.35}
$$

For (VII), by Assumption 1(b), we have

$$
(VII) \leq Cn^{-1/2} \text{ a.s.} \tag{A.36}
$$

Combining (A.29)–(A.36), we have, with probability greater than $1 - p^{-a}$,

$$
\|\widehat{I\beta} - I\beta_0\|_{\max} \leq C \left[s_p^{2-\delta} n^{(-2+\delta)/4} (\log p)^{(5-3\delta)/4} + s_p n^{-1/2} (\log p)^{3/2} \right]. \tag{A.37}
$$

∎

A.3 Proof of Theorem 3

Proof of Theorem 3. By (3.8), there exists h_n such that, with probability greater than $1 - p^{-a}$,

$$\|\widehat{I\beta} - I\beta_0\|_{\max} \leq h_n/2.$$

Thus, it is enough to show the statement under the event $\{\|\widehat{I\beta} - I\beta_0\|_{\max} \leq h_n/2\}$. Similar to the proofs of Theorem 1 (Kim and Shin, 2022), we can show

$$\|\widetilde{I\beta} - I\beta_0\|_1 \leq C s_p h_n^{1-\delta}.$$

∎

Table 4: The symbols of 54 futures in Section 5.

Type	Symbol	Description
Commodity	CA	Cocoa
	CL	Crude Oil WTI
	GC	Gold
	HG	Copper
	HO	NY Harbor ULSD (Heating Oil)
	ML	Milling Wheat
	NG	Henry Hub Natural Gas
	OJ	Orange Juice
	PA	Palladium
	PL	Platinum
	RB	RBOB Gasoline
	RM	Robusta Coffee
	RS	Canola
	SI	Silver
	ZC	Corn
	ZL	Soybean Oil
	ZM	Soybean Meal
	ZO	Oats
	ZR	Rough Rice
	ZW	Wheat
Currency	A6	Australian Dollar
	AD	Canadian Dollar
	B6	British Pound
	BR	Brazilian Real
	DX	US Dollar Index
	E1	Swiss Franc
	E6	Euro FX
	J1	Japanese Yen
	RP	Euro/British Pound
	RU	Russian Ruble
Interest rate	BTP	Euro BTP Long-Bond
	ED	Eurodollar
	G	10-Year Long Gilt
	GG	Euro Bund
	HR	Euro Bobl
	US	30-Year US Treasury Bond
	ZF	5-Year US Treasury Note
	ZN	10-Year US Treasury Note
	ZQ	30-Day Fed Funds
	ZT	2-Year US Treasury Note
Stock market index	DY	DAX
	ES	E-mini S&P 500
	EW	E-mini S&P 500 Midcap
	FX	Euro Stoxx 50
	MME	MSCI Emerging Markets Index
	MX	CAC 40
	NQ	E-mini Nasdaq 100
	RTY	E-mini Russell 2000
	VX	VIX
	X	FTSE 100
	XAE	E-mini Energy Select Sector
	XAF	E-mini Financial Select Sector
	XAI	E-mini Industrial Select Sector
	YM	E-mini Dow