High-Dimensional Time-Varying Coefficient Estimation

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

January 16, 2024

Abstract

In this paper, we develop a novel high-dimensional time-varying coefficient estimation method, based on high-dimensional Itô diffusion processes. To account for high-dimensional time-varying coefficients, we first estimate local (or instantaneous) coefficients using a time-localized Dantzig selection scheme under a sparsity condition, which results in biased local coefficient estimators due to the regularization. To handle the bias, we propose a debiasing scheme, which provides well-performing unbiased local coefficient estimators. With the unbiased local coefficient estimators, we estimate the integrated coefficient, and to further account for the sparsity of the coefficient process, we apply thresholding schemes. We call this Thresholding dEbiased Dantzig (TED). We establish asymptotic properties of the proposed TED estimator. In the empirical analysis, we apply the TED procedure to analyzing high-dimensional factor models using high-frequency data.

Keywords: Dantzig selection, debiased, diffusion process, factor model, high-frequency data, sparsity

*Corresponding author. Address: College of Business, KAIST, Seoul 02455, South Korea. Phone: +82-02-958-3448. E-mail: donggyukim@kaist.ac.kr.

1 Introduction

To explain various data types, numerous regression-based models have been developed. Especially, advances in technology provide us big data, which causes the curse of dimensionality problem. To tackle this problem in the high-dimensional regression, we usually assume the sparsity of variables, that is, the number of significant coefficients is small. To accommodate the sparsity condition, we often employ the LASSO procedure (Tibshirani, 1996), SCAD (Fan and Li, 2001), and Dantzig selector (Candes and Tao, 2007). The works of Belloni et al. (2014); Feng et al. (2020); Yuan and Lin (2006); Zou (2006) are useful for further reading. There are numerous related papers that can be found in the above literature. These estimation methods result in sparse coefficients, and under the sparsity condition, they are consistent estimators (Negahban et al., 2012). Under the diffusion process, Ciolek et al. (2022) studied the properties of the LASSO estimator of the drift component and Gaïffas and Matulewicz (2019) proposed the estimation procedure for the drift parameter in the high-dimensional Ornstein–Uhlenbeck (OU) process.

On the other hand, in high-frequency finance, we often observe that coefficients in the regression model are time-varying. For example, Andersen et al. (2021) investigated the intra-day variation of the local coefficients, which are called the market betas, between the individual assets and market index. To account for the time-varying feature, Mykland and Zhang (2009) computed the market beta as the aggregation of the market betas estimated over local blocks. To evaluate the coefficients of multi factor models, Aït-Sahalia et al. (2020) proposed an integrated coefficient approach using the local coefficient. See also Chen (2018); Oh et al. (2022). We call this high-frequency regression. Recently, Chen et al. (2023) proposed the high-dimensional market beta estimation procedure with large dependent variables and almost finite common factors. However, in the field of finance, there are hundreds of potential factor candidates that explain the cross section of expected stock returns (Cochrane, 2011; Harvey et al., 2016; Hou et al., 2020; McLean and Pontiff, 2016). Thus, we also encounter the curse of dimensionality in high-frequency regressions, so the estimation methods developed for the finite dimension fail to estimate the coefficients consistently. To overcome this issue, we can consider the high-dimensional regression methods such as the LASSO (Tibshirani, 1996), Dantzig selector (Candes and Tao, 2007), and SCAD (Fan and Li, 2001). However, the direct application of these methods cannot explain the time-varying feature of the coefficient process and may suffer from the model errors. Thus, to fully benefit from the utilization of high-frequency financial data in the high-dimensional regression, we need to develop methodologies that can handle both the curse of dimensionality as well as the time-varying coefficients.

In this paper, we introduce a novel high-dimensional high-frequency regression estimation procedure which can accommodate the sparse and time-varying coefficient processes. To model the high-frequency data, we employ diffusion processes whose stochastic difference equations have a time series regression structure. We also assume that the coefficient process $\boldsymbol{\beta}_t$ follows a diffusion process. In this paper, the parameter of interest is the integrated coefficient, $\int_0^1 \boldsymbol{\beta}_t dt$, that represents the average relationship between variables. To handle the curse of dimensionality, we assume that the coefficient processes are sparse, and to account for the sparsity of the time-varying coefficient process, we employ the Dantzig selector procedure (Candes and Tao, 2007). Specifically, due to the time-varying phenomena, we cannot estimate the integrated coefficient directly, and so we first estimate the instantaneous (or local) coefficients using the time-localized Dantzig selector procedure, based on the definition of $\boldsymbol{\beta}_t$. Then, to mitigate the bias coming from the regularization of the Dantzig selector, we propose a debiasing scheme and estimate the integrated coefficient with the debiased Dantzig instantaneous coefficients. With the debiasing scheme, we can obtain more accurate estimators in terms of the element-wise convergence rate; however, the estimated integrated coefficient is not sparse. Thus, to accommodate the sparsity, we further regularize the estimated integrated coefficient. We call this Thresholding dEbiased Dantzig (TED). We also establish its asymptotic properties.

The rest of paper is organized as follows. Section 2 introduces the model set-up. Section 3 proposes the TED estimation procedure and establishes its asymptotic properties. In Section 4, we conduct a simulation study to check the finite sample performance of the TED estimation procedure, and in Section 5, we apply the TED to the high-frequency financial data. The conclusion is presented in Section 6, and we collect all of the proofs in the supplementary materials.

2 The model set-up

We consider the following non-parametric time series regression diffusion model:

$$dY_t = \boldsymbol{\beta}_t^\top d\mathbf{X}_t + dZ_t, \tag{2.1}$$

where Y_t is a dependent process, $\mathbf{X}_t$ is a p-dimensional multivariate covariate process, $\boldsymbol{\beta}_t$ is a coefficient process, and Z_t is a residual process. The p-dimensional covariate process $\mathbf{X}_t$ and residual process Z_t satisfy

$$d\mathbf{X}_t = \boldsymbol{\mu}_t dt + \boldsymbol{\sigma}_t d\mathbf{B}_t \quad \text{and} \quad dZ_t = \nu_t dW_t,$$

where $\boldsymbol{\mu}_t$ is a drift process, $\boldsymbol{\sigma}_t$ and ν_t are instantaneous volatility processes, $\mathbf{B}_t$ and W_t are p-dimensional and one-dimensional standard Brownian motions, respectively, and $\mathbf{B}_t$ and W_t are independent. The processes $\boldsymbol{\mu}_t$, $\boldsymbol{\beta}_t$, $\boldsymbol{\sigma}_t$, and ν_t are predictable. To account for the time-varying coefficient, we further assume that the coefficient $\boldsymbol{\beta}_t$ satisfies the following diffusion process:

$$d\boldsymbol{\beta}_t = \boldsymbol{\mu}_{\beta,t} dt + \boldsymbol{\nu}_{\beta,t} d\mathbf{W}_t^{\beta}, \tag{2.2}$$

where $\boldsymbol{\mu}_{\beta,t}$ and $\boldsymbol{\nu}_{\beta,t}$ are predictable, and $\mathbf{W}_t^{\beta}$ is p-dimensional standard Brownian motion. To figure out the average relationship between the covariate and dependent processes, we consider the integrated coefficient:

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

In finance, hundreds of potential factor candidates have been proposed in order to explain the cross section of expected stock returns (Cochrane, 2011; Harvey et al., 2016; Hou et al., 2020; McLean and Pontiff, 2016). That is, the dimensionality, p, of the covariate process is large. Thus, we often run into the curse of dimensionality problem when handling financial data. However, all of them may not be significant; thus, to account for this, we assume that the coefficient process $\boldsymbol{\beta}_t = (\beta_{1t}, \dots, \beta_{pt})^\top$ satisfies the following sparsity condition:

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

where 0^0 is defined as 0, $\delta \in [0, 1)$, and s_p is diverging slowly with respect to p, for example, $\log p$. We investigate asymptotic properties under this general sparsity case. However, in practice, it is harmless to assume $\delta = 0$. That is, we can assume that several factors are significant, while others do not affect on the expected returns. We note that with the randomness of the coefficient process, in general, the sparsity condition (2.3) is satisfied with high probability. However, for simplicity, we assume that the sparsity condition is satisfied almost surely. The sparsity condition is widely employed in the high-frequency finance literature (Ciolek et al., 2022; Gaïffas and Matulewicz, 2019; Kim et al., 2016, 2018; Tao et al., 2013; Wang and Zou, 2010).

3 High-dimensional high-frequency regression

3.1 Estimation procedure

In this section, we propose an estimation procedure for large integrated coefficients. We first fix some notations. For any given p_1 by p_2 matrix $\mathbf{U} = (U_{ij})$, let

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

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

From the model (2.1), the instantaneous coefficient $\boldsymbol{\beta}_t$ satisfies the following equation:

$$\frac{d}{dt}[Y, \mathbf{X}]_t = \boldsymbol{\beta}_t^\top \frac{d}{dt}[\mathbf{X}, \mathbf{X}]_t \text{ a.s.},$$

where $[\cdot, \cdot]$ denotes the quadratic variation. The coefficient process $\boldsymbol{\beta}_t$ is a function of instantaneous volatilities of $\mathbf{X}$ and Y as follows:

$$\boldsymbol{\beta}_t = \boldsymbol{\Sigma}_t^{-1} \boldsymbol{\Sigma}_{XY,t} \text{ a.s.}, \tag{3.1}$$

where $\boldsymbol{\Sigma}_t = \boldsymbol{\sigma}_t \boldsymbol{\sigma}_t^\top$ and $\boldsymbol{\Sigma}_{XY,t} = \frac{d}{dt}[\mathbf{X}, Y]_t$. Thus, the instantaneous coefficient can be estimated by the instantaneous volatility estimators. For the finite dimensional case, the instantaneous volatility-based estimation procedure works well (Aït-Sahalia et al., 2020). However, this approach cannot explain the sparse structure (2.3). Furthermore, when the dimensionality of the covariate $\mathbf{X}$ is larger than the sample size, this approach fails to consistently estimate the instantaneous coefficient. Therefore, the procedure developed for the finite dimension is neither effective nor efficient. In contrast, the direct application of the high-dimensional regression procedure such as the LASSO (Tibshirani, 1996) and Dantzig selector (Candes and Tao, 2007) cannot consistently estimate the integrated coefficients. Specifically, we can rewrite (2.1) as follows:

$$dY_t = \boldsymbol{\beta}^\top d\mathbf{X}_t + (\boldsymbol{\beta}_t - \boldsymbol{\beta})^\top d\mathbf{X}_t + dZ_t.$$

Due to the time-variation of the coefficient process, we have a non-negligible dependent structure between $\boldsymbol{\beta}^\top d\mathbf{X}_t$ and $(\boldsymbol{\beta}_t - \boldsymbol{\beta})^\top d\mathbf{X}_t$. This produces a bias for the usual high-dimensional regression methods. To mitigate the dependency and accommodate the sparse structure of the coefficient process in (2.3), we employ the time-localized Dantzig selection method as follows. Let $\Delta_i^n A = A_{i\Delta_n} - A_{(i-1)\Delta_n}$ for $1 \le i \le 1/\Delta_n$, where $\Delta_n = 1/n$ is the distance between adjacent observation time points. Define

$$\mathcal{Y}_i = \begin{pmatrix} \Delta_{i+1}^n Y \\ \Delta_{i+2}^n Y \\ \vdots \\ \Delta_{i+k_n}^n Y \end{pmatrix}, \quad \mathcal{X}_i = \begin{pmatrix} \Delta_{i+1}^n \mathbf{X}^\top \\ \Delta_{i+2}^n \mathbf{X}^\top \\ \vdots \\ \Delta_{i+k_n}^n \mathbf{X}^\top \end{pmatrix}, \quad \text{and} \quad \mathcal{Z}_i = \begin{pmatrix} \Delta_{i+1}^n Z \\ \Delta_{i+2}^n Z \\ \vdots \\ \Delta_{i+k_n}^n Z \end{pmatrix},$$

where k_n is the number of observations in each window to calculate the local regression. Then, we estimate the sparse instantaneous coefficient as follows:

$$\widehat{\boldsymbol{\beta}}_{i\Delta_n} = \arg\min \|\boldsymbol{\beta}\|_1 \quad \text{s.t.} \quad \left\| \frac{1}{k_n\Delta_n} \mathcal{X}_i^\top \mathcal{X}_i \boldsymbol{\beta} - \frac{1}{k_n\Delta_n} \mathcal{X}_i^\top \mathcal{Y}_i \right\|_{\max} \le \lambda_n, \tag{3.2}$$

where λ_n is a tuning parameter which converges to zero. We specify λ_n in Theorem 1. With the appropriate λ_n, we can show that the proposed Dantzig instantaneous coefficient estimator $\widehat{\boldsymbol{\beta}}_{i\Delta_n}$ is a consistent estimator (see Theorem 1). To estimate the integrated coefficient $I\beta$ with this consistent estimator, we usually consider the sum of the instantaneous volatility estimators $\widehat{\boldsymbol{\beta}}_{i\Delta_n}$'s.

However, the Dantzig estimator is biased, so their summation cannot enjoy the law of large number properties. For example, the error of the sum of the Dantzig instantaneous coefficient estimators is dominated by the bias terms, and so it has the same convergence rate as that of $\widehat{\boldsymbol{\beta}}_{i\Delta_n}$. To reduce the effect of the bias, we use a debiasing scheme as follows. We first estimate the inverse matrix of the instantaneous volatility matrix $\boldsymbol{\Sigma}_{i\Delta_n}$ using the constrained ℓ_1-minimization for inverse matrix estimation (CLIME) (Cai et al., 2011). Let $\widehat{\boldsymbol{\Omega}}_{i\Delta_n}$ be the solution of the following optimization problem:

$$\min \|\boldsymbol{\Omega}\|_1 \quad \text{s.t.} \quad \|\frac{1}{k_n\Delta_n}\mathcal{X}_i^\top \mathcal{X}_i \boldsymbol{\Omega} - \mathbf{I}\|_{\max} \le \tau_n, \tag{3.3}$$

where τ_n is the tuning parameter specified in Theorem 2. With the CLIME estimator, we adjust the Dantzig instantaneous coefficient estimator as follows:

$$\widetilde{\boldsymbol{\beta}}_{i\Delta_n} = \widehat{\boldsymbol{\beta}}_{i\Delta_n} + \frac{1}{k_n\Delta_n}\widehat{\boldsymbol{\Omega}}_{i\Delta_n}^\top \mathcal{X}_i^\top (\mathcal{Y}_i - \mathcal{X}_i \widehat{\boldsymbol{\beta}}_{i\Delta_n}). \tag{3.4}$$

Then, the debiased Dantzig instantaneous coefficient estimator satisfies

$$\widetilde{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_{0,i\Delta_n} = \frac{1}{k_n\Delta_n}\boldsymbol{\Omega}_{0,i\Delta_n}\left(\mathcal{X}_i^\top \mathcal{Z}_i + \mathcal{A}_i\right) + R_i \text{ a.s.},$$

where the subscript 0 represents the true parameters, $\mathcal{A}_i$ is a martingale difference defined in (A.9), and R_i is a negligible remaining error term (see Theorem 3). We note that the debiasing scheme is usually employed to derive asymptotic normality and to conduct the confidence interval construction or hypothesis test (Javanmard and Montanari, 2018; Van de Geer et al., 2014; Zhang and Zhang, 2014). However, we adopt the debiasing scheme to improve the integrated coefficient estimation. Specifically, the debiasing scheme helps enjoy the law of large number property when averaging the instantaneous coefficient estimators. The integrated coefficient estimator is defined by

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

As discussed above, the debiasing scheme helps improve the element-wise convergence rate of the debiased Dantzig integrated coefficient estimator. However, the debiased Dantzig integrated coefficient estimator does not satisfy the sparsity condition (2.3) due to the bias adjustment. To accommodate the sparsity of the integrated coefficient, we apply the thresholding scheme as follows:

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

where $\mathbf{1}(\cdot)$ is an indicator function, the thresholding function $s(\cdot)$ satisfies that $|s(x)-x| \leq h_n$, and h_n is a thresholding level specified in Theorem 4. Examples of the thresholding function $s(x)$ include the hard thresholding function $s(x) = x$ and the soft thresholding function $s(x) = x - \text{sign}(x)h_n$. For the empirical study, we employed the hard thresholding function. We call this the Thresholded dEbiased Dantzig (TED) estimator. We summarize the TED estimation procedure in Algorithm 1.

3.2 Asymptotic results

In this section, we establish the asymptotic properties for the proposed TED estimation procedure. To investigate the asymptotic properties, we need the following technical conditions.

Assumption 1.

(a) The volatility process $\boldsymbol{\Sigma}_t = (\Sigma_{ijt})_{i,j=1,\ldots,p}$ *satisfies the following Hölder condition:*

$$|\Sigma_{ijt} - \Sigma_{ijs}| \leq C|t-s|^{1/2} \;\; a.s.$$

Algorithm 1 TED estimation procedure

Step 1 Estimate the instantaneous coefficient:

$$\widehat{\boldsymbol{\beta}}_{i\Delta_n} = \arg\min \|\boldsymbol{\beta}\|_1 \quad \text{s.t.} \quad \left\| \frac{1}{k_n \Delta_n} \mathcal{X}_i^\top \mathcal{X}_i \boldsymbol{\beta} - \frac{1}{k_n \Delta_n} \mathcal{X}_i^\top \mathcal{Y}_i \right\|_{\max} \le \lambda_n,$$

where $\lambda_n = C_\lambda s_p \sqrt{\log p} \left(\sqrt{k_n \Delta_n} + k_n^{-1/2} \right)$ and $k_n = c_k n^{1/2}$ for some large constants C_λ and c_k.

Step 2 Estimate the inverse instantaneous volatility matrix:

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

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

Step 3 Debias the Dantzig instantaneous coefficient estimator:

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

Step 4 Estimate the integrated coefficient:

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

Step 5 Threshold the debiased Dantzig integrated coefficient estimator:

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

where $\mathbf{1}(\cdot)$ is an indicator function, the thresholding function $s(\cdot)$ satisfies that $|s(x) - x| \le h_n$, $h_n = C_h b_n$ for some constant C_h, and b_n is defined in Theorem 3.

(b) $\boldsymbol{\mu}_t$, $\boldsymbol{\mu}_{\beta,t}$, $\boldsymbol{\beta}_t$, ν_t, $\boldsymbol{\Sigma}_t$, *and* $\boldsymbol{\Sigma}_{\beta,t} = \boldsymbol{\nu}_{\beta,t} \boldsymbol{\nu}_{\beta,t}^\top$ *are almost surely bounded, and* $\|\boldsymbol{\Sigma}_t^{-1}\|_1 \le C$ *a.s.*

(c) The drift process $\boldsymbol{\mu}_{\beta,t} = (\mu_{\beta,1t}, \dots, \mu_{\beta,pt})^\top$ *and the volatility process* $\boldsymbol{\Sigma}_{\beta,t} = (\Sigma_{\beta,ijt})_{i,j=1,\dots,p}$ *satisfy the following sparsity condition for* $\delta \in [0,1)$*:*

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

(d) $n^{c_1} \le p \le c_2 \exp(n^{c_3})$ *for some positive constants* c_1, c_2, *and* $c_3 < 1/8$, *and* $s_p^2 \log p \Delta_n k_n \to 0$ *as* $n, p \to \infty$.

(e) *The inverse matrix of the volatility matrix process,* $\boldsymbol{\Sigma}_t^{-1} = \boldsymbol{\Omega}_t = (\omega_{ijt})_{i,j=1,\dots,p}$, *satisfies the following sparsity condition:*

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

where $q \in [0,1)$ *and* $s_{\omega,p}$ *is diverging slowly with respect to* p*, for example,* $\log p$.

Remark 1. To investigate estimators of time-varying processes, we need continuity conditions such as Assumption 1(a) and the diffusion process structures for $\mathbf{X}_t$, Y_t and $\boldsymbol{\beta}_t$ in Section 2. Even if Assumption 1(a) is replaced by the condition that $\boldsymbol{\Sigma}_t$ has a continuous Itô diffusion process structure with bounded drift and instantaneous volatility processes, we can obtain the same theoretical results with up to $\sqrt{\log p}$ order. For simplicity, we put Assumption 1(a). The boundedness condition Assumption 1(b) provides sub-Gaussian tails which are often required to investigate high-dimensional inferences. On the other hand, when we investigate the asymptotic behaviors of volatility estimators such as their convergence rate, the boundedness condition can be relaxed to the locally boundedness condition (see Aït-Sahalia and Xiu (2017)). Specifically, Jacod and Protter (2011) showed in Lemma 4.4.9 that if the asymptotic result, such as stable convergence in law or convergence in probability, is satisfied under the boundedness condition, it is also satisfied under the locally boundedness condition. Thus, the asymptotic results established in this paper also hold for the locally boundedness condition. The sparsity condition for the coefficient process, Assumption 1(c), is the technical condition for investigating the discretization error of Dantzig instantaneous coefficient estimator $\widehat{\boldsymbol{\beta}}_{i\Delta_n}$. Finally, to investigate asymptotic properties of the CLIME estimator, we need the sparse inverse matrix condition Assumption 1(e) (Cai et al., 2011). Furthermore, if the smallest eigenvalue of $\boldsymbol{\Sigma}_t$ is strictly bigger than zero, the Frobenius norm of $\boldsymbol{\Omega}_t$ is bounded by $C\sqrt{p}$. This implies that the inverse matrix, $\boldsymbol{\Omega}_t$, is not dense. Since the strict positiveness of the smallest eigenvalue is the minimum requirement to investigate the regression-based models, Assumption 1(e) is reasonable.

In Theorems 1 and 2 below, we establish asymptotic properties for the sparse instantaneous coefficient and inverse matrix. Note that we use subscript 0 for the true parameters.

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

$$\max_i \|\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_{0,i\Delta_n}\|_{\max} \leq C\lambda_n \quad \text{and} \quad \max_i \|\widehat{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_{0,i\Delta_n}\|_1 \leq C s_p \lambda_n^{1-\delta}, \tag{3.5}$$

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

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

$$\max_i \|\widehat{\boldsymbol{\Omega}}_{i\Delta_n} - \boldsymbol{\Omega}_{0,i\Delta_n}\|_{\max} \leq C\tau_n \quad \text{and} \quad \max_i \|\widehat{\boldsymbol{\Omega}}_{i\Delta_n} - \boldsymbol{\Omega}_{0,i\Delta_n}\|_1 \leq C s_{\omega,p} \tau_n^{1-q}, \tag{3.6}$$

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

Remark 2. Theorems 1 and 2 show that by choosing $c = 1/2$, the estimators for the instantaneous coefficient and inverse matrix have element-wise convergence rates of $n^{-1/4}$ and ℓ_1 convergence rates $n^{-(1-\delta)/4}$ and $n^{-(1-q)/4}$, respectively, with the log order term and the sparsity level term. We note that when choosing the sub-interval length $k_n = c_k n^{1/2}$ to estimate the instantaneous processes, we have the same order convergence rates of the statistical estimation and time-varying instantaneous process approximation errors. That is, the order $n^{-1/4}$ is optimal for estimating each element of the instantaneous process; thus, the convergence rates are optimal up to log order.

The Dantzig instantaneous coefficient estimator has a near-optimal convergence rate as shown in Theorem 1. However, as discussed in the previous section, it is a biased estimator, which causes some non-negligible estimation errors when estimating the integrated coefficient. To tackle this problem, we employ debiasing schemes with the consistent CLIME estimator as in (3.4), and in the following theorem, we investigate its asymptotic benefits.

Theorem 3. *Under the assumptions in Theorems 1–2, we choose* $k_n = c_k n^{1/2}$ *for some constant* c_k. *Then, we have for all* i,

$$\widetilde{\boldsymbol{\beta}}_{i\Delta_n} - \boldsymbol{\beta}_{0,i\Delta_n} = \frac{1}{k_n \Delta_n} \boldsymbol{\Omega}_{0,i\Delta_n} \left(\mathcal{X}_i^\top \mathcal{Z}_i + \mathcal{A}_i \right) + R_i, \tag{3.7}$$

where $\mathcal{A}_i$ *is defined in* (A.9) *and*

$$\max_i \|R_i\|_{\max} \leq C \left\{ s_p^{2-\delta} (\log p / n^{1/2})^{(2-\delta)/2} + s_p s_{\omega,p} (\log p / n^{1/2})^{(2-q)/2} + s_p (\log p)^{3/2} / n^{1/2} \right\}, \tag{3.8}$$

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

$$\|\widehat{I\beta} - I\beta_0\|_{\max} \leq C b_n, \tag{3.9}$$

where $b_n = s_p^{2-\delta} (\log p / n^{1/2})^{(2-\delta)/2} + s_p s_{\omega,p} (\log p / n^{1/2})^{(2-q)/2} + s_p (\log p)^{3/2} / n^{1/2}$.

Remark 3. The debiased Dantzig instantaneous coefficient is decomposed by the martingale difference term $\mathcal{X}_i^\top \mathcal{Z}_i + \mathcal{A}_i$ and the non-martingale remaining term R_i. The martingale difference term can enjoy the law of large number property, so the integrated coefficient estimator has a faster convergence rate than the Dantzig instantaneous coefficient estimator. The remaining non-martingale terms have the same order as those of the martingale terms for the integrated coefficient estimator.

Unlike the biased Dantzig estimator, the non-martingale remaining terms do not impact on the integrated coefficient estimator.

Remark 4. Theorem 3 shows the element-wise convergence rate for the debiased Dantzig integrated coefficient. When we have the exact sparse coefficient and inverse matrix processes, that is, $\delta = q = 0$, the debiased Dantzig integrated coefficient estimator has the convergence rate $s_p(s_p + s_{\omega,p})(\log p)^{3/2}/n^{1/2}$. The $n^{1/2}$ term is related with the sample size, which is known as the optimal rate. The $(\log p)^{3/2}$ term comes from handling the high-dimensional error bound. Usually, in high-dimensional literature, we have $\sqrt{\log p}$, but the debiased Dantzig integrated coefficient estimator has $(\log p)^{3/2}$ due to the handling of the high-dimensional error bounds for estimating two coefficients, such as the instantaneous coefficient and the integrated coefficient, and bounding the random processes. Finally, the s_p and $s_{\omega,p}$ terms represent the sparsity levels for the coefficient and inverse volatility matrix. High-dimensional literature commonly assumes the sparsity level to be negligible; hence, we have the convergence rate $n^{-1/2}$ with up to $\log p$ order.

Theorem 3 indicates that, using the debiasing scheme, we obtain well-performing input integrated coefficient estimator $\widehat{\boldsymbol{\beta}}$. As described in Section 3.1, with the input integrated coefficient estimator $\widehat{\boldsymbol{\beta}}$, we apply the thresholding scheme to account for the sparsity and obtain the TED estimator. In the following theorem, we establish the ℓ_1 convergence rate of the TED estimator.

Theorem 4. *Under the assumptions in Theorems 1–2, let $k_n = c_k n^{1/2}$ for some constant c_k. For any given positive constant a, choose $h_n = C_{h,a} b_n$ for some constant $C_{h,a}$, where b_n is defined in Theorem 3. Then, we have, with probability greater than $1 - p^{-a}$,*

$$\|\widetilde{I\beta} - I\beta_0\|_1 \leq C s_p b_n^{1-\delta}. \tag{3.10}$$

Theorem 4 shows that the TED estimator is a consistent estimator in terms of the ℓ_1 norm under the sparsity condition (2.3). When estimating the integrated coefficient without the debiasing step, we can obtain the convergence rate $s_p(s_p\sqrt{\log p}n^{-1/4})^{1-\delta}$. The benefit of applying the debiasing scheme is the difference between b_n and $s_p\sqrt{\log p}n^{-1/4}$. Under the sparsity condition, b_n is $n^{-\{2-(\delta\vee q)\}/4}$ with $\log p$ order for $\delta, q \in [0,1)$, which is faster than the convergence rate of the Dantzig integrated coefficient estimator. Therefore, the TED estimator has the faster convergence rate.

3.3 Extension to jump diffusion processes

In financial practice, we often observe jumps. To reflect this, we can extend the continuous diffusion process (2.1) to the jump diffusion process as follows:

$$\begin{aligned} dY_t &= dY_t^c + dY_t^J, \\ dY_t^c &= \boldsymbol{\beta}_t^\top d\mathbf{X}_t^c + dZ_t, \quad \text{and} \quad dY_t^J = J_t^y d\Lambda_t^y, \end{aligned} \tag{3.11}$$

where Y_t^c and $\mathbf{X}_t^c$ are the continuous part of Y_t and $\mathbf{X}_t$, respectively, J_t^y is the jump size, and Λ_t^y is the Poisson process with the bounded intensity. The covariate process $\mathbf{X}_t$ is

$$d\mathbf{X}_t = d\mathbf{X}_t^c + d\mathbf{X}_t^J, \quad d\mathbf{X}_t^c = \boldsymbol{\mu}_t dt + \boldsymbol{\sigma}_t d\mathbf{B}_t, \quad \text{and} \quad d\mathbf{X}_t^J = \mathbf{J}_t d\boldsymbol{\Lambda}_t, \tag{3.12}$$

where $\mathbf{J}_t = (J_{1t}, \dots, J_{pt})^\top$ is a jump size process and $\boldsymbol{\Lambda}_t = (\Lambda_{1t}, \dots, \Lambda_{pt})^\top$ is a p-dimensional Poisson process with bounded intensities. We assume that the Poisson processes Λ_t^y and $\boldsymbol{\Lambda}_t$ are independent of $\boldsymbol{\sigma}_t$ and $\boldsymbol{\beta}_t$. Under this jump diffusion model, we can still use the proposed estimation procedure, but we cannot observe the continuous diffusion process. To tackle this problem, we first detect the jumps from the observed stock log-return data. For example, we use the truncation method as follows. Define

$$\widehat{\mathcal{Y}}_i^c = \begin{pmatrix} \Delta_{i+1}^n \widehat{Y}^c \\ \Delta_{i+2}^n \widehat{Y}^c \\ \vdots \\ \Delta_{i+k_n}^n \widehat{Y}^c \end{pmatrix} \quad \text{and} \quad \widehat{\mathcal{X}}_i^c = \begin{pmatrix} \Delta_{i+1}^n \widehat{\mathbf{X}}^{c\top} \\ \Delta_{i+2}^n \widehat{\mathbf{X}}^{c\top} \\ \vdots \\ \Delta_{i+k_n}^n \widehat{\mathbf{X}}^{c\top} \end{pmatrix}, \tag{3.13}$$

where $\mathbf{1}_{\{\cdot\}}$ is an indicator function, k_n is the number of observations in each window used to calculate the local regression,

$$\Delta_i^n \widehat{Y}^c = \Delta_i^n Y \, \mathbf{1}_{\{|\Delta_i^n Y| \le u_n\}}, \quad \Delta_i^n \widehat{\mathbf{X}}^c = \left(\Delta_i^n X_j \, \mathbf{1}_{\{|\Delta_i^n X_j| \le v_{j,n}\}} \right)_{j=1,\dots,p},$$

and u_n and $v_{j,n}$, $j = 1, \dots, p$, are the truncation levels. We employ $u_n = C_u s_p \sqrt{\log p} n^{-\varrho}$ and $v_{j,n} = C_{j,v} \sqrt{\log p} n^{-\varrho}$ for $\varrho \in [15/32, 1/2)$ and some constants C_u and $C_{j,v}$, $j = 1, \dots, p$. In the numerical study, we adopt the usual choice in the literature (Aït-Sahalia et al., 2020; Aït-Sahalia and Xiu, 2019). That is, we use

$$u_n = 3n^{-0.47} \sqrt{BV^Y} \quad \text{and} \quad v_{j,n} = 3n^{-0.47} \sqrt{BV_j}, \tag{3.14}$$

where the bipower variations $BV^Y = \frac{\pi}{2} \sum_{i=2}^n |\Delta_{i-1}^n Y| \cdot |\Delta_i^n Y|$ and $BV_j = \frac{\pi}{2} \sum_{i=2}^n |\Delta_{i-1}^n X_j| \cdot |\Delta_i^n X_j|$. Then, to estimate the integrated coefficient $I\beta$, we employ the estimation method in Section 3.1 using $\widehat{\mathcal{Y}}_i^c$ and $\widehat{\mathcal{X}}_i^c$ instead of $\mathcal{Y}_i$ and $\mathcal{X}_i$. We denote the jump-adjusted TED estimator by $\widetilde{I\beta}^c$. In the following theorem, we investigate the asymptotic property of the jump-adjusted TED estimator.

Theorem 5. *Under the models* (3.11)–(3.12)*, let assumptions in Theorem 4 hold. Then, we have, with probability greater than* $1 - p^{-a}$ *for any given positive constant* a*,*

$$\|\widetilde{I\beta}^c - I\beta_0\|_1 \le C s_p b_n^{1-\delta}. \tag{3.15}$$

Theorem 5 shows that the jump-adjusted TED estimator has the same convergence rate obtained in Theorem 4. Therefore, we conclude that the jumps can be detected well and that their effects can be mitigated.

3.4 Discussion on the tuning parameter selection

To implement the TED estimation procedure, we need to choose the tuning parameters. In this section, we discuss how to select the tuning parameters for the numerical studies. We first obtain $\Delta_i^n \widehat{Y}^c$ and $\Delta_i^n \widehat{\mathbf{X}}^c$ with the truncation levels defined in (3.14). Then, to handle the scale issue, we standardize each column of $\widehat{\mathcal{Y}}_i^c$ and $\widehat{\mathcal{X}}_i^c$ to have a mean of 0 and a variance of 1. The re-scaling is conducted after obtaining the TED estimator. For the local regression stage (3.2), we choose $k_n = [n^{1/2}]$. Also, we select

$$\lambda_n = c_\lambda n^{-1/4} (\log p)^{3/2}, \quad \tau_n = c_\tau n^{-1/4} \sqrt{\log p}, \quad \text{and} \quad h_n = c_h n^{-1/2} (\log p)^{3/2}, \tag{3.16}$$

where c_λ, c_τ, and c_h are tuning parameters. In the simulation and empirical studies, we choose $c_\lambda \in [0.1, 10]$ that minimizes the corresponding Bayesian information criterion (BIC). Also, we select $c_\tau \in [0.1, 10]$ by minimizing the following loss function:

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

where $\mathbf{I}_p$ is the p-dimensional identity matrix. Finally, we choose c_h by minimizing the corresponding mean squared prediction error (MSPE), and the result is $c_h = 0.5$. Details can be found in Section 5.

4 A simulation study

In this section, we conducted simulations to check the finite sample performance of the proposed TED estimator. We generated the data with frequency $1/n^{all}$ and considered the following time series regression jump diffusion model:

$$dY_t = \boldsymbol{\beta}_t^\top d\mathbf{X}_t^c + dZ_t + J_t^y d\Lambda_t^y,$$
$$d\mathbf{X}_t = d\mathbf{X}_t^c + d\mathbf{X}_t^J, \quad d\mathbf{X}_t^c = \boldsymbol{\sigma}_t d\mathbf{B}_t, \quad d\mathbf{X}_t^J = \mathbf{J}_t d\boldsymbol{\Lambda}_t, \quad dZ_t = \nu_t dW_t,$$

where $\mathbf{B}_t$ and W_t are p-dimensional and one-dimensional independent Brownian motions, respectively, $\mathbf{J}_t = (J_{1t}, \ldots, J_{pt})^\top$ and J_t^y are jump sizes, and $\boldsymbol{\Lambda}_t = (\Lambda_{1t}, \ldots, \Lambda_{pt})^\top$ and Λ_t^y are the Poisson processes with the intensities $(20, \ldots, 20)^\top$ and 15, respectively. The jump sizes J_{it} and J_t^y were independently generated from the Gaussian distribution with a mean of 0 and standard deviation of 0.05. We set the initial values X_{i0} and Y_0 to 0, while ν_t follows the Ornstein–Uhlenbeck process

$$d\nu_t = 3\left(0.12 - \nu_t\right) dt + 0.03 d\mathbf{W}_t^\nu,$$

where $\nu_0 = 0.15$ and $\mathbf{W}_t^\nu$ is one-dimensional independent Brownian motion. The instantaneous volatility process $\boldsymbol{\sigma}_t$ was taken to be a Cholesky decomposition of $\boldsymbol{\Sigma}_t = (\Sigma_{ijt})_{1 \leq i,j \leq p}$, where $\Sigma_{ijt} = \xi_t 0.8^{|i-j|}$ and ξ_t satisfies

$$d\xi_t = 5\left(0.3 - \xi_t\right) dt + 0.12 d\mathbf{W}_t^\xi,$$

where $\xi_0 = 0.5$ and $\mathbf{W}_t^\xi$ is one-dimensional independent Brownian motion. For the coefficient process $\boldsymbol{\beta}_t$, we considered the time-varying coefficient and constant coefficient processes, where $[s_p]$ factors are only significant. We first generated the time-varying coefficient process as follows:

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

where $\boldsymbol{\mu}_{\beta,t} = (\mu_{1,\beta,t}, \ldots, \mu_{p,\beta,t})^{\top}$, $\boldsymbol{\nu}_{\beta,t} = (\nu_{i,j,\beta,t})_{1 \leq i,j \leq p}$, and $\mathbf{W}_t^{\beta}$ is p-dimensional independent Brownian motion. We set the process $(\nu_{i,j,\beta,t})_{1 \leq i,j \leq [s_p]}$ as $\zeta_t \mathbf{I}_{[s_p]}$, where $\mathbf{I}_{[s_p]}$ is the $[s_p]$-dimensional identity matrix and ζ_t was generated as follows:

$$d\zeta_t = 3\,(0.5 - \zeta_t)\,dt + 0.2 d\mathbf{W}_t^{\zeta},$$

where $\zeta_0 = 0.4$ and $\mathbf{W}_t^{\zeta}$ is one-dimensional independent Brownian motion. For $i = 1, \ldots, [s_p]$, we took the initial value β_{i0} as 1 and $\mu_{i,\beta,t} = 0.05$ for $0 \leq t \leq 1$. We set β_{it}, $i = [s_p] + 1, \ldots, p$, as zero. In contrast, for the constant coefficient process, we set $\beta_{it} = 1$ for $i = 1, \ldots, [s_p]$ and $0 \leq t \leq 1$, while the other β_{it}'s were set to 0. We chose $p = 100$, $s_p = \log p$, $n^{all} = 4000$, and we varied n from 1000 to 4000. To implement the TED estimation procedure, we used the hard thresholding function $s\,(x) = x$ and employed the tuning parameter selection method discussed in Section 3.4.

For the purposes of comparison, we considered the integrated coefficient estimator proposed by Aït-Sahalia et al. (2020). We note that, for small p, one can account for the time variation of the coefficient process. We call this the AKX estimator. Specifically, the AKX estimator is calculated as follows:

$$\widehat{\boldsymbol{\beta}}_{i\Delta_n}^{\mathrm{AKX}} = (\widehat{\mathcal{X}}_i^{c\top} \widehat{\mathcal{X}}_i^{c})^{-1} \widehat{\mathcal{X}}_i^{c\top} \widehat{\mathcal{Y}}_i^{c} \quad \text{and} \quad \widetilde{I\beta}^{\mathrm{AKX}} = \sum_{i=0}^{[1/(K_n\Delta_n)]-1} \widehat{\boldsymbol{\beta}}_{iK_n\Delta_n}^{\mathrm{AKX}} K_n \Delta_n, \tag{4.1}$$

where $\widehat{\mathcal{X}}_i^{c}$ and $\widehat{\mathcal{Y}}_i^{c}$ are defined in (3.13) and we used $K_n = [n^{0.47}]$ instead of $k_n = [n^{0.5}]$. For $\widehat{\mathcal{X}}_i^{c\top} \widehat{\mathcal{X}}_i^{c}$ in (4.1), we added $10^{-4}\mathbf{I}_p$ to avoid the singularity coming from the ultra high-dimensionality. We also employed the LASSO estimator (Tibshirani, 1996), which is able to explain the sparsity of the high-dimensional coefficient process. However, the LASSO estimator is designed for the constant coefficient process; thus, it fails to account for the time-varying coefficient process. We estimated the LASSO estimator as follows:

$$\widetilde{I\beta}^{\text{LASSO}} = \operatorname{argmin}_{\boldsymbol{\beta}}\left\{\sum_{i=0}^{n-1}\left(\Delta_{i+1}^{n}\widehat{Y}^{c} - \Delta_{i+1}^{n}\widehat{\mathbf{X}}^{c\top}\boldsymbol{\beta}\right)^{2} + \lambda^{\text{LASSO}}\|\boldsymbol{\beta}\|_{1}\right\}, \tag{4.2}$$

where $\Delta_{i+1}^{n}\widehat{Y}^{c} = \Delta_{i+1}^{n}Y\,\mathbf{1}_{\{|\Delta_{i+1}^{n}Y|\le u_n\}}$, $\Delta_{i+1}^{n}\widehat{\mathbf{X}}^{c} = \left(\Delta_{i+1}^{n}X_j\,\mathbf{1}_{\{|\Delta_{i+1}^{n}X_j|\le v_{j,n}\}}\right)_{j=1,\ldots,p}$, and the regularization parameter $\lambda^{\text{LASSO}} \in [0.1, 10]$ was chosen by minimizing the corresponding Bayesian information criterion (BIC). We calculated the average estimation errors under the max norm, ℓ_1 norm, and ℓ_2 norm by 1000 simulation procedures.

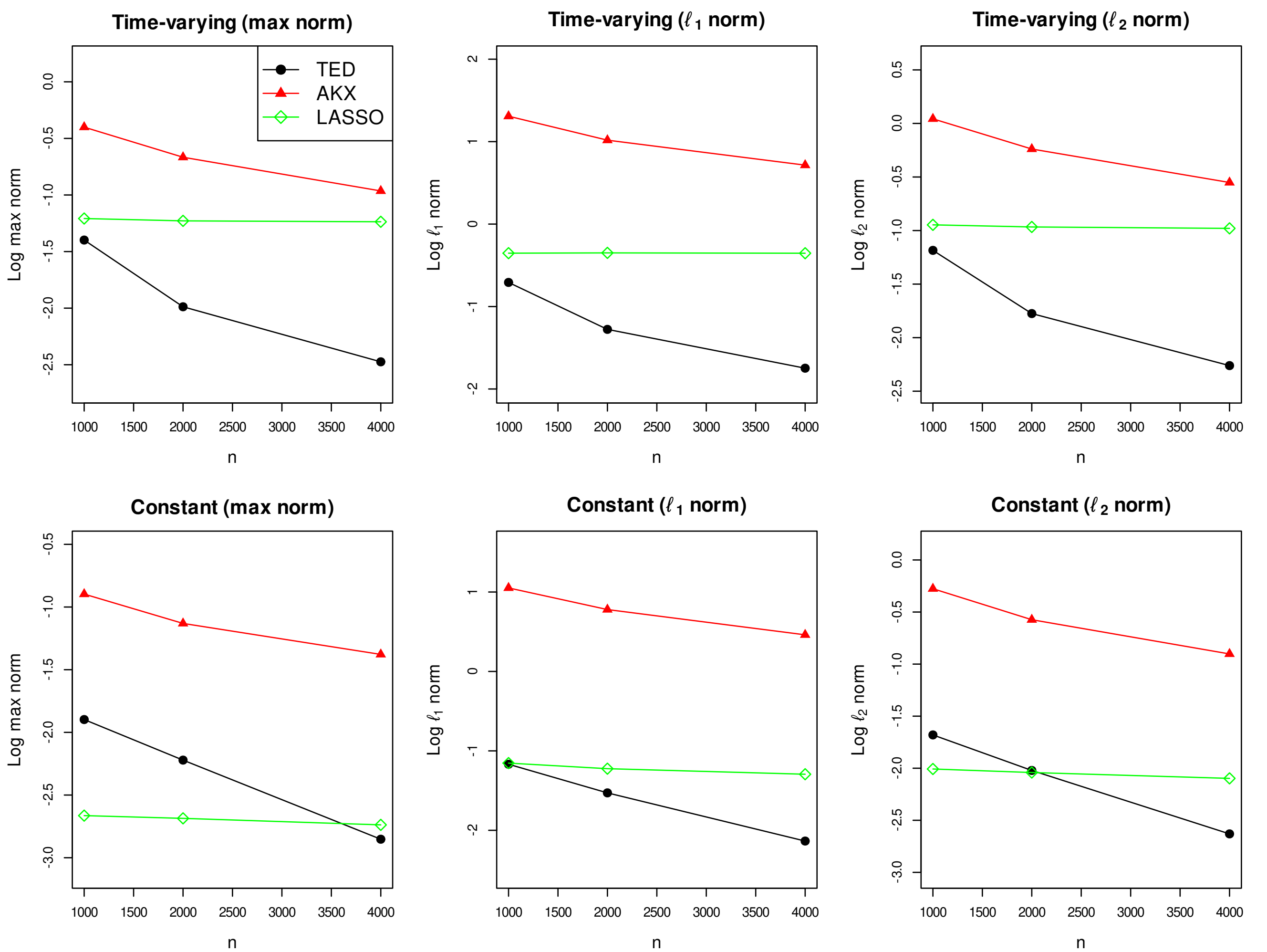

Figure 1: The log max, ℓ_1, and ℓ_2 norm error plots of the TED (black dot), AKX (red triangle), and LASSO (green diamond) estimators for $p = 100$ and $n = 1000, 2000, 4000$.

Figure 1 plots the log max, ℓ_1, and ℓ_2 norm errors of the TED, AKX, and LASSO estimators for the time-varying and constant coefficient processes with $p = 100$ and $n = 1000, 2000, 4000$. From Figure 1, we find that the estimation errors of the TED estimator are decreasing as the number of high-frequency observations increases. For the time-varying coefficient process, the TED estimator outperforms other estimators. This may be because the proposed TED estimation method can account for both time variation and the high-dimensionality of the coefficient process, while the AKX and LASSO estimators fail to explain one of them. When comparing the AKX and LASSO estimators, the LASSO estimator shows better performance. This may be because the errors from the curse of dimensionality are much more significant than those from the time-varying coefficient in this simulation study. For the constant coefficient process, the TED and LASSO estimators outperform the AKX estimator. This is probably due to the fact that only the AKX estimator is unable to handle the curse of dimensionality. We note that the error of the TED estimator decreases faster than that of the LASSO estimator. This may be because the model complexity of the TED estimator is relatively large. From these results, we can conjecture that the TED estimator accounts for the time variation and high-dimensionality of the coefficient process and is robust to the coefficient process structure.

5 An empirical study

We applied the proposed TED estimator to real high-frequency trading data from January 2013 to December 2019. We took stock price data from the End of Day website (https://eoddata.com/), firm fundamentals from the Center for Research in Security Prices (CRSP)/Compustat Merged Database, and futures price data from the FirstRate Data website. For each stock and futures, we obtained 5-min log-price data using the previous tick scheme (Wang and Zou, 2010; Zhang, 2011), where the half trading days were excluded. We considered the log-prices of the five assets as the dependent processes. Specifically, we selected Apple Inc. (AAPL), Berkshire Hathaway Inc.

(BRK.B), General Motors Company (GM), Alphabet Inc. (GOOG), and Exxon Mobil Corporation (XOM). These firms are the top market value stocks in five global industrial classification standards (GICS) sectors: information technology, financials, consumer discretionary, communication services, and energy sectors. For the covariate process, we collected the price data of 54 futures that represent market macro variables. For example, we selected 20 commodity data, 10 currency data, 10 interest rate data, and 14 stock market index data. We listed the symbols of 54 futures in Table B.1 in the supplementary materials. Furthermore, we considered Fama-French five factors in Fama and French (2015) and the momentum factor in Carhart (1997). We denoted market, value, size, profitability, investment, and momentum factors by MKT, HML, SMB, RMW, CMA, and MOM, respectively. We constructed these factors with high-frequency data similar to the scheme in Aït-Sahalia et al. (2020) as follows. First, we obtained the monthly portfolio constituents for above six factors with the stocks listed on NYSE, NASDAQ, and AMEX. Specifically, the MKT is the return of a value-weighted portfolio of whole assets, while the other factors are 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 were classified by the market equity, while high (H), medium (M), and low (L) portfolios were classified by their ratio of book equity to market equity. Also, we classified robust (R), neutral (N), and weak (W) portfolios according to their profitability, while conservative (C), neutral (N), and aggressive (A) portfolios were classified by their investment. Finally, we classified up (U), flat (F), and down (D) portfolios according to the momentum of the return. The details of this process can be found in Aït-Sahalia et al. (2020). Then, we calculated each portfolio return with a frequency of five minutes using the portfolio weights adjusted at a five-minute frequency. Specifically, we obtained the return of any portfolios, $WRet_{d,i}$, for the dth day and ith time interval 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 N_d is the number of stocks for the portfolio on the day d, the superscript j represents the jth stock of the portfolio, and $w_{d,i}^j$ is obtained 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 calculated using the close price of the jth stock on the day $d-1$, and $Ret_{d,0}^j$ is the overnight return from the $(d-1)$th day to the dth day. In sum, we utilized the five assets and 60 factors for the dependent processes and covariate processes, respectively.

For the choice of the tuning parameter c_h, we calculated the mean squared prediction error (MSPE) from the data in 2013. Specifically, we first defined

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

where $\widetilde{I\beta}^{m,j}(c_h)$ is the TED estimator obtained using the tuning parameter c_h and $\widehat{I\beta}^{m,j}$ is the debiased Dantzig integrated coefficient estimator for the mth month of 2013 and jth stock. Then, we selected c_h which minimizes $\Lambda(c_h)$ over $c_h \in \{l/10 \mid 0 \le l \le 5,\, l \in \mathbb{Z}\}$. The result is $c_h = 0.5$. We note that the stationarity assumption is reasonable for the coefficient process, which justifies the proposed tuning parameter choice procedure. Then, for each of the five assets, we employed the TED, AKX, and LASSO estimation procedures to obtain the monthly integrated coefficients.

The tuning parameters were selected based on Section 3.4 and Section 4. Since the AKX estimator is designed for the finite dimension, we also employed the AKX-SIX estimator. The AKX-SIX estimator employs the same estimation method as the AKX estimator except that it only uses MKT, HML, SMB, RMW, CMA, and MOM as factor candidates. We note that these six factors are commonly used in finance practice (Asness et al., 2013; Barroso and Santa-Clara, 2015; Carhart, 1997; Fama and French, 2015, 2016). For each estimation procedure, the coefficients for the non-trading period were estimated to be zero.

Table 1: The annual average in-sample and out-of-sample R^2 for the TED, AKX, AKX-SIX, and LASSO estimators across the five assets.

	In-sample R^2			
	Estimator			
	TED	AKX	AKX-SIX	LASSO
whole period	0.272	0.179	0.053	0.249
2013	0.237	0.163	0.038	0.232
2014	0.246	0.157	0.040	0.217
2015	0.305	0.220	0.067	0.286
2016	0.282	0.197	0.065	0.245
2017	0.211	0.086	0.017	0.180
2018	0.369	0.264	0.094	0.349
2019	0.256	0.170	0.047	0.236
	Out-of-sample R^2			
	Estimator			
	TED	AKX	AKX-SIX	LASSO
whole period	0.266	0.169	0.049	0.243
2014	0.239	0.144	0.034	0.211
2015	0.286	0.203	0.060	0.269
2016	0.267	0.190	0.063	0.240
2017	0.200	0.079	0.015	0.173
2018	0.353	0.234	0.069	0.341
2019	0.251	0.165	0.052	0.226

To investigate the performances of the TED, AKX, AKX-SIX, and LASSO estimators, we first calculated the monthly in-sample and out-of-sample R^2 from the monthly integrated coefficient estimates. We obtained the out-of-sample R^2 using the integrated coefficient estimates from the previous month. The out-of-sample R^2 was calculated from 2014 to 2019 since the tuning parameters were selected from the data in 2013. Then, we calculated the annual average R^2 over the five assets and twelve months. Table 1 reports the annual average in-sample and out-of-sample

R^2 for the TED, AKX, AKX-SIX, and LASSO estimators. From Table 1, we can find that the high-dimensional regression models (TED and LASSO) show better performance than the finite-dimensional regression model. This may be because as we know, the high-dimensional models can overcome the curse of dimensionality. When comparing TED and LASSO, the TED estimator shows the best result for all periods. This is probably due to the fact that only the TED estimator can account for both the high-dimensionality and time-varying property of the coefficient process.

Now, we investigate the TED estimation results. Figure 2 depicts the monthly integrated coefficient estimates for the five assets and 60 factors, and Figure 3 plots the nonzero frequency of monthly integrated coefficients for the five groups, such as the commodity futures group, currency futures group, interest rate futures group, stock market index futures group, and market factor group. From Figures 2–3, we find that the value of the integrated coefficient varies over time and the significant coefficients also change over time. Furthermore, the stock market index futures group had non-zero integrated coefficient estimates more often than the other futures groups. This may be because the stock market index futures can partially explain the market factors. This finding is consistent with the multi-factor models (Asness et al., 2013; Carhart, 1997; Fama and French, 1992, 2015). On the other hand, there are several individual factors that played a significant role in most periods. Thus, to investigate the coefficient behavior in greater details, we draw the integrated coefficients for the five assets and the three most frequent factors illustrated in Figure 4. For example, AAPL has NQ (E-mini Nasdaq-100), YM (E-mini Dow), and ES (E-mini S&P 500); BRK.B has MKT, ES, and YM; GM has MKT, MOM, and EW (E-mini S&P 500 Midcap); GOOG has NQ, BTP (Euro BTP Long-Bond), and ES; and XOM has MKT, RMW, and MOM. Among the factors, either the NQ factor or MKT factor is the most frequently significant factor. Moreover, the coefficient values of the three most frequent factors vary over time, while other factors are significant only for some periods. From these results, we can infer that the coefficient processes are sparse and time-varying. Hence, incorporating these features is important to account

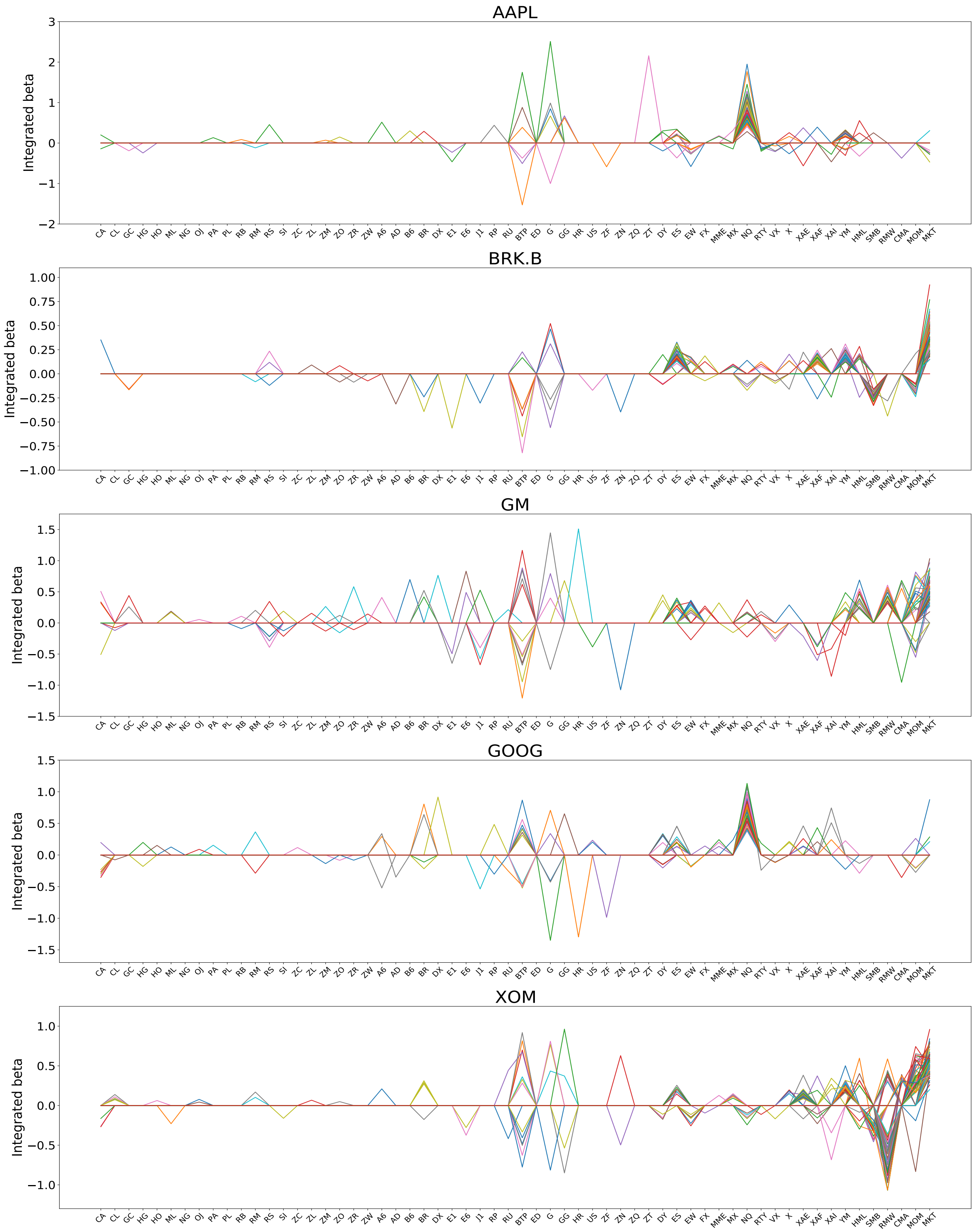

Figure 2: The monthly integrated coefficients from the TED estimation procedure for the five assets and 60 factors. Each line represents the integrated coefficient estimates for each month.

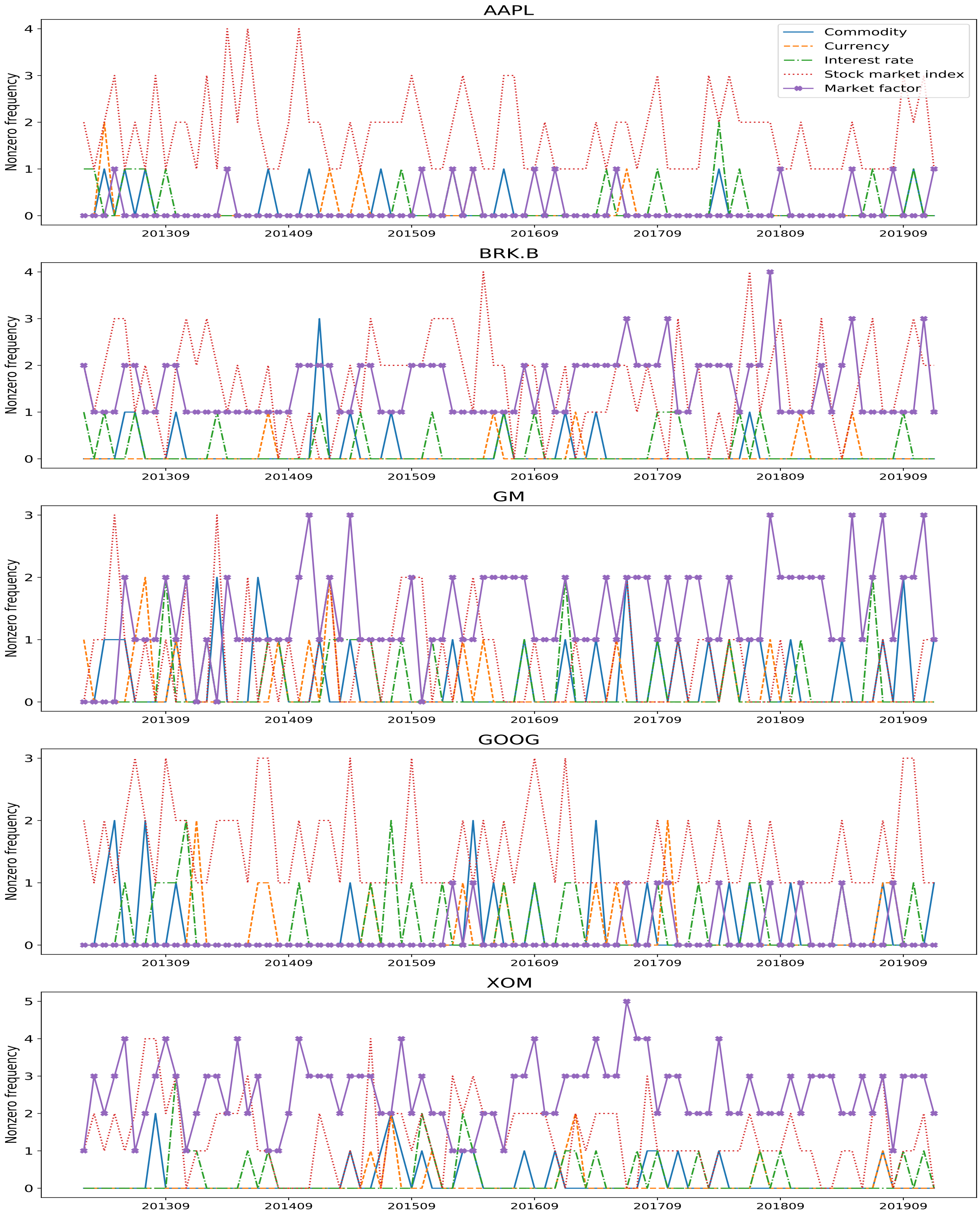

Figure 3: The nonzero frequency of the monthly integrated coefficients from the TED estimation procedure for the five assets and five groups. The five groups are the commodity futures group, currency futures group, interest rate futures group, stock market index futures group, and market factor group.

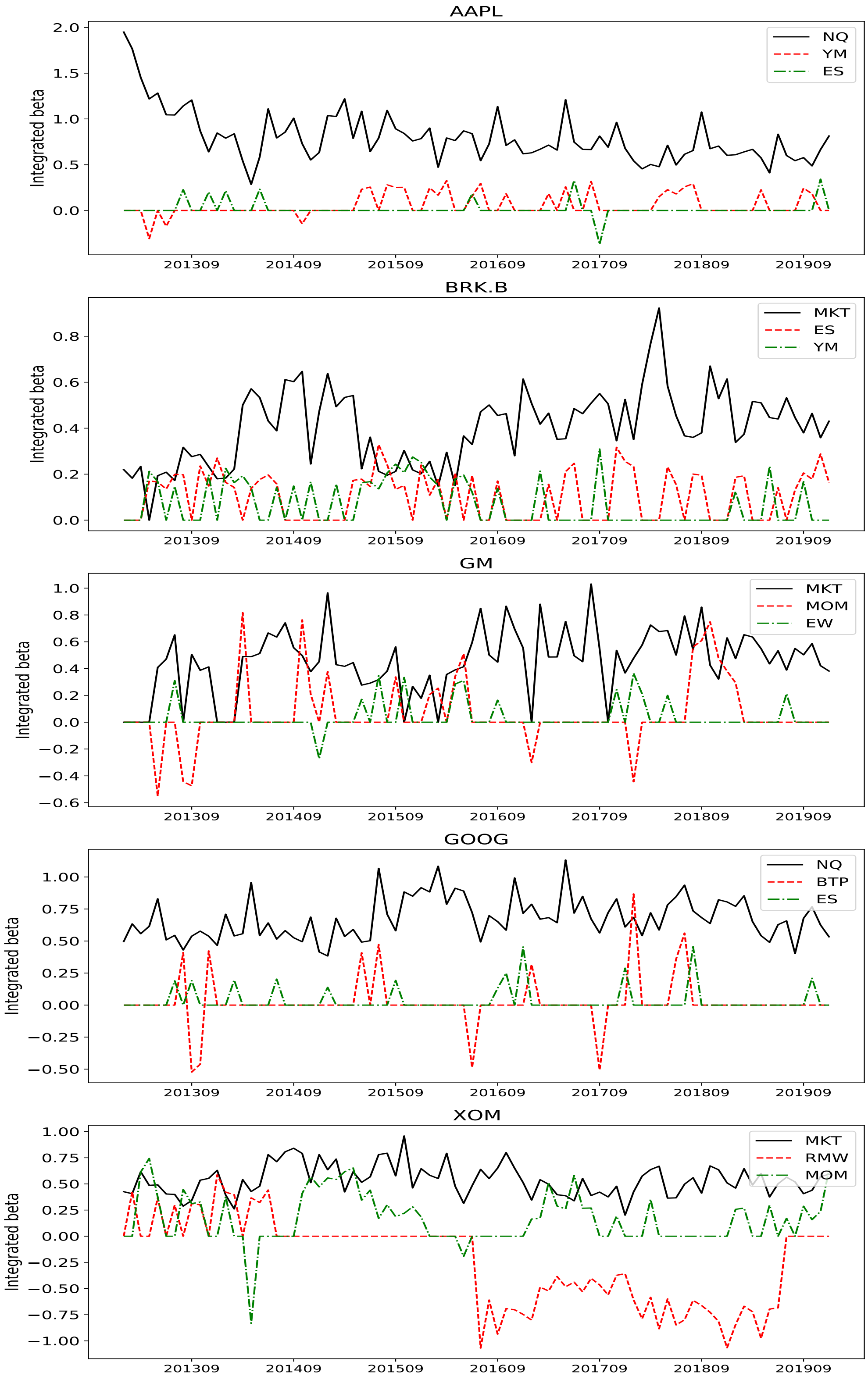

Figure 4: The integrated coefficients from the TED estimation procedure for the three most frequent factors among the 60 factors for each of the five assets.

for market dynamics. The proposed TED procedure can provide a good tool to deal with these issues when analyzing market dynamics using high-frequency data.

In finance practice, the six factors (Fama-French five factors and the momentum factor) are most frequently used (Asness et al., 2013; Barroso and Santa-Clara, 2015; Carhart, 1997; Fama and French, 2015, 2016). Thus, we investigate their integrated coefficient behaviors. Figure 5 depicts the estimates of the monthly integrated coefficient for MKT, HML, SMB, RMW, CMA, and MOM with their non-zero frequency. We find that the MKT factor was significant for BRK.B, GM, and XOM, which may indicate that these firms can be adequately explained by the market movements. Other market factors are also significant for some periods; thus, these factors can explain expected stock returns for BRK.B, GM, and XOM. In contrast, for technology companies such as AAPL and GOOG, the integrated coefficient estimates for the MKT factor are usually small. This may be because the NQ (E-mini Nasdaq-100) factor played a significant role for the technology stocks as shown in Figure 4. Furthermore, AAPL and GOOG cannot be satisfactorily explained using the common six factors. One possible explanation is that, over the last twenty years, the technology companies have led the U.S. economy, with AAPL and GOOG as the most successful companies in the same time frame. Thus, these six factors may not work well for the period when we studied them.

6 Conclusion

In this paper, we proposed a novel Thresholding dEbiased Dantzig (TED) estimation procedure which can accommodate the sparse and time-varying coefficient process in the high-dimensional set-up. Specifically, to account for the sparse and time-varying coefficient process, we applied the Dantzig procedure to the instantaneous coefficient estimator, which results in a biased estimator. To reduce the bias, we proposed a debiased estimation procedure. We estimated the integrated

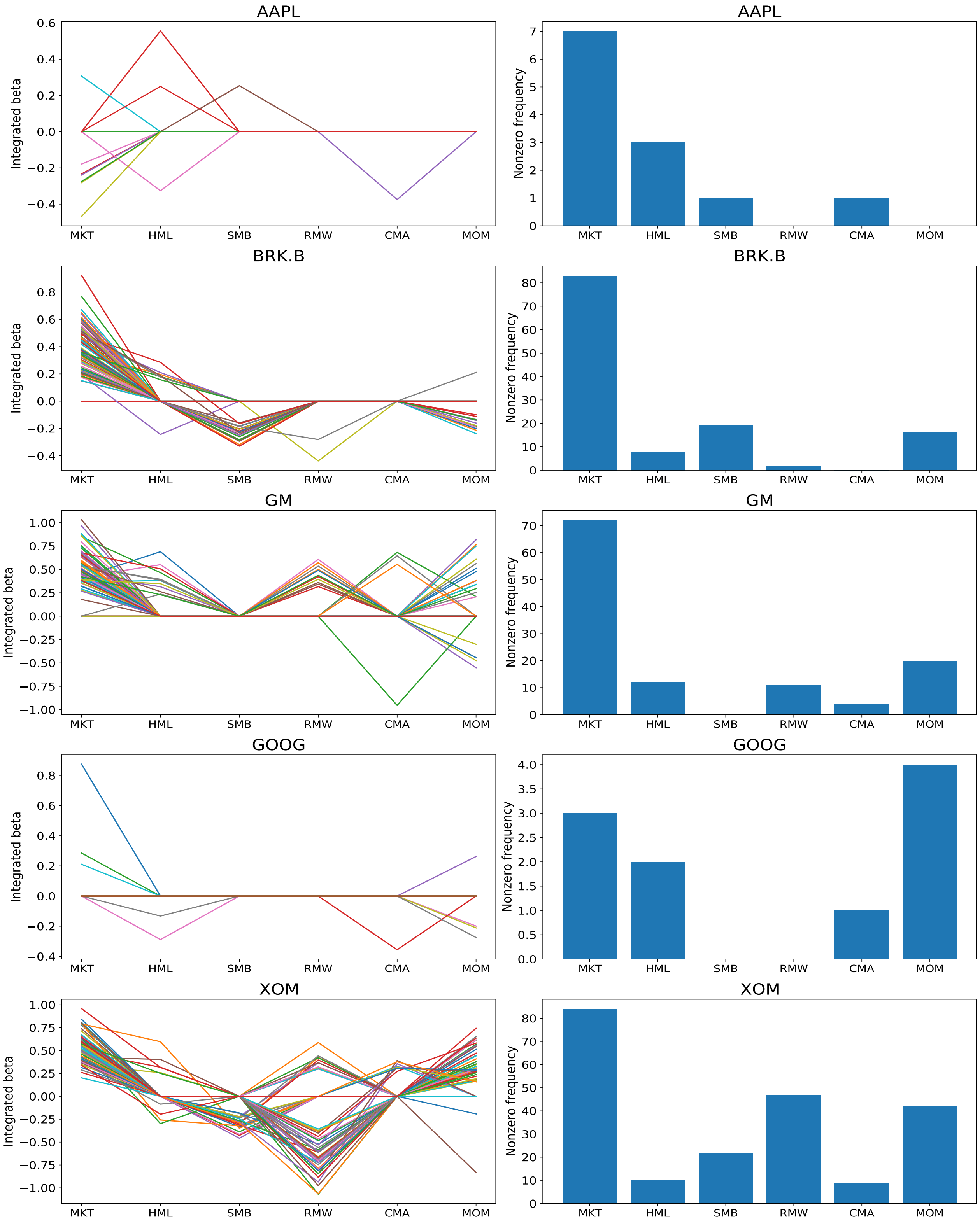

Figure 5: The integrated coefficient estimates from the TED estimation procedure (left) and the nonzero frequency (right) for the six factors, MKT, HML, SMB, RMW, CMA, and MOM. Each line (left) represents the integrated coefficient estimates for each month.

coefficient with this new debiased instantaneous coefficient estimator. We showed that the Dantzig procedure can handle the sparsity of the instantaneous coefficient and that the debiased scheme mitigates the errors from the bias of the instantaneous coefficient estimator. To accommodate the sparsity of the integrated coefficient, we further regularized the coefficient estimator. Finally, we showed that the proposed TED estimator can obtain the near-optimal convergence rate.

In the empirical study, the TED estimator outperforms other estimators in terms of both in-sample and out-of-sample R^2. Furthermore, we found that the coefficient process is sparse and time-varying. These findings revealed that, when analyzing the high-dimensional high-frequency regression, the TED estimator is a useful tool which can handle the curse of dimensionality and the time-varying coefficient. That is, in practice, the TED procedure makes it possible to analyze the stock market with relatively short period using high-frequency data.

References

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

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

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

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

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

Barroso, P. and Santa-Clara, P. (2015). Momentum has its moments. *Journal of Financial Economics*, 116(1):111–120.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 81(2):608–650.

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

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

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

Chen, D., Mykland, P. A., and Zhang, L. (2023). Realized regression with asynchronous and noisy high frequency and high dimensional data. *Journal of Econometrics*.

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

Ciolek, G., Marushkevych, D., and Podolskij, M. (2022). On lasso estimator for the drift function in diffusion models. *arXiv preprint arXiv:2209.05974*.

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

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

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

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

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

Feng, G., Giglio, S., and Xiu, D. (2020). Taming the factor zoo: A test of new factors. *The Journal of Finance*, 75(3):1327–1370.

Gaïffas, S. and Matulewicz, G. (2019). Sparse inference of the drift of a high-dimensional ornstein–uhlenbeck process. *Journal of Multivariate Analysis*, 169:1–20.

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

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

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

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

Kim, D., Kong, X.-B., Li, C.-X., and Wang, Y. (2018). Adaptive thresholding for large volatility matrix estimation based on high-frequency financial data. *Journal of Econometrics*, 203(1):69–79.

Kim, D., Wang, Y., and Zou, J. (2016). Asymptotic theory for large volatility matrix estimation based on high-frequency financial data. *Stochastic Processes and their Applications*, 126:3527—3577.

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

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

Negahban, S. N., Ravikumar, P., Wainwright, M. J., and Yu, B. (2012). A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers. *Statistical science*, 27(4):538–557.

Oh, M., Kim, D., and Wang, Y. (2022). Dynamic realized beta models using robust realized integrated beta estimator. *arXiv preprint arXiv:2204.06914*.

Tao, M., Wang, Y., and Zhou, H. H. (2013). Optimal sparse volatility matrix estimation for high-dimensional itô processes with measurement errors. *The Annals of Statistics*, 41(4):1816–1864.

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

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

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

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67.

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

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

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429.