*Econometric Theory*, **0**, 2025, 1–45.
doi:10.1017/S0266466625100236

# ROBUST HIGH-DIMENSIONAL TIME-VARYING COEFFICIENT ESTIMATION

MINSEOK SHIN
*Pohang University of Science and Technology (POSTECH)*

DONGGYU KIM
*Department of Economics, University of California, Riverside*

In this article, we develop a novel high-dimensional coefficient estimation procedure based on high-frequency data. Unlike usual high-dimensional regression procedures such as LASSO, we additionally handle the heavy-tailedness of high-frequency observations as well as time variations of coefficient processes. Specifically, we employ the Huber loss and a truncation scheme to handle heavy-tailed observations, while $\ell_1$-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 $\ell_1$-regularization. Thus, when estimating integrated coefficients, we propose a debiasing scheme to enjoy the law of large numbers 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.

## 1. INTRODUCTION

With the wide availability of high-frequency financial data, researchers have developed financial models that can incorporate high-frequency data, and 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; Engle and Gallo, 2006; Corsi, 2009; Shephard and Sheppard, 2010; Hansen, Huang, and Shek, 2012; Kim and Wang, 2016; Kim and Fan, 2019; Song et al., 2021). Empirical studies have demonstrated that

The research of M.S. was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (RS-2025-24535699), and in part by the Institute of Information & Communications Technology Planning & Evaluation (IITP)-Global Data-X Leader HRD program grant funded by the Korean government (MSIT) (IITP-2025-RS-2024-00441244). The research of D.K. was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (RS-2024-00343129). Address correspondence to Donggyu Kim, University of California Riverside, Riverside, CA, USA, e-mail: donggyu.kim@ucr.edu.

capturing the auto-regressive structures of high-frequency measures helps explain financial market dynamics. On the other hand, we often employ realized volatility estimators when analyzing regression models, such as the Capital Asset Pricing Model (CAPM) (Sharpe, 1964; Lintner, 1965) 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), and Reiß, Todorov, and Tauchen (2015) for the related literature. Li, Todorov, and Tauchen (2017) derived the asymptotic efficiency bound for betas in a linear continuous-time regression model. In addition, empirical studies have shown the time-varying feature of the beta process (Ferson and Harvey, 1999; Ang and Kristensen, 2012; Reiß et al., 2015; Kong and Liu, 2018; Kalnina, 2022; Kong et al., 2023; Oh, Kim, and Wang, 2024). To address this issue, Aït-Sahalia, Kalnina, and Xiu (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 (Campbell, Hilscher, and Szilagyi, 2008; Bali, Cakici, and Whitelaw, 2011; Cochrane, 2011; Harvey, Liu, and Zhu, 2016; McLean and Pontiff, 2016; Hou, Xue, and Zhang, 2020). 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, Oh, and Shin (2025) 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 high-frequency financial data exhibit heavy tails (Cont, 2001; Fan and Kim, 2018; Mao and Zhang, 2018; Shin, Kim, and Fan, 2023). Under the heavy-tailedness assumption, the existing estimation methods, including the TED estimator (Kim et al., 2025), cannot consistently estimate the time-varying betas. Specifically, they fail to control the tail behavior of the local beta estimator and the bias adjustment term, which can lead to large estimation errors. 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.

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

In this article, 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 only a finite $(2+\zeta)$-th moment condition for an arbitrarily small $\zeta > 0$. 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 $\ell_1$-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 $\ell_1$-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 the article 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 $\mathbf{A} = \left(A_{ij}\right)$, 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 $\|\mathbf{A}\|_F = \sqrt{tr(\mathbf{A}^\top \mathbf{A})}$ 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 $\mathbf{X}(t) = \left(X_1(t), \ldots, X_p(t)\right)^\top$ be the dependent process and $p$-dimensional multivariate covariate process, respectively. We employ the

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

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 $\mathbf{X}^c(t) = \left(X_1^c(t), \ldots, X_p^c(t)\right)^\top$ 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, $\boldsymbol{\beta}(t) = \left(\beta_1(t), \ldots, \beta_p(t)\right)^\top$ 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 $\mathbf{X}(t)$, $\mathbf{J}(t) = \left(J_1(t), \ldots, J_p(t)\right)^\top$ 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. In this article, we do not assume that $\nu(t)$ is bounded. Instead, we only impose the finite moment condition on the residual process in Assumption 1(a), which allows the residual process to exhibit heavy tails. We assume that the coefficient $\boldsymbol{\beta}(t) = \left(\beta_1(t), \ldots, \beta_p(t)\right)^\top$ 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 an $r$-dimensional independent Brownian motion, and $\boldsymbol{\mu}_\beta(t)$ and $\boldsymbol{\nu}_\beta(t)$ are predictable. The main interest of this article is to investigate the latent regression diffusion process. From 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.

**Remark 1.** In this article, we focus on $\int_0^1 \boldsymbol{\beta}(t)dt$ rather than $\int_0^1 |\boldsymbol{\beta}(t)|dt$. Then, the variable selection is based on the average effect of the covariate process on the dependent process. For example, suppose that the beta process smoothly oscillates around zero. In this case, $\int_0^1 \boldsymbol{\beta}(t)\,dt$ would be close to zero, whereas

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

$\int_0^1 |\boldsymbol{\beta}(t)|\,dt$ can be large. Such factors with zero average effect are excluded in applications that emphasize the overall long-term effect. We note that the drift term $\boldsymbol{\mu}_\beta(t) = \big(\mu_{\beta,1}(t), \ldots, \mu_{\beta,p}(t)\big)^\top$ can play a significant role in the difference between $\int_0^1 \boldsymbol{\beta}(t)dt$ and $\int_0^1 |\boldsymbol{\beta}(t)|\,dt$. For example, suppose that $\beta_i(0) = 0$. When $\mu_{\beta,i}(t)$ changes sign over time, $\beta_i(t)$ may oscillate around zero, which can lead to a significant difference between the two measures. In contrast, when $\mu_{\beta,i}(t)$ maintains the same sign and $\beta_i(t)$ fluctuates around that level due to the stochastic Brownian motion component, the difference is relatively small. On the other hand, when the beta process exhibits discontinuities or nonsmoothness, $\int_0^1 |\boldsymbol{\beta}(t)|\,dt$ can serve as a more relevant measure. However, addressing such cases is beyond the scope of this article, and theoretically, the localization scheme does not work. Thus, we leave this issue for a future study.

In the regression-based financial models, there are hundreds of potential factor candidates (Campbell et al., 2008; Bali et al., 2011; Cochrane, 2011; Harvey et al., 2016; McLean and Pontiff, 2016; Hou et al., 2020). To account for this, we allow that 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:

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

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., $\delta = 0$. The exact sparsity condition implies that only several factors are significant, while most factors do not affect the dependent process. Thus, we assume that the relatively small number of factors is 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. Thus, even without the almost sure sparsity condition, the results in this article hold with high probability. However, for simplicity, we assume that the sparsity condition holds almost surely.

## 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) and (2.2). Recently, under the sub-Gaussian assumption, Kim et al. (2025) proposed the integrated beta estimator that can handle the curse of dimensionality and time-varying betas. 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), which leads to the inconsistency of the integrated beta estimator. To accommodate heavy-tailedness, we impose the finite moment condition on the

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

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

$$v_{j,n} = \sqrt{BV_j \log p} n^{-1/2}, \tag{3.1}$$

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 literature (Aït-Sahalia and Xiu, 2019; Aït-Sahalia et al., 2020) except for the $\log p$ term, which is used to bound the continuous parts of the covariate processes with high probability. 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 $Y(t)$.

**Remark 2.** In this article, we handle jumps using thresholding and robustification methods under the finite activity assumption. When extending this assumption to allow for jumps of infinite activity or infinite variation, a major challenge arises from the presence of numerous small jumps. To address this issue, we can apply truncation methods. Specifically, using truncation techniques, we can establish a CLT for the estimated volatility functionals in the presence of infinite activity but finite variation jumps (Mancini, 2009, 2017). For the infinite variation case, we can still obtain consistency, although the convergence rate is determined by the jump activity indices (Mancini, 2017). We note that these CLT and consistency results are established in the finite-dimensional setting. However, it is a theoretically demanding task to extend these results to the high-dimensional case. Thus, we leave this issue for future study.

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

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 procedures 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 $l_\tau$ (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 $l_\tau$ 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 $\ell_1$-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

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

In Theorem 1, we show that the proposed instantaneous beta estimator $\widehat{\boldsymbol{\beta}}_{i\Delta_n}$ is consistent with appropriate $\tau$ and $\eta$. 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 numbers property 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 a 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 $\ell_1$-minimization for inverse matrix estimation (CLIME) (Cai, Liu, and Luo, 2011):

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

where $\lambda$ 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:

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

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

The above adjustment is based on the debiasing scheme, which is widely employed in high-dimensional literature (Javanmard and Montanari, 2014, 2018; Van de Geer et al., 2014; Zhang and Zhang, 2014). This scheme reduces the bias from $\ell_1$-regularization by adding a bias-correction term to the original estimator, where the correction uses an approximate inverse volatility matrix of the covariates to re-weight the residuals. Specifically, it examines how much each parameter was shrunk by the $\ell_1$-penalty and then adjusts the estimates to remove that excess shrinkage. We note that this debiasing scheme performs well under the sub-Gaussian assumption (Javanmard and Montanari, 2014, 2018; Van de Geer et al., 2014; Kim et al., 2025). However, $\Delta_i^n Z^c$ has only finite $(2+\zeta)$-th moment for an arbitrarily small $\zeta > 0$; thus, the debiased instantaneous beta estimator has heavy-tails. To handle this issue, we employ the Winsorization method as follows. Define the truncation (Winsorization) function

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

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 $\varpi$ 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 numbers property when obtaining the integrated beta estimator. Usually, the debiasing scheme is employed to obtain 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 article, 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:

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

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

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 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 $C_\tau$, $C_\eta$, and $c_k$.

**Step 2** Obtain the inverse instantaneous volatility matrix estimator:

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

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

**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 $C_\varpi$.

**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:

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

where $s(\cdot)$ satisfies $|s(x) - x| \leq h_n$, $h_n = C_h b_n$ 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.

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

**Assumption 1.**

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

$$\max_{1\le i\le n} \mathbb{E}\left\{ |\sqrt{n}\Delta_i^n Z^c|^\gamma \,\middle|\, \mathcal{F}_{(i-1)\Delta_n} \right\} \le C,$$

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

where $\gamma = 2+\zeta$ for an arbitrarily small $\zeta > 0$.

(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 $\|\boldsymbol{\Sigma}^{-1}(t)\|_1 \le C$ a.s.

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

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

(d) $n^{c_1} \le p \le c_2 \exp(n^{c_3})$ 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 $\mathcal{W}_t = \left\{ \mathbf{w} \in \mathbb{R}^p : \left\|\mathbf{w}_{S_t^c}\right\|_1 \le 3\left\|\mathbf{w}_{S_t}\right\|_1 + 4\left\|(\boldsymbol{\beta}_0(t))_{S_t^c}\right\|_1 \right\}$, where $\mathbf{w}_{S_t^c}$ is the subvector obtained by stacking $\left\{\mathbf{w}_j : j \in S_t^c\right\}$, $\mathbf{w}_{S_t}$ is the subvector obtained by stacking $\left\{\mathbf{w}_j : j \in S_t\right\}$, $(\boldsymbol{\beta}_0(t))_{S_t^c}$ is the subvector obtained by stacking $\left\{(\boldsymbol{\beta}_0(t))_j : j \in S_t^c\right\}$, and $S_t = \{j : \text{ jth element of } |\boldsymbol{\beta}_0(t)| > n\eta\}$. Then, there exists a positive constant $\kappa$ such that the following inequality holds for some $D = (8 + 48/\kappa)s_p(n\eta)^{1-\delta}$ and $0 \le i \le n - k_n$, where the specific value of $\eta$ is given in Theorem 1:

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

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

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

**Remark 3.** Assumption 1(a) is the finite $(2+\zeta)$-th moment condition for an arbitrarily small $\zeta > 0$, which allows the dependent process $Y(t)$, covariate process $\mathbf{X}(t)$, and residual process $Z^c(t)$ to have heavy-tailed distributions. We note that in financial applications, the finite second moment condition is not restrictive (Cont, 2001; Gabaix et al., 2003). We also note that the moment condition for $Z^c(t)$ is satisfied when $\Delta_i^n Z^c$ is an independent random variable and $\mathbb{E}\{|\Delta_i^n Z^c|^\gamma\} \le Cn^{-\gamma/2}$, or $\sup_{0\le t\le 1}\sup_{t\le s\le 1}\mathbb{E}\{|\nu(s)|^\gamma \,|\, \mathcal{F}_t\} \le C$ a.s. 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 local boundedness condition by Lemma 4.4.9 in Jacod and Protter (2011). Specifically, if the asymptotic result, such as stable convergence

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

in law or convergence in probability, is satisfied under the boundedness condition, it is also satisfied under the local boundedness condition. We note that the local boundedness condition is usually satisfied in financial data. 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. The sparsity assumptions are well established in financial modeling, where only a small number of factors explain the asset returns. In Assumption 1(d), we allow the dimension $p$ to grow with the number of observations $n$ exponentially, which accommodates high-dimensional data settings. 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, Zhou, and Fan, 2020). This implies strictly positive restricted eigenvalues (REs) over a local neighborhood. In high-dimensional applications, global RE conditions are often restrictive due to strong correlations among covariates. By focusing on local neighborhoods around the true parameter, the LRE condition provides a more realistic assumption for real data. We note that $n\eta$ converges to zero for the choice of $\eta$ in Theorems 1 and 2. When the coefficient process $\boldsymbol{\beta}(t)$ satisfies the exact sparsity condition, i.e., $\delta = 0$, $\mathcal{W}_t$ is replaced by an $\ell_1$-cone $\{\mathbf{w}\in\mathbb{R}^p : \|\mathbf{w}_{S_t^c}\|_1 \leq 3\|\mathbf{w}_{S_t}\|_1\}$, 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 $\mathbf{\Sigma}(t)$ follows a continuous Itô diffusion process with bounded drift and instantaneous volatility processes. We also note that this condition is generally reasonable in financial markets, except the cases of volatility spikes. However, such spikes can be separately modeled as jumps and handled under the finite activity assumption.

The following theorem derives the asymptotic properties of the 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 $\eta = C_{\eta,a}[s_p n^{-3/2}\sqrt{k_n \log p} + n^{-1}k_n^{-1/2}(\log p)^{3/4}]$ for some large constants $C_{\tau,a}$ and $C_{\eta,a}$. Then, we have, for large $n$,*

$$\max_i \|\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0(i\Delta_n)\|_1 \leq C s_p (n\eta)^{1-\delta} \quad \text{and}$$

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

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

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

**Remark 4.** Theorem 1 shows the $\ell_1$ and $\ell_2$ norm error bounds of the instantaneous beta estimator. We note that as $k_n$ increases, the statistical estimation error decreases, and the 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 $\ell_1$ convergence rate of $n^{-(1-\delta)/4}$ and $\ell_2$ 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 numbers 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 $C_\varpi$. Then, we have, with probability greater than $1 - p^{-a}$,*

$$\|\widehat{I\beta} - I\beta_0\|_{\max} \le C b_n, \tag{3.8}$$

*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 5.** 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 convergence 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 $\ell_1$ convergence rate of the RED-LASSO estimator.

THEOREM 3. *Under the assumptions in Theorem 2, for any given positive constant a, choose $h_n = C_{h,a} b_n$ for some constant $C_{h,a}$. Then, we have, with probability greater than $1 - p^{-a}$,*

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

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

Theorem 3 shows that the proposed RED-LASSO estimator is consistent in terms of the $\ell_1$ norm. We note that under the sub-Gaussian assumption on the log-return data, Kim et al. (2025) proposed the integrated beta estimator that has the $\ell_1$ 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. We also note that the regularized approximate quadratic (RA-Lasso) estimator (Fan, Li, and Wang, 2017) and the regularized adaptive Huber estimator (Sun et al., 2020) are robust to heavy-tailed distributions. Under the constancy of the beta process, these estimators achieve the $\ell_1$ convergence rate of $s_p n^{-1/2} \sqrt{\log p}$. In contrast, the proposed RED-LASSO estimator accommodates both heavy-tailedness and time-varying beta processes, with an additional cost of at most $\log p$ and $s_p$ orders.

### 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 $k_n = [n^{1/2}]$ for the simulation study. The choice of $k_n$ for the empirical study is presented in Section 5. Also, we choose

$$\begin{aligned} &\tau = c_\tau n^{-1/4} (\log p)^{-3/4}, \quad \eta = c_\eta n^{-5/4} (\log p)^{3/4}, \\ &\lambda = c_\lambda n^{-1/4} \sqrt{\log p}, \quad \varpi = c_\varpi (\log p)^{1/4}, \quad \text{and} \quad h_n = c_h n^{-1/2} (\log p)^{3/2}, \end{aligned} \tag{3.10}$$

where $c_\tau, c_\eta, c_\lambda, c_\varpi$, and $c_h$ are tuning parameters. For the simulation and empirical studies, we choose $c_\varpi$ and $c_h$ that minimize the corresponding mean squared prediction error (MSPE). The results are $c_\varpi = 0.025$, and $c_h = 0.1$. Details can be found in Section 5. Also, we select $c_\tau, c_\eta \in [0.1, 10]$ via five-fold cross-validation based on the following mean squared error (MSE):

$$\frac{1}{k_n^{\text{test}}} \left\| \mathcal{Y}_i^{\text{test}} - \left( \mathcal{X}_i^{\text{test}} \right) \widehat{\boldsymbol{\beta}}_{i\Delta_n}^{\text{train}} \right\|_2^2,$$

where $\mathcal{Y}_i^{\text{test}}$ and $\mathcal{X}_i^{\text{test}}$ are obtained from $\mathcal{Y}_i$ and $\mathcal{X}_i$, respectively, by selecting the rows corresponding to the test data, $k_n^{\text{test}}$ is the number of rows in $\mathcal{Y}_i^{\text{test}}$ and $\mathcal{X}_i^{\text{test}}$, and $\widehat{\boldsymbol{\beta}}_{i\Delta_n}^{\text{train}}$ is the instantaneous beta estimator calculated from the training data. The test and training data are chosen based on the five-fold cross-validation. This choice procedure is similar to that of Sun et al. (2020) and Tan, Sun, and Witten (2023), which employ robust regularization approaches. Similarly, we choose $c_\lambda \in [0.1, 10]$ via five-fold cross-validation based on the following loss function

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

(Cai et al., 2011):

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

where $\widehat{\mathbf{\Omega}}_{i\Delta_n}^{\mathrm{train}}$ is the inverse instantaneous volatility matrix estimator obtained from the training data and $\mathbf{I}_p$ is the $p$-dimensional identity matrix. We note that $\widehat{\boldsymbol{\beta}}_{i\Delta_n}^{\mathrm{train}}$ and $\widehat{\mathbf{\Omega}}_{i\Delta_n}^{\mathrm{train}}$ enjoy the same theoretical properties as $\widehat{\boldsymbol{\beta}}_{i\Delta_n}$ and $\widehat{\mathbf{\Omega}}_{i\Delta_n}$, which can be shown similar to the proofs of Theorems 1 and 2. This holds because the time gaps used in each local estimation are sufficiently short, which allows us to control the errors introduced by subsampling.

## 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) and (2.2), we generated the data using the heavy-tailed and sub-Gaussian processes with frequency $1/n^{all}$. Specifically, we employed the following time-series regression jump-diffusion model:

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

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 $\mathbf{\Lambda}(t) = \left(\Lambda_1(t), \ldots, \Lambda_p(t)\right)^\top$ 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 $\infty$ 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 $v(t)$ as follows:

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

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

$$dv'(t) = 3\left(0.8 - v'(t)\right)dt + 0.24d\mathbf{W}^v(t),$$

where $v'(0) = 1$ and $\mathbf{W}^v(t)$ is an independent Brownian motion. We note that the process $v(t)$ is not realistic. However, to investigate the effect of the heavy-tailedness of the return process, the structure of $v(t)$ is imposed. To generate the volatility process $\boldsymbol{\sigma}(t)$, we first generated the Ornstein–Uhlenbeck process $u(t)$ as follows:

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

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

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 $\boldsymbol{\Sigma}(t) = \left(\Sigma_{ij}(t)\right)_{1\le i,j\le p}$, where $\Sigma_{ij}(t) = u(t)0.6^{|i-j|}$. To generate the coefficient $\boldsymbol{\beta}(t)$, we considered the exact sparse process, i.e., $\beta_i(t) = 0$ for $[s_p]+1 \le i \le p$. 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 $\boldsymbol{\mu}_\beta(t) = \left(\mu_{\beta,1}(t), \ldots, \mu_{\beta,p}(t)\right)^\top$, $\boldsymbol{\nu}_\beta(t) = \left(\nu_{\beta,i,j}(t)\right)_{1\le i,j\le p}$, and $\mathbf{W}_\beta(t)$ is a $p$-dimensional independent Brownian motion. For $1 \le i \le [s_p]$, the initial value $\beta_i(0) = 1$ and $\mu_{\beta,i}(t) = 0.1$ for $0 \le t \le 1$. The process $\left(\nu_{\beta,i,j}(t)\right)_{1\le i,j\le [s_p]}$ 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:

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

where $\xi(0) = 0.15$ and $\mathbf{W}^\xi(t)$ is an independent Brownian motion. We chose $p = 100$, $s_p = \log p$, and $n^{all} = 4,000$, and we varied $n$ from 1,000 to 4,000. 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

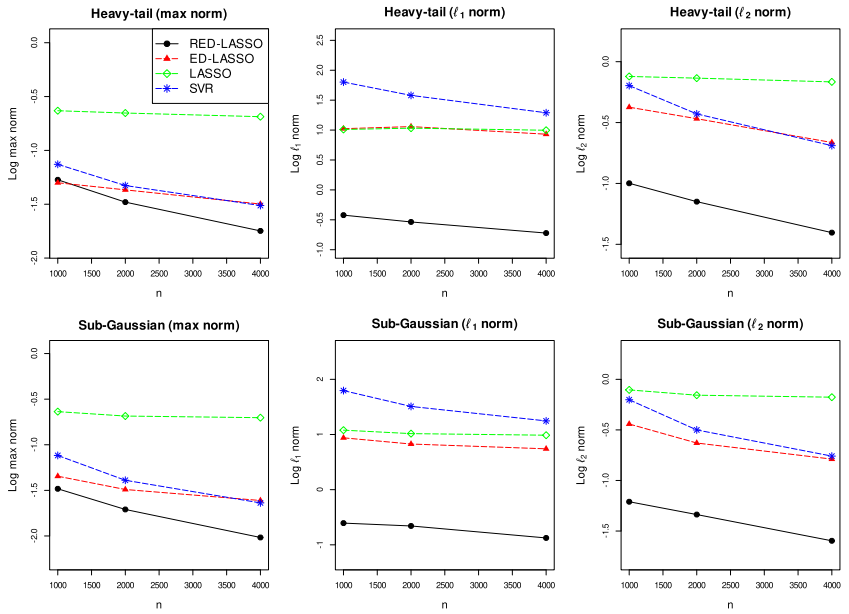
$$\mathcal{Y}_i' = \begin{pmatrix} \Delta_{i+1}^n \widehat{Y}^c \\ \Delta_{i+2}^n \widehat{Y}^c \\ \vdots \\ \Delta_{i+k_n}^n \widehat{Y}^c \end{pmatrix} \quad \text{and} \quad \Delta_i^n \widehat{Y}^c = \Delta_i^n Y \mathbf{1}_{\{|\Delta_i^n Y| \le u_n\}}. \tag{4.1}$$

In the simulation and empirical studies, we choose $u_n = \sqrt{BV^Y \log p}\, n^{-1/2}$, where the bipower variation $BV^Y = \frac{\pi}{2}\sum_{i=2}^n |\Delta_{i-1}^n Y| \cdot |\Delta_i^n Y|$. 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 via five-fold cross-validation based on the MSE. We also employed the support vector

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press




The log max, $\ell_1$, and $\ell_2$ norm error plots (corresponding to three columns) of the RED-LASSO (black dot), ED-LASSO (red triangle), LASSO (green diamond), and SVR (blue star) estimators for $p = 100$ and $n = 1,000, 2,000, 4,000$.

regression (SVR) estimator with a linear kernel as follows:

$$\widetilde{I\beta}^{\mathrm{SVR}} = \mathrm{argmin}_{\boldsymbol{\beta}} \left\{ C_s \sum_{i=1}^{n} \max\left\{ |\Delta_i^n \widehat{Y}^c - \Delta_i^n \widehat{\mathbf{X}}^{c\top} \boldsymbol{\beta}| - \epsilon, 0 \right\} + \frac{1}{2} \|\boldsymbol{\beta}\|_2^2 \right\}, \tag{4.3}$$

where the cost parameter $C_s \in [10^{-4}, 1]$ and insensitivity parameter $\epsilon \in [10^{-4}, 1]$ were selected by five-fold cross-validation using the MSE. We note that the SVR estimator can partially mitigate the effect of heavy-tailed distributions by employing the above epsilon-insensitive loss function instead of a squared loss function. However, it cannot fully account for heavy-tails, nor can it handle the time-varying property of the beta process. After obtaining the RED-LASSO, ED-LASSO, LASSO, and SVR estimators, the average estimation errors under the max norm, $\ell_1$ norm, and $\ell_2$ norm were computed by 1,000 simulations.

Figure 1 plots the log max, $\ell_1$, and $\ell_2$ norm errors of the RED-LASSO, ED-LASSO, LASSO, and SVR estimators with $n = 1,000, 2,000, 4,000$ for the heavy-tailed 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-tailed process. This may be because the RED-LASSO estimator can fully explain the heavy-tailedness, while other estimators cannot. For the sub-Gaussian process,

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

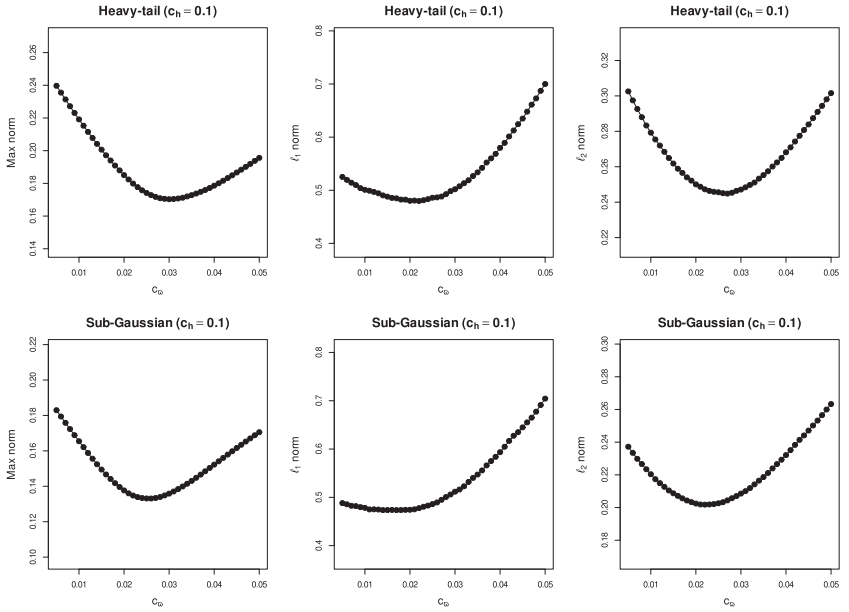



**FIGURE 2.** The max, $\ell_1$, and $\ell_2$ norm error plots of the RED-LASSO estimator for $p = 100$ and $n = 4,000$ with the heavy-tailed and sub-Gaussian processes. Note that $c_h = 0.1$ is fixed and $c_\varpi$ is varied from 0.005 to 0.05.

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. We also note that the ED-LASSO estimator does not outperform the SVR estimator for both heavy-tailed and sub-Gaussian processes. This may be because, although the SVR estimator cannot account for the time-varying property of the beta process, it handles heavy-tailed distributions better than the ED-LASSO estimator.

On the other hand, we set the tuning parameters $c_\varpi = 0.025$ and $c_h = 0.1$ in the numerical study. To investigate the robustness of the RED-LASSO estimator with respect to the choice of $c_\varpi$ and $c_h$, we calculated the estimation errors by varying $c_\varpi$ and $c_h$. Figures 2 and 3 show the max, $\ell_1$, and $\ell_2$ norm errors of the RED-LASSO estimator with $n = 4,000$ for the heavy-tailed and sub-Gaussian processes. In Figure 2, $c_h = 0.1$ is fixed and $c_\varpi$ is varied from 0.005 to 0.05, while $c_\varpi = 0.025$ is fixed and $c_h$ is varied from 0.05 to 0.5 in Figure 3. As seen in Figures 2 and 3, the

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

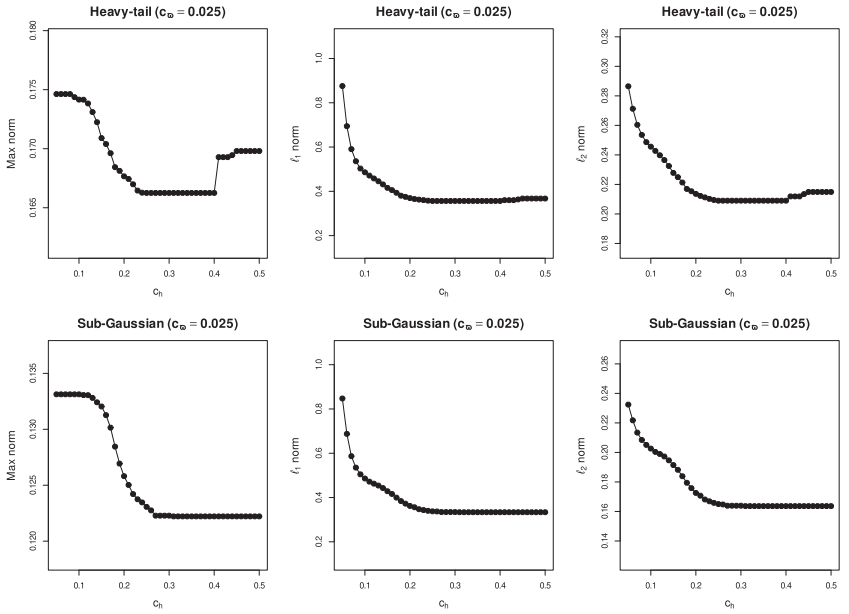



**FIGURE 3.** The max, $\ell_1$, and $\ell_2$ norm error plots of the RED-LASSO estimator for $p = 100$ and $n = 4,000$ with the heavy-tailed and sub-Gaussian processes. Note that $c_\varpi = 0.025$ is fixed and $c_h$ is varied from 0.05 to 0.5.

errors of the RED-LASSO estimator usually do not critically depend on the choice of $c_\varpi$ and $c_h$. This result supports the practical applicability of the RED-LASSO procedure. An exception is the $\ell_1$ norm error for small $c_h$. This may be because the RED-LASSO estimator with small $c_h$ cannot effectively handle the non-sparse structure of the debiased integrated beta estimator.

## 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 used 5-minute data to mitigate the effect of microstructure noise. Specifically, we obtained 5-minute log-price data with the previous tick scheme (Wang and Zou, 2010; Zhang, 2011) and processed the data similarly to the procedure in Kim et al. (2025). 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), Amazon.com, Inc. (AMZN), Alphabet Inc. (GOOG), and Exxon Mobil

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

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 futures data, 10 currency futures data, 10 interest rate futures data, and 14 stock market index futures data. The specific list is presented in Table A1 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 stock market analysis. We note that 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:

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

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-minute 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 $d$th day and $i$th time interval, $N_d$ is the number of portfolio components on the $d$th day, the superscript $j$ is used to represent the $j$th 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 $j$th 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 et al. (2025).

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

When obtaining the RED-LASSO estimator, we selected the tuning parameters based on Sections 3.3 and 4. Also, we set $k_n = 78$ so that the instantaneous betas were estimated on a daily basis, while the integrated betas were estimated on a monthly basis. Finally, we chose the tuning parameters $c_\varpi$ and $c_h$ using the cross-validation scheme on multi-period high-frequency financial data in 2013. Since the integrated beta estimator is obtained at the monthly level and the stationarity assumption is reasonable for the beta process, we employed one-month-ahead prediction error to evaluate performance for each tuning parameter selection (Wang and Zou, 2010). Specifically, we first defined the following MSPE:

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

where $\widetilde{I\beta}^{j,s}(c_\varpi, c_h)$ is the RED-LASSO estimator with the tuning parameters $c_\varpi$ and $c_h$ for the $j$th month in 2013 and $s$th stock. We note that $\widetilde{I\beta}^{j,s}(\infty,\infty)$ is the RED-LASSO estimator obtained without truncation or thresholding. Then, we selected $c_\varpi$ and $c_h$ by minimizing $\Lambda(c_\varpi, c_h)$ over $c_\varpi \in \{0.005l \mid 1 \le l \le 10, l \in \mathbb{Z}\}$ and $c_h \in \{0.05l \mid 1 \le l \le 10, l \in \mathbb{Z}\}$. The results are $c_\varpi = 0.025$ and $c_h = 0.1$. Then, using the RED-LASSO, ED-LASSO, LASSO, and SVR estimation procedures, we obtained the monthly integrated betas for each of the five assets. For the non-trading period, we set the beta estimates as zero.

We first compare the performances of the RED-LASSO, ED-LASSO, LASSO, and SVR estimators. To do this, we calculated the monthly in-sample and out-of-sample $R^2$ with the monthly integrated beta estimates. We note that since the integrated beta reflects the average effect of covariate movements on the dependent process, the higher in-sample $R^2$ implies a better approximation of this average effect. This indicates how well the estimator captures the overall time-varying relationship between the dependent and covariate processes. That is, $R^2$ measures the goodness-of-fit for the proposed time-varying linear model. However, the in-sample $R^2$ can cause the overfitting issue. To overcome this, we use the out-of-sample $R^2$. Theoretically, under a stationarity condition of the beta process, the best predictor of the beta process is the one from the previous period. From this point of view, the out-of-sample $R^2$ indicates the performance of explaining this average relationship over future data. Specifically, from the high out-of-sample $R^2$, we can conjecture that the proposed method can account for the dynamics of the beta process. 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 12 months. Table 1 shows the average in-sample and out-of-sample $R^2$ of the RED-LASSO, ED-LASSO, LASSO, and SVR 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.

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

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

| | In-sample $R^2$ | | | |
|---|---|---|---|---|
| | Estimator | | | |
| | RED-LASSO | ED-LASSO | LASSO | SVR |
| Whole period | 0.233 | 0.228 | 0.197 | 0.226 |
| 2013 | 0.229 | 0.222 | 0.195 | 0.225 |
| 2014 | 0.202 | 0.195 | 0.165 | 0.192 |
| 2015 | 0.246 | 0.240 | 0.211 | 0.238 |
| 2016 | 0.223 | 0.221 | 0.190 | 0.223 |
| 2017 | 0.180 | 0.180 | 0.151 | 0.177 |
| 2018 | 0.320 | 0.313 | 0.280 | 0.307 |
| 2019 | 0.228 | 0.225 | 0.191 | 0.223 |
| | Out-of-sample $R^2$ | | | |
| | Estimator | | | |
| | RED-LASSO | ED-LASSO | LASSO | SVR |
| Whole period | 0.215 | 0.206 | 0.190 | 0.199 |
| 2014 | 0.183 | 0.170 | 0.157 | 0.151 |
| 2015 | 0.229 | 0.217 | 0.205 | 0.214 |
| 2016 | 0.206 | 0.198 | 0.183 | 0.182 |
| 2017 | 0.165 | 0.158 | 0.143 | 0.160 |
| 2018 | 0.299 | 0.290 | 0.267 | 0.287 |
| 2019 | 0.209 | 0.204 | 0.186 | 0.202 |

Table 2 reports the monthly average proportion of non-zero integrated beta estimates across six factor groups and five assets over 84 months for the RED-LASSO, ED-LASSO, LASSO, and SVR estimators. Detailed non-zero frequencies for each individual factor are presented in Table A2 in the Appendix. As seen in Table 2, the RED-LASSO estimator can better account for the sparsity of the integrated betas than the ED-LASSO, LASSO, and SVR estimators. From this result, we can conjecture that the proposed RED-LASSO provides more sparse beta estimates, which is an important property in practice. Furthermore, as discussed above, the RED-LASSO estimator shows 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 (Fama and French, 1992, 2015; Carhart, 1997; Asness, Moskowitz, and Pedersen, 2013) since the market factors can be partially explained by the stock market index futures factors.

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

**TABLE 2.** The average proportion of non-zero monthly integrated beta estimates across factor groups and assets for the RED-LASSO (RED), ED-LASSO (ED), LASSO, and SVR estimators

| | AAPL | | | | BRK.B | | | | AMZN | | | | GOOG | | | | XOM | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Type | RED | ED | LASSO | SVR | RED | ED | LASSO | SVR | RED | ED | LASSO | SVR | RED | ED | LASSO | SVR | RED | ED | LASSO | SVR |
| All factors | 0.18 | 0.42 | 0.30 | 0.95 | 0.24 | 0.44 | 0.50 | 0.95 | 0.21 | 0.44 | 0.39 | 0.95 | 0.21 | 0.43 | 0.35 | 0.95 | 0.25 | 0.47 | 0.59 | 0.95 |
| Commodity | 0.05 | 0.33 | 0.12 | 0.96 | 0.05 | 0.32 | 0.31 | 0.96 | 0.05 | 0.34 | 0.21 | 0.96 | 0.05 | 0.33 | 0.19 | 0.96 | 0.14 | 0.39 | 0.46 | 0.96 |
| Currency | 0.08 | 0.39 | 0.18 | 0.97 | 0.10 | 0.40 | 0.40 | 0.97 | 0.11 | 0.41 | 0.30 | 0.97 | 0.12 | 0.39 | 0.23 | 0.97 | 0.12 | 0.41 | 0.49 | 0.97 |
| Interest rate | 0.06 | 0.29 | 0.15 | 0.84 | 0.08 | 0.29 | 0.38 | 0.84 | 0.07 | 0.29 | 0.25 | 0.84 | 0.07 | 0.31 | 0.20 | 0.84 | 0.06 | 0.30 | 0.38 | 0.84 |
| Stock market index | 0.50 | 0.62 | 0.70 | 0.99 | 0.58 | 0.63 | 0.85 | 0.99 | 0.52 | 0.62 | 0.74 | 0.99 | 0.53 | 0.63 | 0.74 | 0.99 | 0.42 | 0.60 | 0.86 | 0.99 |
| Market factor | 0.25 | 0.55 | 0.37 | 1.00 | 0.61 | 0.78 | 0.71 | 1.00 | 0.39 | 0.61 | 0.52 | 1.00 | 0.37 | 0.60 | 0.46 | 1.00 | 0.79 | 0.87 | 0.86 | 1.00 |

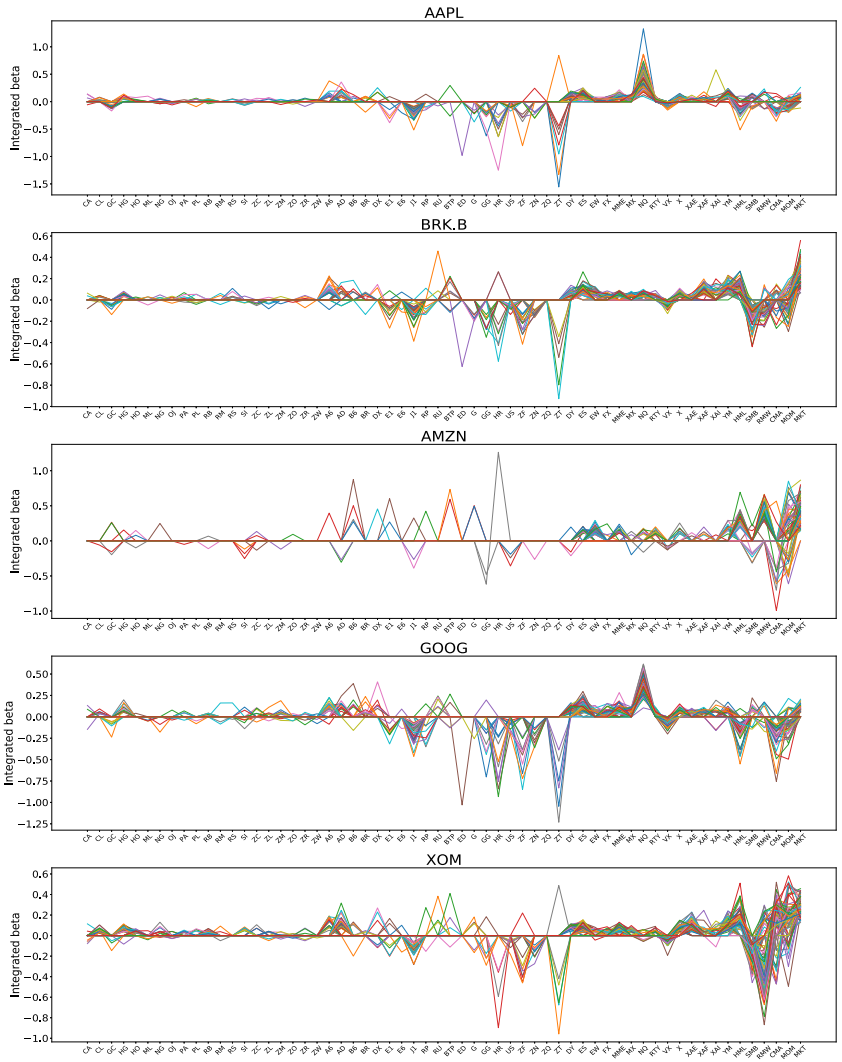



**FIGURE 4.** The monthly integrated betas from the RED-LASSO estimator for the five assets and 60 factors. Each line connects the 60 integrated beta estimates for each month.

Now, we investigate the result of the RED-LASSO estimator. Figure 4 shows the monthly integrated betas from the RED-LASSO estimator for the five assets and 60 factors. Figure 5 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 4 and 5, we see that integrated betas change over time,

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

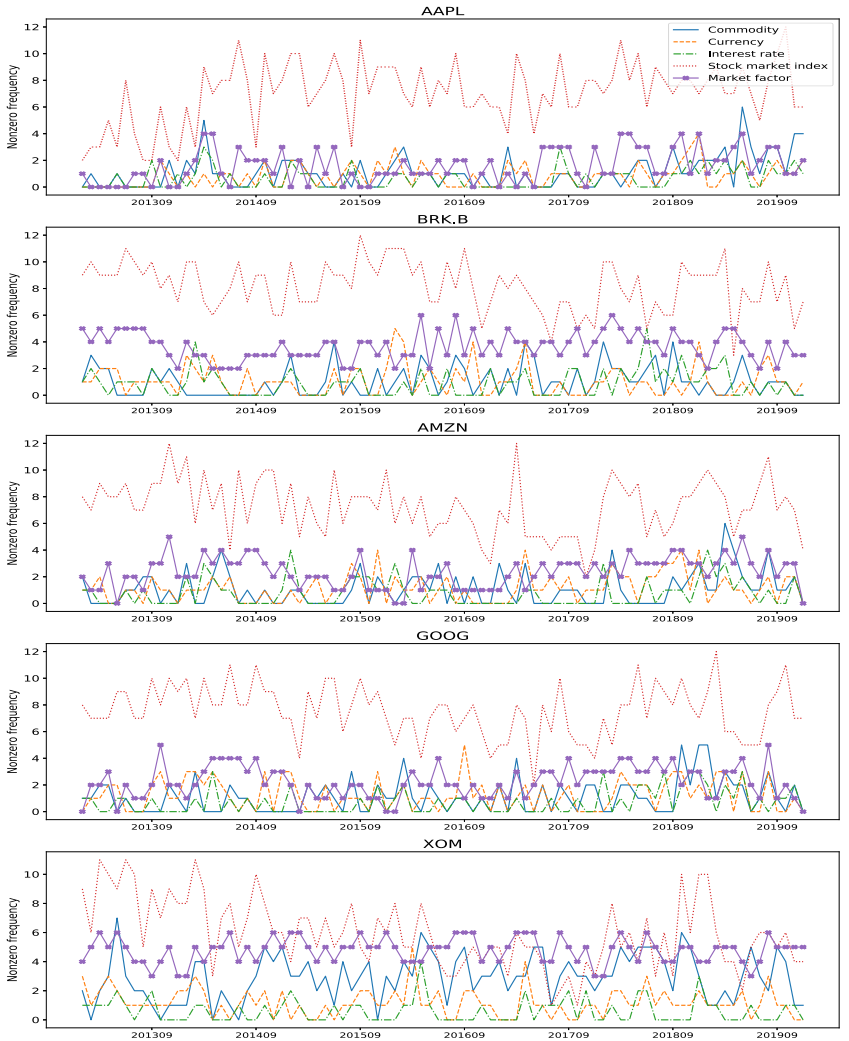



**FIGURE 5.** The non-zero frequency of the monthly integrated betas from the RED-LASSO estimator for the five assets and five groups. The five groups consist of the commodity futures group, currency futures group, interest rate futures group, stock market index futures group, and market factor group.

and only a small number of factors had non-zero integrated betas in most periods. To investigate the 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 6. The AAPL has NQ (E-mini Nasdaq 100), ES (E-mini S&P 500), and YM (E-mini Dow); BRK.B has MKT, YM, and ES; AMZN has MKT, MOM, and RMW; GOOG has NQ, ES, and VX (VIX); and XOM has MKT, XAE (E-

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press


**Figure 6.** 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.

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

mini Energy Select Sector), and YM. 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 (Carhart, 1997; Asness et al., 2013; Fama and French, 2015, 2016). To investigate their beta behaviors in more detail, we present the box plots of the integrated betas from the RED-LASSO and ED-LASSO estimators for these six factors in Figure 7. As expected, the MKT factor played a significant role for BRK.B and XOM. Specifically, the MKT factor had non-zero integrated betas for all 84 months as shown in Table A2 in the Appendix. In contrast, the six factors frequently had zero integrated betas for AAPL, AMZN, and GOOG. This may be because technology companies, such as AAPL, AMZN, and GOOG, have recently shown outstanding performance in the U.S. market, potentially reducing their dependence on these six factors. We note that the results of the two estimators are similar, but the RED-LASSO estimator has a more stable result. Thus, we can conjecture that considering both the heavy-tailed distribution and the time variation of the beta process helps better explain the beta dynamics.

## 6. CONCLUSION

In this article, 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 $\ell_1$-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 $\ell_1$-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 numbers 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, the proposed tuning parameter selection procedure does not guarantee the theoretical properties. It would be interesting to develop a robust tuning parameter selection method with rigorous theoretical guarantees and practical applicability. However, developing a procedure that satisfies both theoretical and

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

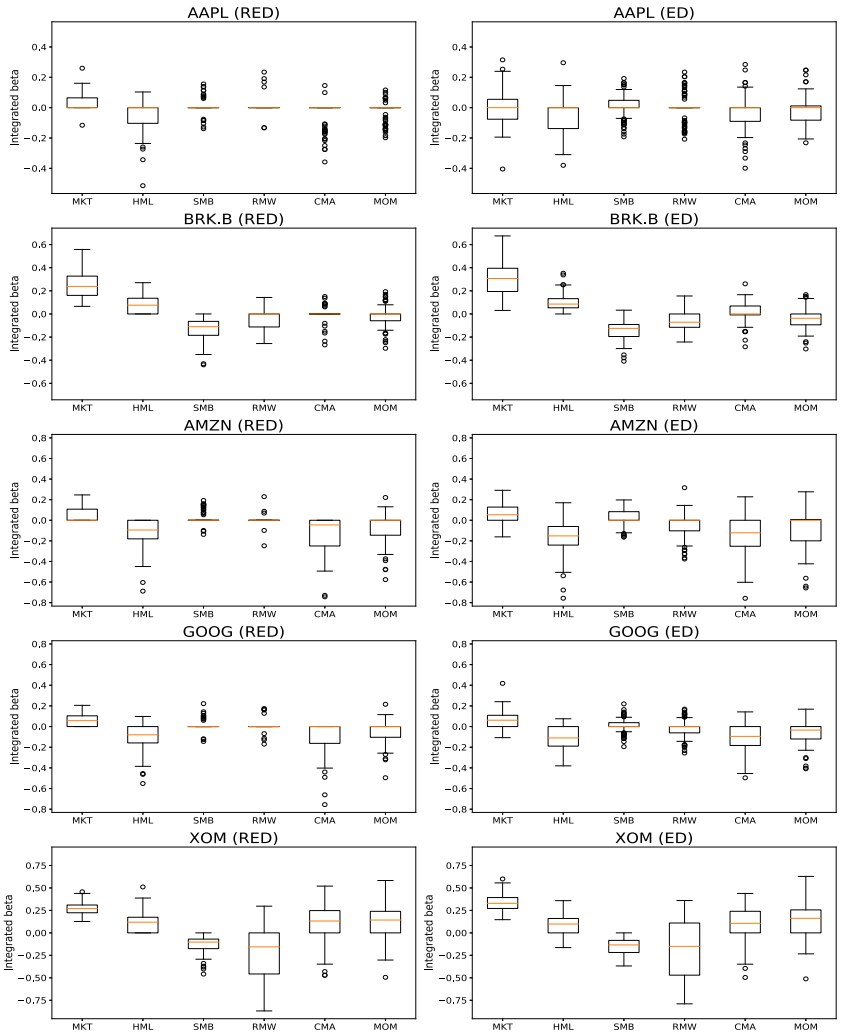



**FIGURE 7.** The box plots of the monthly integrated betas from the RED-LASSO and ED-LASSO estimators for MKT, HML, SMB, RMW, CMA, and MOM over 84 months for each of the five assets.

practical criteria may be challenging. In addition, in this article, 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

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

important to develop a robust estimation method that can handle microstructure noises. Finally, for each local sparse beta estimation, thanks to the bias adjustment, the bias is negligible, and we can obtain an asymptotic normality. However, the parameter of interest is the integrated beta, and the integrated beta estimator has a faster convergence rate than that of the local beta estimator, due to the property of the law of large numbers. Unfortunately, the convergence rate of the integrated beta estimator can be dominated by that of the bias term. Thus, we still have a bias issue for the integrated beta inference. It would be interesting but difficult to develop a novel bias adjustment scheme to manage this non-negligible bias term. We leave these topics for future study.

## 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 A1. *Under the assumptions in Theorem 1, we have*

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

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

**Proof of Proposition A1.** Define

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

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 $1 \le j \le p$, we have

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

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

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

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

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

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

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

Consider the first term. Similar to the proofs of Theorem 1 (Kim et al., 2024), we can show, for any constant $b \geq 1$,

$$\Pr\left\{ \sup_{1\leq h\leq k_n} \mathbb{E}\left\{ |\widetilde{\mathcal{X}}_{ih}|^b \middle| \mathcal{F}_{(i+h-1)\Delta_n} \right\} \leq \left( Cs_p \Delta_n \sqrt{bk_n \log p} \right)^b \right\} \geq 1 - p^{-2-a} \quad \text{and}$$

$$\sup_{1\leq h\leq k_n} \mathbb{E}\left\{ |\Delta^n_{i+h} X^c_j|^b \middle| \mathcal{F}_{(i+h-1)\Delta_n} \right\} \leq (C\Delta_n b)^{b/2} \text{ a.s.} \tag{A.6}$$

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

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

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

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

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

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

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

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

Then, using the fact that

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

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

$$\begin{aligned}
&\sup_{1\leq h\leq k_n} \left| \mathbb{E}\left\{ f(h) - \left( \mathcal{Z}_{ih} + \widetilde{\mathcal{X}}_{ih} \right) \Delta^n_{i+h} X^c_j \mathbf{1}\left( |\Delta^n_{i+h} X^c_j| \leq C_X \sqrt{\log p} n^{-1/2} \right) \middle| \mathcal{F}_{(i+h-1)\Delta_n} \right\} \right| \\
&\leq \tau^{-1} \sup_{1\leq h\leq k_n} \mathbb{E}\left\{ |\mathcal{Z}_{ih} + \widetilde{\mathcal{X}}_{ih}|^2 |\Delta^n_{i+h} X^c_j| \middle| \mathcal{F}_{(i+h-1)\Delta_n} \right\} \\
&\leq \tau^{-1} \sup_{1\leq h\leq k_n} \\
&\quad \left( \mathbb{E}\left\{ |\mathcal{Z}_{ih} + \widetilde{\mathcal{X}}_{ih}|^\gamma \middle| \mathcal{F}_{(i+h-1)\Delta_n} \right\} \right)^{2/\gamma} \left( \mathbb{E}\left\{ |\Delta^n_{i+h} X^c_j|^{\gamma/(\gamma-2)} \middle| \mathcal{F}_{(i+h-1)\Delta_n} \right\} \right)^{(\gamma-2)/\gamma} \\
&\leq C\tau^{-1} n^{-3/2},
\end{aligned} \tag{A.9}$$

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

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

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

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

$$\begin{aligned}&\Pr\left\{(I)_j^{(3)} \ge s \text{ and } \sum_{h=1}^{k_n}\mathbb{E}\left\{(f(h))^2\middle|\mathcal{F}_{(i+h-1)\Delta_n}\right\} \le Cn^{-2}k_n\right\}\\ &\le 2\exp\left\{-Ck_n^2s^2\Big/\left(\tau\sqrt{\log p}n^{-1/2}k_ns+n^{-2}k_n\right)\right\}.\end{aligned} \tag{A.11}$$

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

$$\begin{aligned}&\sup_{1\le h\le k_n}\left|\mathbb{E}\left\{(f(h))^2\middle|\mathcal{F}_{(i+h-1)\Delta_n}\right\}\right|\\ &\le \sup_{1\le h\le k_n}\mathbb{E}\left\{|\mathcal{Z}_{ih}+\widetilde{\mathcal{X}}_{ih}|^2|\Delta_{i+h}^nX_j^c|^2\middle|\mathcal{F}_{(i+h-1)\Delta_n}\right\}\\ &\le \left(\sup_{1\le h\le k_n}\mathbb{E}\left\{|\mathcal{Z}_{ih}+\widetilde{\mathcal{X}}_{ih}|^\gamma\middle|\mathcal{F}_{(i+h-1)\Delta_n}\right\}\right)^{2/\gamma}\\ &\quad\times\left(\sup_{1\le h\le k_n}\mathbb{E}\left\{|\Delta_{i+h}^nX_j^c|^{2\gamma/(\gamma-2)}\middle|\mathcal{F}_{(i+h-1)\Delta_n}\right\}\right)^{(\gamma-2)/\gamma}\\ &\le Cn^{-2},\end{aligned}$$

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

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

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

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

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

$$\begin{aligned}&Q_1=\{\max_{i,j}|\Delta_i^nX_j^c|\le C\sqrt{\log p}n^{-1/2}\},\\ &Q_2=\{\max_{i,j}\int_{i\Delta_n}^{(i+k_n)\Delta_n}d\Lambda_j(t)\le C\log p\}\cap\{\max_i\int_{i\Delta_n}^{(i+k_n)\Delta_n}d\Lambda^y(t)\le C\log p\},\\ &Q_3=\{\max_{i,j}\sum_{k=1}^{k_n}\mathbf{1}_{\{|\Delta_{i+k}^nX_j|>v_{j,n}\}}\le C\log p\}.\end{aligned}$$

By (A.3), we have

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

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

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

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

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

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

which implies

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

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

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

□

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

**Proof of Theorem 1** By Proposition A1, it is enough to show the statement under (A.1). First, we investigate $\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0$. Since

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

we have

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

Then, we have

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

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

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

Thus, by (A.17) and (A.18), we have

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

where the second inequality is due to 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

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

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

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

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

For the first and second terms, we have

$$\begin{aligned}
&\eta\left(\|\widehat{\boldsymbol{\beta}}_{i\Delta_n}\|_1 - \|\boldsymbol{\beta}_0\|_1\right) + \langle \nabla\mathcal{L}_{\tau,i}(\boldsymbol{\beta}_0), \widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0\rangle \\
&\geq -\eta\|\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0\|_1 - \|\nabla\mathcal{L}_{\tau,i}(\boldsymbol{\beta}_0)\|_{\max}\|\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0\|_1 \\
&\geq -\frac{12\kappa+72}{\kappa+12}\sqrt{s_p}n^{-\delta/2}\eta^{1-\delta/2}\|\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_0\|_2, \qquad \textbf{(A.23)}
\end{aligned}$$

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

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

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 $\ell_2$ norm error bound. Then, by (A.19), we can show the $\ell_1$ norm error bound. □

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

**TABLE A1.** 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 | U.S. 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 U.S. Treasury Bond |
| | ZF | 5-Year U.S. Treasury Note |
| | ZN | 10-Year U.S. Treasury Note |
| | ZQ | 30-Day Fed Funds |
| | ZT | 2-Year U.S. 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 |

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

**TABLE A2.** The number of non-zero monthly integrated beta estimates across 60 factors and five assets over 84 months for the RED-LASSO (RED), ED-LASSO (ED), LASSO, and SVR estimators

| | | AAPL | | | | BRK.B | | | | AMZN | | | | GOOG | | | | XOM | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Type | Symbol | RED | ED | LASSO | SVR | RED | ED | LASSO | SVR | RED | ED | LASSO | SVR | RED | ED | LASSO | SVR | RED | ED | LASSO | SVR |
| Commodity | CA | 4 | 28 | 9 | 84 | 3 | 34 | 21 | 84 | 2 | 23 | 15 | 84 | 3 | 21 | 15 | 84 | 6 | 32 | 33 | 84 |
| | CL | 5 | 36 | 13 | 84 | 6 | 27 | 38 | 84 | 7 | 35 | 27 | 84 | 6 | 39 | 21 | 84 | 64 | 65 | 79 | 84 |
| | GC | 15 | 36 | 16 | 84 | 11 | 36 | 40 | 84 | 8 | 41 | 32 | 84 | 6 | 42 | 21 | 84 | 3 | 41 | 32 | 84 |
| | HG | 21 | 36 | 27 | 84 | 17 | 21 | 46 | 84 | 13 | 38 | 30 | 84 | 13 | 31 | 29 | 84 | 25 | 46 | 57 | 84 |
| | HO | 4 | 33 | 11 | 84 | 2 | 32 | 34 | 84 | 3 | 27 | 24 | 84 | 1 | 30 | 20 | 84 | 38 | 47 | 76 | 84 |
| | ML | 1 | 19 | 4 | 49 | 3 | 17 | 10 | 49 | 0 | 14 | 11 | 49 | 3 | 22 | 11 | 49 | 3 | 11 | 16 | 49 |
| | NG | 6 | 37 | 9 | 84 | 3 | 28 | 23 | 84 | 4 | 24 | 17 | 84 | 2 | 18 | 11 | 84 | 9 | 29 | 35 | 84 |
| | OJ | 2 | 23 | 9 | 84 | 4 | 26 | 21 | 84 | 7 | 22 | 16 | 84 | 3 | 26 | 12 | 84 | 4 | 30 | 29 | 84 |
| | PA | 1 | 26 | 11 | 84 | 6 | 30 | 28 | 84 | 4 | 32 | 19 | 84 | 5 | 26 | 14 | 84 | 8 | 28 | 38 | 84 |
| | PL | 5 | 28 | 8 | 84 | 5 | 27 | 23 | 84 | 7 | 30 | 12 | 84 | 2 | 28 | 15 | 84 | 5 | 30 | 41 | 84 |
| | RB | 7 | 35 | 13 | 84 | 1 | 28 | 37 | 84 | 2 | 38 | 24 | 84 | 5 | 32 | 19 | 84 | 44 | 40 | 73 | 84 |
| | RM | 0 | 15 | 7 | 52 | 3 | 11 | 15 | 52 | 4 | 11 | 13 | 52 | 3 | 14 | 12 | 52 | 3 | 21 | 20 | 52 |
| | RS | 2 | 25 | 9 | 84 | 5 | 24 | 25 | 84 | 3 | 19 | 14 | 84 | 2 | 15 | 13 | 84 | 0 | 17 | 25 | 84 |
| | SI | 4 | 27 | 13 | 84 | 3 | 26 | 30 | 84 | 4 | 33 | 14 | 84 | 6 | 26 | 16 | 84 | 6 | 31 | 38 | 84 |
| | ZC | 2 | 28 | 10 | 84 | 5 | 24 | 17 | 84 | 2 | 34 | 17 | 84 | 6 | 35 | 20 | 84 | 3 | 32 | 36 | 84 |
| | ZL | 3 | 21 | 8 | 84 | 4 | 27 | 22 | 84 | 4 | 32 | 17 | 84 | 6 | 27 | 14 | 84 | 9 | 33 | 34 | 84 |
| | ZM | 3 | 24 | 8 | 84 | 2 | 36 | 22 | 84 | 2 | 33 | 17 | 84 | 9 | 30 | 15 | 84 | 6 | 36 | 30 | 84 |
| | ZO | 3 | 27 | 11 | 84 | 1 | 37 | 27 | 84 | 4 | 36 | 14 | 84 | 5 | 29 | 13 | 84 | 4 | 34 | 28 | 84 |
| | ZR | 3 | 36 | 12 | 84 | 4 | 23 | 22 | 84 | 5 | 23 | 15 | 84 | 5 | 40 | 13 | 84 | 5 | 25 | 29 | 84 |
| | ZW | 3 | 30 | 7 | 84 | 0 | 26 | 21 | 84 | 6 | 41 | 19 | 84 | 3 | 29 | 17 | 84 | 1 | 30 | 33 | 84 |
| Currency | A6 | 10 | 40 | 23 | 84 | 17 | 43 | 46 | 84 | 14 | 30 | 29 | 84 | 17 | 37 | 26 | 84 | 18 | 44 | 59 | 84 |
| | AD | 14 | 38 | 23 | 84 | 12 | 32 | 41 | 84 | 13 | 37 | 33 | 84 | 13 | 38 | 27 | 84 | 32 | 34 | 68 | 84 |
| | B6 | 3 | 37 | 12 | 84 | 4 | 29 | 32 | 84 | 1 | 38 | 22 | 84 | 6 | 31 | 16 | 84 | 4 | 38 | 38 | 84 |
| | BR | 5 | 23 | 13 | 83 | 5 | 34 | 23 | 83 | 4 | 27 | 18 | 83 | 8 | 31 | 15 | 83 | 3 | 31 | 35 | 83 |
| | DX | 3 | 42 | 10 | 84 | 2 | 31 | 32 | 84 | 4 | 38 | 20 | 84 | 5 | 32 | 20 | 84 | 6 | 35 | 35 | 84 |
| | E1 | 5 | 29 | 13 | 84 | 12 | 42 | 36 | 84 | 6 | 33 | 25 | 84 | 7 | 36 | 22 | 84 | 6 | 38 | 33 | 84 |
| | E6 | 1 | 33 | 12 | 84 | 3 | 39 | 31 | 84 | 4 | 36 | 22 | 84 | 2 | 32 | 18 | 84 | 2 | 37 | 38 | 84 |
| | J1 | 24 | 41 | 38 | 84 | 26 | 40 | 58 | 84 | 40 | 56 | 44 | 84 | 35 | 46 | 34 | 84 | 25 | 36 | 49 | 84 |
| | RP | 4 | 30 | 8 | 84 | 7 | 25 | 27 | 84 | 5 | 24 | 23 | 84 | 8 | 21 | 13 | 84 | 2 | 32 | 34 | 84 |
| | RU | 1 | 18 | 7 | 67 | 2 | 23 | 14 | 67 | 6 | 30 | 19 | 67 | 5 | 31 | 10 | 67 | 3 | 25 | 30 | 67 |

(Continued)

**TABLE A2.** (continued)

| | | AAPL | | | | BRK.B | | | | AMZN | | | | GOOG | | | | XOM | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Type | Symbol | RED | ED | LASSO | SVR | RED | ED | LASSO | SVR | RED | ED | LASSO | SVR | RED | ED | LASSO | SVR | RED | ED | LASSO | SVR |
| Interest rate | BTP | 2 | 24 | 5 | 84 | 8 | 30 | 27 | 84 | 8 | 34 | 16 | 84 | 4 | 27 | 12 | 84 | 4 | 29 | 31 | 84 |
| | ED | 1 | 2 | 2 | 36 | 1 | 2 | 10 | 36 | 2 | 3 | 8 | 36 | 1 | 3 | 7 | 36 | 0 | 3 | 11 | 36 |
| | G | 4 | 34 | 14 | 84 | 7 | 26 | 35 | 84 | 7 | 35 | 14 | 84 | 1 | 41 | 17 | 84 | 5 | 29 | 35 | 84 |
| | GG | 12 | 31 | 17 | 84 | 7 | 36 | 42 | 84 | 5 | 31 | 27 | 84 | 9 | 34 | 20 | 84 | 5 | 36 | 45 | 84 |
| | HR | 8 | 29 | 16 | 84 | 8 | 28 | 31 | 84 | 6 | 27 | 25 | 84 | 9 | 29 | 18 | 84 | 4 | 31 | 36 | 84 |
| | US | 7 | 36 | 23 | 84 | 11 | 34 | 51 | 84 | 10 | 31 | 39 | 84 | 11 | 33 | 26 | 84 | 7 | 32 | 36 | 84 |
| | ZF | 5 | 33 | 20 | 84 | 15 | 33 | 48 | 84 | 6 | 33 | 31 | 84 | 10 | 36 | 28 | 84 | 9 | 31 | 45 | 84 |
| | ZN | 6 | 37 | 20 | 84 | 12 | 34 | 49 | 84 | 11 | 31 | 28 | 84 | 9 | 36 | 30 | 84 | 11 | 31 | 42 | 84 |
| | ZQ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | ZT | 8 | 25 | 15 | 82 | 5 | 21 | 30 | 82 | 8 | 24 | 22 | 82 | 6 | 25 | 18 | 82 | 6 | 34 | 40 | 82 |
| Stock market index | DY | 49 | 59 | 73 | 84 | 43 | 34 | 76 | 84 | 40 | 54 | 73 | 84 | 43 | 43 | 72 | 84 | 21 | 48 | 76 | 84 |
| | ES | 71 | 70 | 82 | 84 | 81 | 80 | 84 | 84 | 80 | 75 | 84 | 84 | 77 | 74 | 84 | 84 | 57 | 53 | 84 | 84 |
| | EW | 23 | 42 | 62 | 84 | 62 | 65 | 80 | 84 | 38 | 44 | 66 | 84 | 32 | 56 | 60 | 84 | 22 | 42 | 75 | 84 |
| | FX | 30 | 44 | 58 | 84 | 38 | 40 | 70 | 84 | 34 | 39 | 60 | 84 | 40 | 41 | 65 | 84 | 21 | 36 | 75 | 84 |
| | MME | 56 | 57 | 71 | 83 | 56 | 48 | 76 | 83 | 62 | 66 | 77 | 83 | 63 | 62 | 75 | 83 | 49 | 48 | 80 | 83 |
| | MX | 42 | 44 | 60 | 84 | 36 | 53 | 76 | 84 | 40 | 40 | 67 | 84 | 49 | 46 | 67 | 84 | 32 | 44 | 79 | 84 |
| | NQ | 84 | 84 | 84 | 84 | 41 | 52 | 75 | 84 | 84 | 84 | 84 | 84 | 84 | 84 | 84 | 84 | 10 | 53 | 68 | 84 |
| | RTY | 47 | 50 | 74 | 84 | 48 | 44 | 72 | 84 | 61 | 56 | 74 | 84 | 44 | 58 | 67 | 84 | 25 | 45 | 70 | 84 |
| | VX | 57 | 60 | 67 | 84 | 59 | 56 | 76 | 84 | 61 | 61 | 64 | 84 | 65 | 60 | 71 | 84 | 38 | 43 | 74 | 84 |
| | X | 39 | 43 | 58 | 84 | 60 | 55 | 74 | 83 | 33 | 45 | 61 | 84 | 47 | 44 | 68 | 84 | 54 | 63 | 82 | 84 |
| | XAE | 11 | 31 | 20 | 84 | 12 | 29 | 47 | 84 | 10 | 39 | 33 | 84 | 17 | 42 | 29 | 84 | 76 | 80 | 81 | 84 |
| | XAF | 11 | 35 | 20 | 84 | 50 | 65 | 66 | 84 | 12 | 31 | 26 | 84 | 18 | 44 | 31 | 84 | 12 | 37 | 44 | 84 |
| | XAI | 5 | 37 | 19 | 84 | 18 | 44 | 46 | 84 | 8 | 45 | 27 | 84 | 11 | 42 | 29 | 84 | 11 | 38 | 49 | 84 |
| | YM | 65 | 75 | 78 | 84 | 83 | 82 | 84 | 84 | 49 | 52 | 79 | 84 | 45 | 54 | 76 | 84 | 72 | 76 | 84 | 84 |
| Six factors | HML | 28 | 45 | 28 | 84 | 55 | 68 | 63 | 84 | 51 | 67 | 50 | 84 | 45 | 62 | 43 | 84 | 61 | 64 | 79 | 84 |
| | SMB | 17 | 42 | 19 | 84 | 73 | 81 | 68 | 84 | 19 | 40 | 28 | 84 | 13 | 40 | 23 | 84 | 66 | 77 | 64 | 84 |
| | RMW | 6 | 37 | 14 | 84 | 39 | 58 | 52 | 84 | 5 | 36 | 23 | 84 | 8 | 35 | 20 | 84 | 66 | 73 | 72 | 84 |
| | CMA | 21 | 40 | 28 | 84 | 15 | 44 | 35 | 84 | 42 | 56 | 45 | 84 | 38 | 56 | 39 | 84 | 63 | 68 | 70 | 84 |
| | MOM | 29 | 54 | 35 | 84 | 42 | 60 | 56 | 84 | 41 | 61 | 44 | 84 | 38 | 53 | 37 | 84 | 62 | 77 | 68 | 84 |
| | MKT | 30 | 60 | 67 | 84 | 84 | 84 | 84 | 84 | 40 | 52 | 73 | 84 | 46 | 58 | 70 | 84 | 84 | 84 | 84 | 84 |

## 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 $1-p^{-2-a}$,

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

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

For $\widehat{\boldsymbol{\Omega}}_{i\Delta_n}$, similar to the proofs of Theorem 1 (Kim et al., 2025), 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 et al., 2025), 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

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

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

Note that

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

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

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

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

Let

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

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

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

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

where

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

Then, we have

$$\begin{aligned}
\|\widehat{I\beta} - I\beta_0\|_{\max} \le & \left\| \sum_{i=0}^{[1/(k_n\Delta_n)]-2} \left( \widetilde{\boldsymbol{\beta}}_{ik_n\Delta_n} - \widetilde{\boldsymbol{\beta}}^{(2)}_{ik_n\Delta_n} \right) k_n\Delta_n \right\|_{\max} \\
& + \left\| \sum_{i=0}^{[1/(k_n\Delta_n)]-2} \left( \widetilde{\boldsymbol{\beta}}^{(2)}_{ik_n\Delta_n} - \widetilde{\boldsymbol{\beta}}^{(3)}_{ik_n\Delta_n} \right) k_n\Delta_n \right\|_{\max} \\
& + \left\| \sum_{i=0}^{[1/(k_n\Delta_n)]-2} \left( \widetilde{\boldsymbol{\beta}}^{(3)}_{ik_n\Delta_n} - \widetilde{\boldsymbol{\beta}}^{(4)}_{ik_n\Delta_n} \right) k_n\Delta_n \right\|_{\max} \\
& + \left\| \sum_{i=0}^{[1/(k_n\Delta_n)]-2} \left( \widetilde{\boldsymbol{\beta}}^{(4)}_{ik_n\Delta_n} - \widetilde{\boldsymbol{\beta}}^{(5)}_{ik_n\Delta_n} \right) k_n\Delta_n \right\|_{\max} \\
& + \left\| \sum_{i=0}^{[1/(k_n\Delta_n)]-2} \left( \widetilde{\boldsymbol{\beta}}^{(5)}_{ik_n\Delta_n} - \boldsymbol{\beta}_0((i+1)k_n\Delta_n) \right) k_n\Delta_n \right\|_{\max} \\
& + \left\| \sum_{i=0}^{[1/(k_n\Delta_n)]-2} \int_{ik_n\Delta_n}^{(i+1)k_n\Delta_n} \left( \boldsymbol{\beta}_0((i+1)k_n\Delta_n) - \boldsymbol{\beta}_0(t) \right) dt \right\|_{\max} \\
& + \left\| \int_{([1/(k_n\Delta_n)]-1)k_n\Delta_n}^{1} \boldsymbol{\beta}_0(t)dt \right\|_{\max} \\
= & (I) + (II) + (III) + (IV) + (V) + (VI) + (VII). \qquad \textbf{(A.29)}
\end{aligned}$$

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

$$\Pr\left\{ (I) \le C s_p^{2-\delta} n^{(-2+\delta)/4} (\log p)^{(5-3\delta)/4} \right\} \ge 1 - p^{-1-a}. \qquad \textbf{(A.30)}$$

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

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

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

Consider $(III)$. Similar to the proofs of (A.20) in Kim et al. (2025), we can show, with probability at least $1-p^{-1-a}$,

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

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

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

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

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

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

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

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

$$\leq C\left(n^{-1/4}\sqrt{\log p}+n^{-1/2}(\log p)^2+\lambda\right)\times s_p\left(s_p n^{-1/4}(\log p)^{3/4}\right)^{1-\delta}$$
$$\leq C s_p^{2-\delta} n^{(-2+\delta)/4}(\log p)^{(5-3\delta)/4}. \quad \textbf{(A.33)}$$

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

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

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

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

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

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

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

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

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 $1-2p^{-2-a}$,

$$\left\|(V)^{(2)}\right\|_{\max}$$
$$\leq\sup_{0\leq i\leq[1/(k_n\Delta_n)]-2}\sup_{1\leq j\leq p}\Bigg|\mathbb{E}\left\{[\psi_\varpi\left(g(ik_n)\right)]_j\middle|\mathcal{F}_{(i+1)k_n\Delta_n}\right\}$$
$$-\left[\boldsymbol{\beta}_0((i+1)k_n\Delta_n)-\widehat{\boldsymbol{\beta}}_{ik_n\Delta_n}\right]_j\Bigg|$$
$$=\sup_{0\leq i\leq[1/(k_n\Delta_n)]-2}\sup_{1\leq j\leq p}\left|\mathbb{E}\left\{[\psi_\varpi\left(g(ik_n)\right)-g(ik_n)]_j\middle|\mathcal{F}_{(i+1)k_n\Delta_n}\right\}\right|$$

https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

$$\leq \sup_{0\leq i\leq [1/(k_n\Delta_n)]-2} \sup_{1\leq j\leq p} \mathbb{E}\left\{\left|[g(ik_n)]_j\right| \mathbf{1}\left(\left|[g(ik_n)]_j\right| > \varpi\right) \middle| \mathcal{F}_{(i+1)k_n\Delta_n}\right\}$$

$$\leq \sup_{0\leq i\leq [1/(k_n\Delta_n)]-2} \sup_{1\leq j\leq p} \mathbb{E}\left\{\left|[g(ik_n)]_j\right|^2/\varpi \middle| \mathcal{F}_{(i+1)k_n\Delta_n}\right\}$$

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

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

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

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 et al., 2025), we can show

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

□

## REFERENCES


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

Aït-Sahalia, Y., & 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. In T. B. Fomby & D. Terrell (Eds.), *Econometric analysis of financial and economic time series* (pp. 1–39). Emerald Group Publishing Limited.

Ang, A., & Kristensen, D. (2012). Testing conditional factor models. *Journal of Financial Economics*, 106(1), 132–156.

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

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


https://doi.org/10.1017/S0266466625100236 Published online by Cambridge University Press

Barndorff-Nielsen, O. E., & 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., & Luo, X. (2011). A constrained $\ell_1$ minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494), 594–607.

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

Candes, E., & 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. 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., & Gallo, G. M. (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics*, 131(1), 3–27.

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

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

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

Fan, J., & 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., Li, Q., & Wang, Y. (2017). Estimation of high dimensional mean regression in the absence of symmetry and light tail assumptions. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79(1), 247–265.

Fan, J., & 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., & 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., & 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.

Gabaix, X., Gopikrishnan, P., Plerou, V., & Stanley, H. E. (2003). A theory of power-law distributions in financial market fluctuations. *Nature*, 423(6937), 267–270.

Hansen, P. R., Huang, Z., & 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., & Zhu, H. (2016). . . . and the cross-section of expected returns. *The Review of Financial Studies*, 29(1), 5–68.

Hou, K., Xue, C., & 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., & Protter, P. E. (2011). *Discretization of processes* (Vol. 67). Springer Science & Business Media.

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

Javanmard, A., & 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*, 41, 1–12.

Kim, D., & 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., Oh, M., and Shin, M. (2025). High-dimensional time-varying coefficient estimation in diffusion models. Preprint, arXiv:2202.08419.

Kim, D., & 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.

Kong, X.-B., Lin, J.-G., Liu, C., & Liu, G.-Y. (2023). Discrepancy between global and local principal component analysis on large-panel high-frequency data. *Journal of the American Statistical Association*, 118(542), 1333–1344.

Kong, X.-B., & Liu, C. (2018). Testing against constant factor loading matrix with large panel high-frequency data. *Journal of Econometrics*, 204(2), 301–319.

Li, J., Todorov, V., & 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.

Mancini, C. (2009). Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics*, 36(2), 270–296.

Mancini, C. (2017). Truncated realized covariance when prices have infinite variation jumps. *Stochastic Processes and their Applications*, 127(6), 1998–2035.

Mao, G., & 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., & Pontiff, J. (2016). Does academic research destroy stock return predictability? *The Journal of Finance*, 71(1), 5–32.

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

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

Reiß, M., Todorov, V., & 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., & 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., & Fan, J. (2023). Adaptive robust large volatility matrix estimation based on high-frequency financial data. *Journal of Econometrics*, 237(1), 105514.

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

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

Tan, K. M., Sun, Q., & Witten, D. (2023). Sparse reduced rank Huber regression in high dimensions. *Journal of the American Statistical Association*, 118(544), 2383–2393.

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., & Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*, 42(3), 1166–1202.

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

Zhang, C.-H., & 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.