Realized Regression with Asynchronous and Noisy High Frequency and High Dimensional Data*

Dachuan Chen — Nankai University

Per A. Mykland — The University of Chicago

Lan Zhang — University of Illinois at Chicago

Thursday 17th November, 2022

Abstract

We develop regression for high frequency data. This regression is novel in that it can be for both fixed and increasing dimension. Also, the data may have microstructure noise, and observations (trades, or quotes) can be asynchronous, (*i.e.,* the observations do not need to be synchronized across dimensions). As is customary for high-frequency inference methods, we refer to our method as "realized" regression.

In our methodology, spot beta becomes a key quantity in the nonparametric framework of high frequency econometrics. The central contribution of this paper is a feasible estimator of spot beta, which is robust to noise and asynchronicity. With the help of the spot-version of the Smoothed TSRV estimator, spot beta can be consistently estimated. There are two direct applications of the spot beta estimates in the current paper. In the first application, the integrated beta can be consistently estimated by aggregating the spot beta estimates. After a bias-correction procedure, a fixed dimension central limit theorem is established for the bias-corrected estimator, with convergence rate which may be arbitrarily close to $O_p(n^{-1/4})$. In the second application we assume time-varying factor structure and conditional sparsity. The spot beta matrix estimator enables the estimation of high dimensional spot covariance and precision matrices. The latter is obtained by thresholding the spot residual covariance estimates, and convergence rates derived. As an empirical application, this paper explores the hourly change in beta around earnings announcements of the S&P 100 constituents.

Key Words: Asynchronous sampling times, Factor model, High dimensionality, High frequency, Market microstructure noise, Realized regression, Spot beta, Integrated beta, Spot covariance and precision matrices.

*Financial support from the National Science Foundation under grants DMS-2015530 (Zhang), and DMS-2015544 (Mykland), is gratefully acknowledged.

1. Introduction

Regression is a main technique in scientific research, which is widely used in exploring the linear relationship between observable quantities, and in analyzing the structure of variability.

The connection between regression and finance originated from the capital asset pricing model (CAPM, Markowitz (1952, 1959), Sharpe (1964), Lintner (1965), Black (1972)). Over time, the connection has expanded to multiple factors, such as in Fama and MacBeth (1973), and Ross (1976). The literature has gradually split into regression (observed factors) and principal component analysis (PCA, unobserved factors). We are here concerned with the former. For literature reviews, see, *e.g.,* Campbell et al. (1997) and Cochrane (2005). Recent developments in high frequency PCA are reviewed in Chen et al. (2020).

The importance of time-varying betas (regression coefficients) has received increasing attention in the finance and econometrics literature. Such betas reflect time-varying conditional information. Research in this direction includes Hansen and Richard (1987), Bollerslev et al. (1988), Jagannathan and Wang (1996), Boguth et al. (2011), Ang and Kristensen (2012), Engle (2016), and Gagliardini et al. (2016).

WIth the advent of high-frequency data, a literature has started to develop where time-varying betas are estimated from intraday data. Important empirical contributions are Andersen et al. (2006), who investigated the persistence and predictability of time-varying beta estimates, and Patton and Verardo (2012), who explored the effect of information flows on stock returns.

The purpose of this paper is to develop the theory for how to estimate betas in fixed and increasing dimension, for high frequency data. If we let $c_t^{\mathbf{X},\mathbf{X}}$ and $c_t^{\mathbf{X},\mathbf{Y}}$ be the (unobserved) spot (instantaneous) covariance matrices of (latent) efficient prices (or other semi-martingales) $\mathbf{X}$ and $\mathbf{Y}$, the spot and integrated beta are given by[1]

$$\boldsymbol{\beta}_t = \left(c_t^{\mathbf{X},\mathbf{X}}\right)^{-1} c_t^{\mathbf{X},\mathbf{Y}} \text{ and } \int_0^{\mathcal{T}} \beta_t dt, \tag{1.1}$$

where $[0, \mathcal{T}]$ is the fixed interval under observation. By considering data with microstructure noise, as well as letting observations (such as transactions and quotes) happen asyncronously across dimensions, we bring the theory to the point where it can accommodate real data.

[1] *Cf.* the development leading to eq. (3.4), below, as well as B_t in 4.10. Here, Y is a scalar process, and X is a q-dimensional process, where q can be fixed, or tend in infinity with increasing data density.

In finite dimension (Section 3), our theory focuses on the integrated beta. *The integrated beta $\int_0^{\mathcal{T}} \beta_t dt$ is consistently estimated by aggregating estimates of spot beta.* The aggregation is similar to the constructions in the papers cited at the beginning of Section 1.1 We show asymptotic normality in finite dimension (Theorem 2), preceded by a bias correction which is needed for this normality to hold.

In increasing dimension (Section 4), our theory estimates the spot (instantaneous) β_t,[2] and from there estimates the spot precision matrix, which has a role in determining asset allocation, *cf.* Fan et al. (2016a). We derive the rate of convergence as the dimensions of X and Y tend to infinity.

Both these developments take as their points of departure spot covariance matrices that are calculated by the S-TSRV procedure (pre-averaging followed by two-scales, Section 2 in this paper, and Mykland et al. (2019)). The basic pre-averaging is done over time blocks of length $\Delta\tau_n$, and spot covariance matrices are calculated over time blocks of length ΔT_n. To get a sense of the magnitudes we have in mind, in the simulation we have used $\Delta T_n = 2340$ seconds, and $\Delta\tau_n$ is 5, 15 or 60 seconds. In the empirical application, $\Delta\tau_n = 5$ seconds, and ΔT_n is (in most cases) hourly.

On the theoretical side, the rate of convergence in the CLT (Theorem 2) is a_n^{-1}, which is allowed to be arbitrarily close to $n^{-1/4}$. The latter is previously known as the standard efficient rate for covariances in estimation problems with microstructure.[3] A precise explanation of the rate a_n is given in eq. (2.10) and Remark 1 in Section 2.2. As described there, a_n is closely related to $\Delta\tau_n$.

In the increasing dimension setting, the rates of convergence also depend crucially on a_n, but we defer discussion of this to Section 4.

1.1. Sketch of finite dimensional regression

Closely related literature. The theory of estimation the two betas in (1.1) has previously been studied in the case of no microstructure noise and synchronous observations, in Mykland and Zhang (2006, 2008), and Zhang (2012), with a jump-robust version in Aït-Sahalia et al. (2020) and Aït-Sahalia et al. (2021).

[2] There called $\mathbf{B}_t$ to emphasize that it is a matrix.

[3] Going back to Jacod and Protter (1998), Engle (2000), Andersen et al. (2001), Barndorff-Nielsen and Shephard (2002), Zhang et al. (2005), Jacod et al. (2009) and others. Recent contributions include Bibinger and Mykland (2016), Bibinger et al. (2017), and Mykland et al. (2019). An interesting variant over the this estimation problem involves using factor structure to estimate higher dimensional covariance (co-volatility), and relevant literature is discussed in connection with increasing dimension in Section 1.2.

In this setting, the estimator of integrated beta is simply a sum of ordinary least squares regression estimators (Mykland and Zhang, 2009, Section 4.2, pp.1424-1426). More generally, all proposed estimators of (1.1) are local in time, so that covariance at time t is only compared with variance around t. This is also the case for the estimators developed in the current paper.

In the presence of asynchronous and noisy observations, the development of a feasible spot beta estimator has become increasingly necessary. As shown by Monte Carlo simulation (Table 5.1 in Section 5.2), integrated beta estimates become biased when the data is noisy. By applying the spot-version of the smoothed TSRV (S-TSRV), this paper proposes feasible estimators for spot beta under both fixed and increasing dimension.

Bias in the integrated beta. Expanding the Riemann sum of spot beta estimates to higher order, a bias term naturally arises, which is analogous to the aggregated second order expansion term of the non-linear functional of stochastic volatilities in Jacod and Rosenbaum (2013) and Aït-Sahalia and Xiu (2017). This bias term becomes the main barrier to the central limit theorem. By properly selecting the range of the smoothing window ΔT_n over which the spot β is calculated, and then applying the extended bias-correction technique based on Chen et al. (2020), the central limit theorem (CLT) for the bias-corrected estimator (Theorem 2) follows.

An earlier approach to the assessment of integrated beta is to estimate

$$\mathcal{T}\left(\int_0^{\mathcal{T}} c_t^{\mathbf{X},\mathbf{X}} dt\right)^{-1} \int_0^{\mathcal{T}} c_t^{\mathbf{X},Y} dt. \tag{1.2}$$

The theory for the estimation (1.2) would seem to go back to Barndorff-Nielsen and Shephard (2004), and natural estimators were considered empirically by Andersen et al. (2006) and Patton and Verardo (2012). The *advantage* of this formulation is that it permits results for covariance (co-volatility) matrices to be directly extended to the estimation of integrated β. This reduces the problem to one that has been given substantial consideration in the literature, and for which there are now already results that cover noise and asynchronicity. (See Footnote 3.)

A main *disadvantage* of estimating (1.2) is that natural estimators are not local in time: if the time interval is a day, then, for example, covariance at 10:45 am is compared with variance at 3:20 pm.

Notwithstanding the distinction between (1.1) and (1.2), the two quantities are similar if the time span $\mathcal{T}$ is comparatively short. They are also the same if β_t is constant in t. Constancy tests for betas have been proposed by Todorov and Bollerslev (2010), Kalnina (2012), Reiß et al. (2015) and Kong and Liu (2018).

We also point out that the estimators in the current paper are based on the assumption that the latent semi-martingales are continuous. This is substantially more complex for the case where there is microstructure noise and asynchronous observation, and we hope to approach this topic in a later paper.

1.2. Sketch of high (increasing) dimensional regression

When estimating a high dimensional spot (cross-sectional) covariance matrix, the rank of the estimated matrix is bounded by $2\Delta T_n/\Delta\tau_n + b$,[4] by construction. This is a severe constraint, even more so than when estimating an integrated matrix. It is thus possible that the rank of the true spot covariance matrix may grow much faster than the given bound.

To resolve such a contradiction, the main approach in the literature is to rely on sparsity. Our high dimensional realized regression makes use of a time-varying (observed) factor model, where we treshold the residual based on sparsity. This goes back to Bickel and Levina (2008). Our development of a the large spot precision matrix estimator may be regarded as the "realized" and spot (high-frequency) version of Fan et al. (2011).

An estimation theory for high dimensional high frequency *integrated* covariance matrices has been derived with blockwise-diagonal residual covariance structure in Fan et al. (2016a), which was further improved by considering the asynchronous and noisy observations in Dai et al. (2019). In both these papers, the factor loadings are assumed to be time-invariant, which is unlike in the current paper.

1.3. Empirical application

As an application in Section 6, we use high-frequency beta estimation to study the variation of stock betas on earnings announcement days. It is well known in the literature that stock betas tend to be higher around the event days. For example, Ball and Kothari (1991) documented an increase in daily

[4] Here ΔT_n and $\Delta\tau_n$ are as described above, and b is a very slowly growing number, *cf.* eq. 2.7 and Remark 1 in Section 2.2.

average beta during the three-day earnings announcement period. Vijh (1994) found that after being added to S&P 500 index between 1975 and 1989, those stocks displayed higher market beta at daily and weekly frequency. More recently, Patton and Verardo (2012) estimated daily variations in betas around earnings announcements for all the S&P 500 constituent stocks over the period 1996-2006. They found that the beta increase on announcement day was short-lived and it reverted to average levels two to five days later.

We investigate hourly beta variation within 5 days of the earnings announcement. Our study follows the spirit of Patton and Verardo (2012). While the earlier paper uses daily betas, our current technology permits us to find hourly betas, and thus to understand intra-day variation as well as overnight change in beta. Also, the construction of the beta estimate differs. Patton and Verardo (2012) used 25-minute intra-day returns (plus the overnight return from the previous day) to construct daily beta estimates. As the authors mentioned, they used the 25-minute sampling interval to reduce the impact of microstructure noise but at the cost of the accuracy of the estimate.

In the current paper, we construct beta estimates from 5-second pre-averaged returns of S&P 100 constituent stocks from 2007 to 2017, while taking account of the microstructure noise and the cross-sectional asynchronicity. Our hourly betas are unbiased and consistent, thus can more precisely capture the beta dynamics in a shorter time window around the announcements. With the definition of "Day 0" as the calendar day of each earnings announcement, we are able to separate the before- and post-market announcement impact on beta change. When the earnings are released in the morning prior to market open on "Day 0", we observe substantial beta jump in the first hour (i.e. 10am). On the other hand, when the earnings are announced after market close (4pm), we notice a significant beta jump the following day, again at the first hour. Within the 5-day window (from "Day -2" to "Day +2"), most hourly beta stays at the non-earnings level.

1.4. Organization and Notation

This paper is organized as follows: we first set up the general data structure and define the spot-version of the Smoothed TSRV (S-TSRV, Mykland et al. (2019)) estimator in Section 2. For fixed dimension, consistency and asymptotic normality are shown theoretically in Sections **??**-3, and for high dimension, consistency is shown in Section 4. The results are corroborated by Monte Carlo simulation in Section 5.2. Section 7 conducts an empirical study that applies our methodology to the cross-sectional intraday returns of the components of S&P 100 Index.

For a matrix $\mathbf{A}_{p\times q}$, $(\mathbf{A})_{k,\bullet}$ denotes its $k-$th row, $(\mathbf{A})_{\bullet,r}$ denotes its $r-$th column, $\mathbf{A}^{(r,k)}$ denotes its $(r,k)-$th element, $d\mathbf{A}_t = \left\{d\mathbf{A}_t^{(r,k)}\right\}_{1\le r\le p,1\le k\le q}$ and $\mathbf{A}^\intercal$ denotes its transpose. We denote the largest and smallest eigenvalue of matrix $\mathbf{A}$ by $\lambda_{\max}(\mathbf{A})$ and $\lambda_{\min}(\mathbf{A})$, respectively. We denote by $\|\mathbf{A}\|, \|\mathbf{A}\|_1, \|\mathbf{A}\|_{\mathrm{F}}, \|\mathbf{A}\|_{\max}$ the spectral norm, L_1-norm, Frobenius norm and elementwise max norm of matrix $\mathbf{A}$, defined as $\|\mathbf{A}\| = \lambda_{\max}^{1/2}(\mathbf{A}^\intercal\mathbf{A}), \|\mathbf{A}\|_1 = \max_j \Sigma_i \left|\mathbf{A}^{(i,j)}\right|, \|\mathbf{A}\|_{\mathrm{F}} = \mathrm{tr}^{1/2}(\mathbf{A}^\intercal\mathbf{A}), \|\mathbf{A}\|_{\max} = \max_{i,j}\left|\mathbf{A}^{(i,j)}\right|$. If $\mathbf{A}$ is a vector, then $\|\mathbf{A}\|$ and $\|\mathbf{A}\|_{\mathrm{F}}$ are equal to its Euclidean norm. For two sequences, we write $x_n \asymp y_n$ if $x_n = O_p(y_n)$ and $y_n = O_p(x_n)$.

A number of processes, such as the martingale the martingale M, is fully indexed as $M_{n,t}^{(r,s)}$, where the superscript (r,s) refers to matrix element, and the subscript t refers to time, $t\in[0,\mathcal{T}]$, and n is an index referring to the number of observations. In order to not overburden the paper with super- and subscripts, we do on occasion omit one or several of these. (i) $M_{n,t}$ is a matrix martingale. Further notation in this direction is introduced in Section 3. (ii) Meanwhile, we introduce dependence on n when we gradually get close to asymptotics in eq. (2.11)-(2.12), and therefore also in the definition (2.8). However, one should bear in mind that every ingredient in (2.8) depends on sample size n, with the single exception of the latent process (2.1)-(2.2). (iii) In certain equations, such as in Remark 3, the time variable t is omitted in the subscript of the martingale $M_{n,t}^{(r,s)}$, because the the quadratic variation $[\cdot,\cdot]_t$ is an operation on the entire path of the martingale, and t is conventionally moved to become a subscript of the quadratic variation instead. Note in particular that M_∞ (with possibly further indices) always refers to a limit when n has gone to infinity. This is because time t is always finite ($\le \mathcal{T}$). – Similar considerations apply to other stochastic variables and processes in the following.

2. Basic Setup

2.1. Data Description

We here provide a description of the data generating process, as well as assumptions that we make on these processes.

The latent process. For two positive integers $q, d \geq 1$, we work with data discretely sampled from the continuous process

$$(\mathbf{\Xi}_t)_{0\leq t\leq \mathcal{T}} = \left(\underbrace{\Xi_t^{(1)}, \ldots, \Xi_t^{(q)}}_{\text{covariate process } X}, \underbrace{\Xi_t^{(q+1)}, \ldots, \Xi_t^{(q+d)}}_{\text{dependent variable process } Y}\right)_{0\leq t\leq \mathcal{T}}. \tag{2.1}$$

The separation of Ξ_t into an X_t and a Y_t process is irrelevant in this section, which is concerned with the estimation of the covariance (volatility) matrix process for Ξ_t, but it plays a rôle when studying regression in subsequent sections.

We assume that the (Ξ_t) process is a $(q+d)$-dimensional continuous Itô process, *i.e.,* of the following form

$$\mathbf{\Xi}_t = \mathbf{\Xi}_0 + \int_0^t \mu_u du + \int_0^t \sigma_u dW_u, \tag{2.2}$$

where W is a $(q+d)$-dimensional standard $(\mathcal{F}_t)_{0\leq t\leq \mathcal{T}}$-Brownian motion, and X_0 is $\mathcal{F}_0$-measurable. The coefficients μ_u and σ_u are predictable and

$$\mu_t \text{ and } c_t \text{ are locally bounded in } \|\cdot\|_{\max}\text{-norm}, \tag{2.3}$$

where we use

$$c_t = \left(\sigma\sigma^\intercal\right)_t. \tag{2.4}$$

Thus, the integrated covariance matrix of $\mathbf{\Xi}_t$ may be expressed as:

$$\langle \mathbf{\Xi}, \mathbf{\Xi}\rangle_t = \int_0^t c_u du. \tag{2.5}$$

The volatility matrix. We also suppose that $c_t^{(r,s)}$ is itself an Itô process for any $1 \leq r, s \leq q+d$. In other words, it has the same structure as described above, but is a matrix and not a vector.

The Observed Process. For $1 \leq r \leq q+d$, the process $\left(\Xi_t^{(r)}\right)_{0 \leq t \leq \mathcal{T}}$ is observed on the grid $\mathcal{G}^{(r)} = \left\{0 = t_0^{(r)} < t_1^{(r)} < \cdots < t_{n^{(r)}}^{(r)} = \mathcal{T}\right\}$, after contamination by microstructure noise $\epsilon_{t_j^{(r)}}^{(r)}$. This yields an observed process $\mathbf{\Xi}^* = \left(\Xi^{*,(1)}, \ldots, \Xi^{*,(q)}, \Xi^{*,(q+1)}, \ldots, \Xi^{*,(q+d)}\right)$, as follows:

$$\Xi_{t_j^{(r)}}^{*,(r)} = \Xi_{t_j^{(r)}}^{(r)} + \epsilon_{t_j^{(r)}}^{(r)}, \text{ for } 1 \leq r \leq q+d.$$

Our assumptions on the data are summarized as follows:

Condition 1. (Structure of the data.) The data generating process and the observations are as laid out in Section 2.1. The processes Ξ_t, μ_t and σ_t are adapted to a filtration $(\mathcal{F}_t)$. The observation times $t_{n,j}$ are $(\mathcal{F}_t)$-stopping times. For each (n,j), the noise $\epsilon_{n,t_{n,j}}$ is $\mathcal{F}_{t_{n,j}}$-measurable, and $\sup_{n,j} E\epsilon_{n,t_{n,j}}^2 < \infty$, and $E\epsilon_{n,t_{n,j}} = 0$. The signal $\mathbf{\Xi}_t$ may not depend on n.

2.2. *Estimator for the Integrated Covariance Matrix: The S-TSRV and its Decomposition.*

In order to estimate the integrated covariance matrix $\langle \mathbf{\Xi}, \mathbf{\Xi} \rangle_t$, we construct the smoothed TSRV (S-TSRV) estimator $\widehat{\langle \mathbf{\Xi}, \mathbf{\Xi} \rangle}_t$ on a synchronous grid, as follows.

$$\{0 = \tau_{n,0} < \tau_{n,1} < \cdots < \tau_{n,N} = \mathcal{T}\}. \tag{2.6}$$

Denote $\mathcal{M}_{n,i}^{(r)} = \#\left\{j : \tau_{n,i-1} < t_j^{(r)} \leq \tau_{n,i}\right\}$.

For $0 \leq t \leq \mathcal{T}$, $1 \leq r,s \leq q+d$ and a pair (J,K), set

$$K\left[\widetilde{\tilde{\Xi}^{(r)}, \tilde{\Xi}^{(s)}}\right]_t^{(K)} = \left(\frac{1}{2}\sum_{i=1}^{b-K} + \sum_{i=b-K+1}^{N^*(t)-b} + \frac{1}{2}\sum_{i=N^*(t)-b+1}^{N^*(t)-K}\right)\left(\tilde{\Xi}_{i+K}^{(r)} - \tilde{\Xi}_i^{(r)}\right)\left(\tilde{\Xi}_{i+K}^{(s)} - \tilde{\Xi}_i^{(s)}\right),$$

where

$$N^*(t) = \max\{1 \leq i \leq N : \tau_{n,i} \leq t\} \text{ and } b = K + J, \tag{2.7}$$

and where, $1 \leq i \leq N$ and $1 \leq r \leq q+d$, the pre-averaged price is defined as:

$$\tilde{\Xi}_i^{(r)} = \frac{1}{\mathcal{M}_{n,i}^{(r)}} \sum_{\tau_{n,i-1} < t_j^{(r)} \leq \tau_{n,i}} \Xi_{t_j^{(r)}}^{*,(r)}.$$

We similarly define $J\left[\widetilde{\tilde{\Xi}^{(r)},\tilde{\Xi}^{(s)}}\right]^{(J)}$ by switching J and K.

The Smoothed-TSRV is defined as:

$$\langle\widehat{\Xi^{(r)},\Xi^{(s)}}\rangle_{n,t}=\frac{1}{(1-b/N)(K-J)}\left\{K\left[\widetilde{\tilde{\Xi}^{(r)},\tilde{\Xi}^{(s)}}\right]_t^{(K)}-J\left[\widetilde{\tilde{\Xi}^{(r)},\tilde{\Xi}^{(s)}}\right]_t^{(J)}\right\}. \tag{2.8}$$

We assume the following about the block structure (imposed by the econometrician) and its interface with the data.

Condition 2. (Structure of Blocks.) We assume that the block separation times $\tau_{n,i}$ are $(\mathcal{F}_t)$-stopping times that are "exogenous" (independent of the Ξ-process), and that for each n, there are nonrandom $\Delta\tau_n^+$ and $\mathcal{M}_n^-\geq 1$, so that $\Delta\tau_n^+\geq\max_i\Delta\tau_{n,i}$ and $\mathcal{M}_n^-\leq\min_i\mathcal{M}_{n,i}$. Assume that as $n\to\infty$, $\Delta\tau_n^+\propto\mathcal{M}_n^-/n$, in which case the number of blocks $N=N_n$ is of exact order $O\left(n/\mathcal{M}_n^-\right)$. Also assume that $K_n\Delta\tau_n^+\to 0$ as $n\to\infty$, and that $K_n>J_n\geq 1$. Finally suppose that $K_n-J_n=O_p\left(\left(N_n/\mathcal{M}_n^-\right)^{2/3}\right)$, and that

$$N_n/\mathcal{M}_n^-\to\infty. \tag{2.9}$$

See Remark 1 below for some clarification of Condition 2.

Condition 3. (Assumption on the interface between noise and blocks, and on averaged noise) We suppose that $E(\bar{\epsilon}_{n,i}\mid\mathcal{F}_{\tau_{i-J}})=0$, and that $E\sup_i E(\bar{\epsilon}_{n,i}^2\mid\mathcal{F}_{\tau_{i-J}})=o_p(\Delta\tau_n^+(K-J)^{1/2})$. Also let $\bar{\epsilon}_{n,i}=\bar{\epsilon}_i$ be the averaged noise across the block from $\tau_{n,i-1}$ to $\tau_{n,i}$. Assume that the $\epsilon_{n,t_{n,j}}$ process is stationary, exponentially α mixing, and that there is a constant $\kappa>0$ so that $E\epsilon_{n,t_{n,j}}^{4+\kappa}<\infty$.[5]

Define the sequence $\{a_n\}_{n\geq 1}$ by

$$a_n=\left[\left(K_n-J_n\right)\Delta\tau_n^+\right]^{\frac{1}{2}}, \tag{2.10}$$

[5]Condition 3 is one of several ways to to assure $\mathrm{Cov}\left(\bar{\epsilon}_i^{(s_1)},\bar{\epsilon}_i^{(s_2)}\right)=\left(\mathcal{M}_n^-\right)^{-1}\varsigma^{(s_1,s_2)}$ and $\sup_i\mathrm{cum}_4\left(\bar{\epsilon}_i^{(s_1)},\bar{\epsilon}_i^{(s_1)},\bar{\epsilon}_i^{(s_2)},\bar{\epsilon}_i^{(s_2)}\right)=O_p\left(\left(\mathcal{M}_n^-\right)^{-2}\right)$ as $n\to\infty$, *cf.* McLeish (1975), Hall and Heyde (1980, Chapter 5 and Appendix 3), Aït-Sahalia et al. (2011), Zhang (2011), Mykland et al. (2019, Condition 4 and the subsequent discussion on p. 109), and Chen et al. (2020, Assumption 2, p. 1963). For the relationship to the latter, observe that since $E(\bar{\epsilon}_{n,i})=0$, the fourth cumulant $\mathrm{cum}_4\left(\bar{\epsilon}_i^{(s_1)},\bar{\epsilon}_i^{(s_1)},\bar{\epsilon}_i^{(s_2)},\bar{\epsilon}_i^{(s_2)}\right)=\mathrm{Var}\left(\bar{\epsilon}_i^{(s_1)}\bar{\epsilon}_i^{(s_2)}\right)$.

and note that $a_n \to 0$ as the number of observations $n \to \infty$ by Condition 2. Under Condition 1-3, it follows from Mykland et al. (2019, Section 5, pp. 110-111)[6] that

$$\langle \widehat{\Xi^{(r)}, \Xi^{(s)}} \rangle_{n,t} = \int_0^t c_u^{(r,s)} du + M_{n,t}^{(r,s)} + o_p(a_n), \tag{2.11}$$

where $c^{(r,s)}$ is the $(r,s)-$th element of c from (2.4), and there the $M_{n,t}/a_n$ converges stably in law to a continuous martingale limit.

Remark 1. (The meaning and size of K_n, J_n, and a_n.) We here explain that the order of convergence a_n can be up to $n^{-1/4}$, but that this rate cannot be attained within the development of this paper. To see this, return to Condition 2, and consider the simplified case where $\mathcal{M}_{n,i}$ only depends on n, *i.e.*, $\mathcal{M}_{n,i} = \mathcal{M}_n$. In this case, $K_n - J_n = O_p\left((N_n/\mathcal{M}_n^-)^{2/3}\right)$ is desirable since it assures an optimal tradeoff between statistical error due to signal and to noise (Mykland et al., 2019, end of Section 5, p. 111). The same discussion shows that if eq. (2.9) were removed from Condition 2, one might choose N_n and $\mathcal{M}_n$ to be of exact order $O(n^{1/2})$, and K_n and J_n would be finite. In this case, a_n is of exact order $n^{-1/4}$. However, assumption (2.9) is necessary for the representations (2.13)-(2.15), cf. Chen et al. (2020, Appendix A). We believe that it is possible to create an asymptotic development that does not require (2.9), since the finite sample calculations in Mykland et al. (2019) remain valid in this case, but this is beyond the scope of this paper. Meanwhile, the current paper should be read with the understanding that a_n is almost $n^{-1/4}$, and that K_n and J_n are approximately finite (they grow arbitrarily slowly).

Remark 2. The selection of the tuning parameters ("scales") K_n and J_n is an area which remains more art than science. For low dimensional problems, one can proceed through signature plots on estimated volatilities, introduced by Andersen et al. (2000) and their co-authors. Signature plot was used to determine K_n and J_n in multiple dimensions in (Zhang, 2011, Fig. 2, p. 42). For moderate dimension regression problems, one option is the signature plot of integrated beta, as in Fig. 5.2 in Section 5. For truly high dimensional problems, an attractive approach is to use signature plots on eigenvalues (Chen et al., 2020, Fig. 2, p. 13). We have not gone into this detail in this paper, but Figure 5.3 (also Section 5) plots the spectral norm (also an eigenvalue) of the error of the final precision matrix estimator (the

[6] Here and below, the effect of the drift term is negligible, cf. Mykland and Zhang (2009, Section 2.2, pp. 1407-1409).

red curve in the plot). Finally, note that for the S-TSRV, the scales are expected to be approximately finite (Remark 1 above), while for the original TSRV (Zhang et al. (2005)), especially K_n will grow with sample size n.

We shall need a slightly sharper representation under the same assumptions. For $1 \leq r \leq q+d$, define $\Delta\Xi^{(r)}_{\tau_i} = \Xi^{(r)}_{\tau_i} - \Xi^{(r)}_{\tau_{i-1}}$, the estimation error can be written as follows:

$$\widehat{\langle \Xi^{(r)}, \Xi^{(s)} \rangle}_{n,t} - \int_0^t c_u^{(r,s)} du = M_t^{(r,s)} + \tilde{e}_t^{(r,s)} - e_0^{(r,s)}, \tag{2.12}$$

where the subscript n has been omitted on the right hand side, and below until eqn. (2.15), for notational convenience, and where the martingale term may be expressed as:

$$M_t^{(r,s)} = M_t^{X,(r,s)} + M_t^{\epsilon,(r,s)} + o_p\left(a_n\right), \text{ where} \tag{2.13}$$

$$\begin{aligned}
M_t^{X,(r,s)} &= \sum_{p=1}^{K-J-1} \left(\frac{K-J-p}{K-J}\right) \sum_{i=J+p+1}^{N^*(t)} \Delta\Xi^{(r)}_{\tau_{i-p}} \Delta\Xi^{(s)}_{\tau_i}[2], \\
M_t^{\epsilon,(r,s)} &= \frac{1}{K-J} \sum_{i=K+1}^{N^*(t)} \left(\bar{\epsilon}^{(r)}_{i-J} - \bar{\epsilon}^{(r)}_{i-K}\right) \bar{\epsilon}^{(s)}_i[2],
\end{aligned}$$

and the edge effect terms $e_0^{(r,s)}$ and $\tilde{e}_t^{(r,s)}$ has the order of $O_p\left(a_n^2\right)$, which may be further expressed as:

$$\begin{aligned}
e_0^{(r,s)} &= \frac{1}{K-J} \sum_{i=J+1}^{K} \bar{\epsilon}^{(r)}_{i-J} \bar{\epsilon}^{(s)}_i[2] + \sum_{p=1}^{K-J-1} \sum_{i=1}^{K-J-p} \left(\frac{K-J-p-i}{K-J}\right) \Delta\Xi^{(r)}_{\tau_{J+i}} \Delta\Xi^{(s)}_{\tau_{J+i+p}}[2] \\
&+ \sum_{i=1}^{K-J} \left(\frac{K-J-i}{K-J}\right) \Delta\Xi^{(r)}_{\tau_{J+i}} \Delta\Xi^{(s)}_{\tau_{J+i}} + o_p\left(a_n^2\right), \text{ and}
\end{aligned} \tag{2.14}$$

$$\begin{aligned}
\tilde{e}_t^{(r,s)} &= -\frac{1}{K-J} \sum_{i=J}^{K-1} \bar{\epsilon}^{(r)}_{N^*(t)-i-J} \bar{\epsilon}^{(s)}_{N^*(t)-i}[2] - \sum_{p=1}^{K-J-1} \sum_{i=0}^{K-J-p} \left(\frac{K-J-p-i}{K-J}\right) \Delta\Xi^{(r)}_{\tau_{N^*(t)-i-p}} \Delta\Xi^{(s)}_{\tau_{N^*(t)-i}}[2] \\
&- \sum_{i=0}^{K-J} \left(\frac{K-J-i}{K-J}\right) \Delta\Xi^{(r)}_{\tau_{N^*(t)-i}} \Delta\Xi^{(s)}_{\tau_{N^*(t)-i}} + o_p\left(a_n^2\right).
\end{aligned} \tag{2.15}$$

The representation and rates in (2.13)-(2.15) follow from Chen et al. (2020, Appendix A).

Remark 3. (Assumption on Asymptotic Covariance) Let $M_{n,t}^{(r,s)}$ be as defined in (2.12) and (2.13). Since the above development guarantees that $a_n^{-1}M_n^{(r,s)}$, $1 \le r, s \le q+d$, converge jointly in law (as continuous martingales) to a limit $M_\infty^{(r,s)}$. Following Jacod and Shiryaev (2003, Corollary 6.30, p. 385), it is then also the case that for the optional ("observed") quadratic variations,

$$a_n^{-2}\left[M_n^{(r_1,s_1)}, M_n^{(r_2,s_2)}\right]_t \xrightarrow{p} \left[M_\infty^{(r_1,s_1)}, M_\infty^{(r_2,s_2)}\right]_t, \text{ for } 1 \le r_1, s_1, r_2, s_2 \le q+d \text{ and } 0 \le t \le \mathcal{T}.$$

2.3. *The estimation of the Spot Volatility Matrix*

For the simplicity of discussion, we define the spot volatility estimator $\hat{c}_{\Delta T_n,t}^{(r,s)}$ for some $\Delta T_n > 0$ as follows:

$$\hat{c}_{\Delta T_n,t}^{(r,s)} = \frac{1}{\Delta T_n}\left(\widehat{\langle \Xi^{(r)}, \Xi^{(s)} \rangle}_{t+\Delta T_n} - \widehat{\langle \Xi^{(r)}, \Xi^{(s)} \rangle}_t\right), \tag{2.16}$$

where $\{\Delta T_n\}_{n\ge 1}$ is a sequence of positive numbers satisfying

$$a_n^{-2}\Delta T_n \to \infty \text{ and } \Delta T_n \to 0 \text{ as } n \to \infty. \tag{2.17}$$

Moreover, to facilitate the theory development, we define

$$\bar{c}_{\Delta T_n,t}^{(r,s)} = \frac{1}{\Delta T_n}\int_t^{t+\Delta T_n} c_u^{(r,s)} du,\ \bar{\pi}_{\Delta T_n,t}^{(r,s)} = \bar{c}_{\Delta T_n,t}^{(r,s)} - c_t^{(r,s)} \text{ and } \tilde{\pi}_{\Delta T_n,t}^{(r,s)} = \hat{c}_{\Delta T_n,t}^{(r,s)} - \bar{c}_{\Delta T_n,t}^{(r,s)}, \tag{2.18}$$

and

$$\breve{\pi}_{\Delta T_n,t}^{(r,s)} = \frac{1}{\Delta T_n}\left(M_{t+\Delta T_n}^{(r,s)} - M_t^{(r,s)}\right) \text{ and } \eta_{\Delta T_n,t}^{(r_1,s_1,r_2,s_2)} = \tilde{\pi}_{\Delta T_n,t}^{(r_1,s_1)}\tilde{\pi}_{\Delta T_n,t}^{(r_2,s_2)} - \breve{\pi}_{\Delta T_n,t}^{(r_1,s_1)}\breve{\pi}_{\Delta T_n,t}^{(r_2,s_2)}. \tag{2.19}$$

We now list several useful results of spot volatility estimator.

Lemma 1. *Assume Conditions 1-3, as well as Condition (2.17). Then we have: (i)*

$$\sup_{t,r_1,r_2,s_1,s_2}\left|E\left(\breve{\pi}_{\Delta T_n,t}^{(r_1,s_1)}\breve{\pi}_{\Delta T_n,t}^{(r_2,s_2)}\right)\right| = O_p\left(a_n^2\Delta T_n^{-1}\right), \tag{2.20}$$

and

$$\sup_{t,r_1,r_2,s_1,s_2}\left\|\breve{\pi}_{\Delta T_n,t}^{(r_1,s_1)}\breve{\pi}_{\Delta T_n,t}^{(r_2,s_2)}-E\left(\breve{\pi}_{\Delta T_n,t}^{(r_1,s_1)}\breve{\pi}_{\Delta T_n,t}^{(r_2,s_2)}\right)\right\|_2=O_p\left(a_n^2\Delta T_n^{-1}\right). \tag{2.21}$$

(ii)

$$\sup_{t,r_1,r_2,s_1,s_2}\left|E\left(\eta_{\Delta T_n,t}^{(r_1,s_1,r_2,s_2)}\right)\right|=O_p\left(a_n^4\Delta T_n^{-2}\right), \tag{2.22}$$

and

$$\sup_{t,r_1,r_2,s_1,s_2}\left\|\eta_{\Delta T_n,t}^{(r_1,s_1,r_2,s_2)}-E\left(\eta_{\Delta T_n,t}^{(r_1,s_1,r_2,s_2)}\right)\right\|_2=O_p\left(a_n^3\Delta T_n^{-3/2}\right). \tag{2.23}$$

Proof. The proof of this lemma is collected in Appendix A. □

3. Multiple Regression

In multiple regression, it is possible that $q,d>1$ in the definition (2.1) of $(\mathbf{\Xi}_t)_{0\le t\le\mathcal{T}}$. Without loss of generality, we denote $\mathbf{X}=\left(X^{(1)},\ldots,X^{(q)}\right)=\left(\Xi^{(1)},\ldots,\Xi^{(q)}\right)$ and we let Y be a single process, so that $Y=\Xi^{(q+l)}$ for some $1\le l\le d$. It is natural to use the following notations: $\langle\mathbf{X},\mathbf{X}\rangle_t=\left\{\left\langle\Xi^{(r)},\Xi^{(s)}\right\rangle_t\right\}_{1\le r,s\le q}$, $\langle\mathbf{X},Y\rangle_t=\left\{\left\langle\Xi^{(r)},\Xi^{(q+l)}\right\rangle_t\right\}_{1\le r\le q}$, $\widehat{\langle\mathbf{X},\mathbf{X}\rangle}_t=\left\{\widehat{\left\langle\Xi^{(r)},\Xi^{(s)}\right\rangle}_t\right\}_{1\le r,s\le q}$, and $\widehat{\langle\mathbf{X},Y\rangle}_t=\left\{\widehat{\left\langle\Xi^{(r)},\Xi^{(q+l)}\right\rangle}_t\right\}_{1\le r\le q}$.

For the convenience of notation, we define

$$\underbrace{c_t^{\mathbf{X},\mathbf{X}}=\left\{c_t^{(r,s)}\right\}_{1\le r,s\le q}}_{q\times q\text{ matrix process}}\quad\text{and}\quad\underbrace{c_t^{\mathbf{X},Y}=\left\{c_t^{(r,q+1)}\right\}_{1\le r\le q}}_{q\times 1\text{ column vector process}}. \tag{3.1}$$

We analogously define the related matrix and vector quantities for $M,\bar{c},\hat{c},\bar{\pi},\tilde{\pi},\breve{\pi},\check{\varphi},\tilde{e}$ an e.

Suppose that the processes are related by

$$dY_t=\sum_{k=1}^{q}\beta_t^{(k)}dX_t^{(k)}+dZ_t\text{ with }\left\langle X^{(k)},Z\right\rangle_t=0\text{ for all }t\text{ and }k. \tag{3.2}$$

If we assume that $\boldsymbol{\beta}=\left(\beta^{(1)},\ldots,\beta^{(q)}\right)$ is a $q\times 1$ column vector process, then the quadratic variation of the residual process may be expressed as:

$$\begin{aligned}\langle Z,Z\rangle_t &= \langle Y,Y\rangle_t - 2\int_0^t d\langle \mathbf{X},Y\rangle_s \boldsymbol{\beta}_s + \int_0^t \boldsymbol{\beta}_s^\intercal d\langle \mathbf{X},\mathbf{X}\rangle_s \boldsymbol{\beta}_s \\ &= \langle Y,Y\rangle_t - 2\int_0^t \boldsymbol{\beta}_s^\intercal c_s^{\mathbf{X},Y} ds + \int_0^t \boldsymbol{\beta}_s^\intercal c_s^{\mathbf{X},\mathbf{X}} \boldsymbol{\beta}_s ds. \end{aligned} \tag{3.3}$$

To find $\min_{\boldsymbol{\beta}} \langle Z,Z\rangle_{\mathcal{T}}$, and assuming $c_s^{\mathbf{X},\mathbf{X}}$ is positive definite almost surely for all $0 \le t \le \mathcal{T}$, we solve the identity $-2c_s^{\mathbf{X},Y} + 2c_s^{\mathbf{X},\mathbf{X}}\boldsymbol{\beta}_s = 0$, and finally obtain the unique solution as follows:

$$\boldsymbol{\beta}_s = \left(c_s^{\mathbf{X},\mathbf{X}}\right)^{-1} c_s^{\mathbf{X},Y}. \tag{3.4}$$

The spot beta estimator is naturally constructed as:

$$\hat{\boldsymbol{\beta}}_{\Delta T_n,T_{i-1}} = \left(\hat{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1} \hat{c}^{\mathbf{X},Y}_{\Delta T_n,T_{i-1}}. \tag{3.5}$$

The quantity in which we are interested is:

$$\boldsymbol{\theta} = \int_0^{\mathcal{T}} \boldsymbol{\beta}_t dt,$$

and its estimator is given by[7]:

$$\hat{\boldsymbol{\theta}}_n = \sum_{i=1}^{B} \hat{\boldsymbol{\beta}}_{\Delta T_n,T_{i-1}} \Delta T_n.$$

We first show the consistency of $\hat{\boldsymbol{\theta}}_n$. For the simplicity of discussion, we define an intermediate process:

$$\bar{\boldsymbol{\beta}}_{\Delta T_n,T_{i-1}} = \left(\bar{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1} \bar{c}^{\mathbf{X},Y}_{\Delta T_n,T_{i-1}}. \tag{3.6}$$

With this smoothed beta, the estimation error of $\hat{\boldsymbol{\beta}}_{\Delta T_n,T_{i-1}}$ can also be decomposed into two parts,

[7] Cf. Mykland and Zhang (2009, Section 4.2, pp. 1424-1428), Zhang (2012, Section 4).

Moreover, the estimation error may be decomposed as follows:

$$\hat{\theta}_n - \theta = \underbrace{\sum_{i=1}^{B} \left(\hat{\beta}_{\Delta T_n, T_{i-1}} - \beta_{T_{i-1}} \right) \Delta T_n}_{\text{Aggregated error of } \hat{\beta}_{\Delta T_n, T_{i-1}},\ R^{\text{Spot}}} - \underbrace{\sum_{i=1}^{B} \int_{T_{i-1}}^{T_i} \left(\beta_s - \beta_{T_{i-1}} \right) ds}_{\text{Discretization error, } R^{\text{Discrete}}} . \tag{3.7}$$

Then we can show the representations of these two types of estimation error. There representations matter both in the proofs, and also in Section 3.1.

We presently state the consistency of spot beta estimator $\hat{\boldsymbol{\beta}}_{\Delta T_n, T_{i-1}}$. For this, we need an additional assumption about spot covariance matrix.

Condition 4. There are constants $\vartheta_1, \vartheta_2 > 0$ such that $\inf_{0 \le t \le \mathcal{T}} \lambda_{\min} \left(c_t^{\mathbf{X}, \mathbf{X}} \right) > \vartheta_1$ and $\sup_{0 \le t \le \mathcal{T}} \| c_t \|_{\max} < \vartheta_2$ almost surely.

Condition 4 can, obviously, be localized just as in 2.3, cf. Jacod and Protter (2012, Chapter 4.4.1, pp. 114-121) and Mykland and Zhang (2012, Chapter 2.4.5, pp. 160-161).

Lemma 2. *(Consistency of $\hat{\boldsymbol{\theta}}_n$) Assume Conditions 1-4. Assume that the number of regressors q is finite, and ΔT_n satisfies condition (2.17). Then, for any ϵ, $0 < \epsilon < 1/2$, we have:*

$$\sup_i \left\| \hat{\boldsymbol{\beta}}_{\Delta T_n, T_{i-1}} - \boldsymbol{\beta}_{T_{i-1}} \right\| = O_p \left(\Delta T_n^{1/2 - \varepsilon} \right) + O_p \left(\left(a_n^2 \Delta T_n^{-1} \right)^{1/2 - \varepsilon} \right) = o_p(1),$$

and

$$\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} = o_p(1).$$

Proof. The proof of this lemma is collected in the Appendix B. □

3.1. Asymptiotic Bias of the naïve Regression Estimator

When $\Delta T_n \to 0$ and $\inf_n a_n^{-1} \Delta T_n > 0$, the discretization error R^{Discrete} (eqn (3.7)) becomes the dominating term in the estimation error of $\hat{\theta}_n$. However, in this scenario, it cannot achieve the optimal convergence rate. Consequently, we consider the setting of $a_n^{-1} \Delta T_n \to 0$ and $a_n^{-2} \Delta T_n \to \infty$. In this scenario, the aggregated error of $\hat{\beta}_{\Delta T_n, T_{i-1}}$, R^{Spot} becomes the dominating term. By further analyzing the aggregated error R^{Spot}, it is easy to show that there is a bias term arises in R^{Spot}, which has bigger size than the martingale term. In the following theorem, we provide the representation of the bias term so that we can design the bias-corrected estimator in the subsequent subsection. For ease of exposition, this result is stated in the simple regression case only.

Theorem 1. *(Second order behavior of $\hat{\theta}_n$ in the univariate case.) Assume that $q = d = 1$, as well as Conditions 1-4. and also that $a_n^{-1}\Delta T_n \to 0$ and $a_n^{-2}\Delta T_n \to \infty$. Then we have:*

$$a_n^{-2}\Delta T_n \left(\hat{\theta}_n - \theta\right) \xrightarrow{p} -\varphi_{\mathcal{T}},$$

where

$$\varphi_t = \int_0^t \left(c_u^{X,X}\right)^{-2} \left(d\left[M_\infty^{X,X}, M_\infty^{X,Y}\right]_u - \beta_u d\left[M_\infty^{X,X}, M_\infty^{X,X}\right]_u\right).$$

Proof. The proof of this theorem is collected in the Appendix C. □

3.2. Bias corrected estimator and CLT for multiple regression

Similar to the single regressor case, the size of bias term is bigger than the martingale term when $a_n^{-1}\Delta T_n \to 0$. Thus, in order to develop the CLT, we need to contruct the bias corrected estimator. For $1 \le r, s \le q + d$, we define

$$\check{\varphi}^{(r,s)}_{\Delta T_n, T_{n,i-1}} = \frac{1}{2}\left(\hat{c}^{(r,s)}_{\Delta T_n/2,(i-1/2)\Delta T_n} - \hat{c}^{(r,s)}_{\Delta T_n/2,(i-1)\Delta T_n}\right). \tag{3.8}$$

The bias corrected estimator is defined as:

$$\tilde{\boldsymbol{\theta}}_n = \sum_{i=1}^{B}\left[\hat{\boldsymbol{\beta}}_{\Delta T_n, T_{i-1}} + \left(\hat{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n, T_{i-1}}\right)^{-1}\left(\hat{\phi}^{\mathbf{X},\mathbf{X},\mathbf{X},Y}_{\Delta T_n, T_{i-1}} - \hat{\phi}^{\mathbf{X},\mathbf{X},\mathbf{X},\mathbf{X}}_{\Delta T_n, T_{i-1}}\hat{\boldsymbol{\beta}}_{\Delta T_n, T_{i-1}}\right)\right]\Delta T_n, \tag{3.9}$$

where

$$\hat{\phi}^{\mathbf{X},\mathbf{X},\mathbf{X},Y}_{\Delta T_n, T_{i-1}} = \check{\varphi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n, T_{i-1}}\left(\hat{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n, T_{i-1}}\right)^{-1}\check{\varphi}^{\mathbf{X},Y}_{\Delta T_n, T_{i-1}} \text{ and } \hat{\phi}^{\mathbf{X},\mathbf{X},\mathbf{X},\mathbf{X}}_{\Delta T_n, T_{i-1}} = \check{\varphi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n, T_{i-1}}\left(\hat{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n, T_{i-1}}\right)^{-1}\check{\varphi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n, T_{i-1}},$$

where $\breve{\varphi}^{\mathbf{X},Y}_{\Delta T_n,T_{i-1}}$ and $\breve{\varphi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}$ is defined in (3.8) and with the notations "$\mathbf{X}$,Y" and "$\mathbf{X},\mathbf{X}$" that follow from the conventions of (3.1), respectively.

Before stating the Central Limit Theorem (CLT), we introduce the following notation. Recall the definition of $\left[M^{(r_1,s_1)}_{\infty}, M^{(r_2,s_2)}_{\infty}\right]_t$ in Remark 3 in Section 2.2. We set

$$
\begin{aligned}
\mathrm{ACOV}\left(M^{\mathbf{X},Y}, M^{\mathbf{X},Y}\right)^{(r,k)}_t &\triangleq \left[M^{(r,q+l)}_{\infty}, M^{(k,q+l)}_{\infty}\right]_t, \\
\mathrm{ACOV}\left(M^{\mathbf{X},Y}, M^{\mathbf{X},\mathbf{X}}\right)^{(r,k)}_t &\triangleq \left\{\left[M^{(r,q+l)}_{\infty}, M^{(v,k)}_{\infty}\right]_t\right\}_{1\le v\le q} \quad (q\times 1 \text{ vector process}), \text{ and} \\
\mathrm{ACOV}\left(M^{\mathbf{X},\mathbf{X}}, M^{\mathbf{X},\mathbf{X}}\right)^{(r,k)}_t &\triangleq \left\{\left[M^{(r,v)}_{\infty}, M^{(u,k)}_{\infty}\right]_t\right\}_{1\le v,u\le q} \quad (q\times q \text{ matrix process}).
\end{aligned}
\tag{3.10}
$$

and

$$
\mathbf{\Sigma}_t \triangleq \int_0^t \left(c^{\mathbf{X},\mathbf{X}}_u\right)^{-1} d\mathbf{\Lambda}_u \left(c^{\mathbf{X},\mathbf{X}}_u\right)^{-1}, \tag{3.11}
$$

where $d\mathbf{\Lambda}_u = \left\{d\Lambda^{(r,k)}_u\right\}_{1\le r,k\le q}$, and its $(r,k)-$th element is defined as:

$$
d\Lambda^{(r,k)}_u \triangleq d\mathrm{ACOV}\left(M^{\mathbf{X},Y}, M^{\mathbf{X},Y}\right)^{(r,k)}_u - \boldsymbol{\beta}^{\intercal}_u d\mathrm{ACOV}\left(M^{\mathbf{X},Y}, M^{\mathbf{X},\mathbf{X}}\right)^{(r,k)}_u [2] + \boldsymbol{\beta}^{\intercal}_u d\mathrm{ACOV}\left(M^{\mathbf{X},\mathbf{X}}, M^{\mathbf{X},\mathbf{X}}\right)^{(r,k)}_u \boldsymbol{\beta}_u, \tag{3.12}
$$

where [2] denotes the summation by switching r and k. Moreover, the $(r,k)-$th element of $\mathbf{\Sigma}_t$ can be expressed as:

$$
\mathbf{\Sigma}^{(r,k)}_t = \int_0^t \left(\mathbf{A}_u\right)^{\intercal}_{\bullet,r} d\mathbf{\Lambda}_u \left(\mathbf{A}_u\right)_{\bullet,k},
$$

where $\mathbf{A}_t \triangleq \left(c^{\mathbf{X},\mathbf{X}}_t\right)^{-1}$.

Finally, the CLT for $\tilde{\boldsymbol{\theta}}_n$ can be stated as follows.

Theorem 2. *(Central Limit Theorem for $\tilde{\boldsymbol{\theta}}_n$) Assume all conditions in Lemma 2 and further assume that $a_n^{-1}\Delta T_n \to 0$ and $a_n^{-3/2}\Delta T_n \to \infty$. Then we know that there is a $q\times q$ matrix process $(\mathbf{\Sigma}_t)_{0\le t\le \mathcal{T}}$ defined in (3.11), such that*

$$
a_n^{-1}\left(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right) \xrightarrow{\mathscr{L}} \mathbf{N}_q\left(0, \mathbf{\Sigma}_{\mathcal{T}}\right),
$$

where the convergence is stable in law, $\mathbf{N}_q\left(0,\mathbf{\Sigma}_{\mathcal{T}}\right)$ is a $q-$dimensional normal distribution with mean 0 and covariance matrix as $\mathbf{\Sigma}_{\mathcal{T}}$.

Proof. The proof of this theorem is collected in the Appendix E. $\square$

Moreover, following the idea of Mykland and Zhang (2017), it is straightforward to see that the asymptotic variance estimator could be constructed as follows:

$$\hat{\mathbf{\Sigma}}_{\mathcal{T}} = \Delta T_n^2 \sum_{i=1}^{B} \left(\hat{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n, T_{i-1}}\right)^{-1} \hat{\mathbf{\Phi}}_{\Delta T_n, T_{i-1}} \hat{\mathbf{\Phi}}^{\intercal}_{\Delta T_n, T_{i-1}} \left(\hat{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n, T_{i-1}}\right)^{-1}, \tag{3.13}$$

where

$$\hat{\mathbf{\Phi}}_{\Delta T_n, T_{i-1}} = \check{\varphi}^{\mathbf{X},Y}_{\Delta T_n, T_{i-1}} - \check{\varphi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n, T_{i-1}} \hat{\boldsymbol{\beta}}_{\Delta T_n, T_{i-1}}$$

with $\check{\varphi}^{\mathbf{X},Y}_{\Delta T_n, T_{i-1}}$ and $\check{\varphi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n, T_{i-1}}$ being defined in (3.8) and the notations "$\mathbf{X}$,Y" and "$\mathbf{X},\mathbf{X}$" following from the conventions of (3.1), respectively.

4. High Dimensional Factor Model

We again start by adjusting the notation. In the case of high dimensional factor model, we assume that $q, d > 1$, with d typically much larger than q. Specifically, q is asymptotically "almost" finite (see Condition 5 below), while $d \to \infty$ as $n \to \infty$. As foreshadowed in (2.1), denote

$$\mathbf{X} = \left(X^{(1)}, \ldots, X^{(q)}\right) = \left(\Xi^{(1)}, \ldots, \Xi^{(q)}\right), \text{ and } \mathbf{Y} = \left(Y^{(1)}, \ldots, Y^{(d)}\right) = \left(\Xi^{(q+1)}, \ldots, \Xi^{(q+d)}\right).$$

It is then also natural to use the following notations: $\langle \mathbf{X}, \mathbf{X} \rangle_t = \left\{ \langle \Xi^{(r)}, \Xi^{(s)} \rangle_t \right\}_{1 \le r,s \le q}$, $\langle \mathbf{X}, \mathbf{Y} \rangle_t = \left\{ \langle \Xi^{(r)}, \Xi^{(q+l)} \rangle_t \right\}_{1 \le r \le q, 1 \le l \le d}$, $\widehat{\langle \mathbf{X}, \mathbf{X} \rangle}_t = \left\{ \widehat{\langle \Xi^{(r)}, \Xi^{(s)} \rangle}_t \right\}_{1 \le r,s \le q}$, and $\widehat{\langle \mathbf{X}, \mathbf{Y} \rangle}_t = \left\{ \widehat{\langle \Xi^{(r)}, \Xi^{(q+l)} \rangle}_t \right\}_{1 \le r \le q, 1 \le l \le d}$. For the spot quantities, we define $c_t^{\mathbf{X},\mathbf{X}}$ as in (3.1), and define

$$c_t^{\mathbf{X},\mathbf{Y}} = \left\{ c_t^{(r,q+l)} \right\}_{1 \le r \le q, 1 \le l \le d}, \text{ which is a } q \times d \text{ matrix process, and} \tag{4.1}$$

$$c_t^{\mathbf{Y},\mathbf{Y}} = \left\{ c_t^{(q+r,q+s)} \right\}_{1 \le r,s \le d}, \text{ which is a } d \times d \text{ matrix process.} \tag{4.2}$$

Following the similar convention, we define the related matrix and vector quantities for $M, \bar{c}, \hat{c}, \bar{\pi}, \tilde{\pi}, \breve{\pi}, \breve{\varphi}, \tilde{e}$ an e. Then it is easy to see that in matrix form,

$$c_t = \begin{pmatrix} c_t^{\mathbf{X},\mathbf{X}} & c_t^{\mathbf{X},\mathbf{Y}} \\ \left(c_t^{\mathbf{X},\mathbf{Y}}\right)^\intercal & c_t^{\mathbf{Y},\mathbf{Y}} \end{pmatrix} \text{ and } \hat{c}_t = \begin{pmatrix} \hat{c}_t^{\mathbf{X},\mathbf{X}} & \hat{c}_t^{\mathbf{X},\mathbf{Y}} \\ \left(\hat{c}_t^{\mathbf{X},\mathbf{Y}}\right)^\intercal & \hat{c}_t^{\mathbf{Y},\mathbf{Y}} \end{pmatrix}, \tag{4.3}$$

with $\hat{c}_t = \left\{\hat{c}^{(r,s)}_{\Delta T_n,t}\right\}_{1\le r,s\le q+d}$ which is defined in (2.16).

4.1. *Specification of the factor model*

The log-price process $\mathbf{Y}_t = \left(Y_t^{(1)}, Y_t^{(2)}, \ldots, Y_t^{(d)}\right)$ of d stocks is generated from a multiple regression, also known as a "supervised" factor model:

$$d\mathbf{Y}_t = \mathbf{B}_t d\mathbf{X}_t + d\mathbf{Z}_t, \tag{4.4}$$

where $\mathbf{X}_t = \left(X_t^{(1)}, X_t^{(2)}, \ldots, X_t^{(q)}\right)$ is a $q \times 1$ vector process, denoting a set of time-varying common regressors or factors, $\mathbf{B}_t$ is a $d \times q$ matrix process of time-varying factor loadings and $\mathbf{Z}_t = \left(Z_t^{(1)}, Z_t^{(2)}, \ldots, Z_t^{(d)}\right)$ is a $d \times 1$ vector process of idiosyncratic noise components, satisfying

$$\langle \mathbf{X}, \mathbf{Z}\rangle_t = 0 \text{ for all } t, \tag{4.5}$$

cf. Mykland and Zhang (2006). The difference from an *unsupervised factor model,* is that in our current case, the factors $\mathbf{X}_t$ are observed, though with noise, and at asynchronous discrete times, as in Section 2.1.

It is straightforward to see that

$$d\langle \mathbf{Y}, \mathbf{Y}\rangle_t = \mathbf{B}_t d\langle \mathbf{X}, \mathbf{X}\rangle_t \mathbf{B}_t^\intercal + d\langle \mathbf{Z}, \mathbf{Z}\rangle_t \text{ for } 0 \le t \le \mathcal{T}, \tag{4.6}$$

whence also

$$c_t^{\mathbf{Y},\mathbf{Y}} = \mathbf{B}_t c_t^{\mathbf{X},\mathbf{X}} \mathbf{B}_t^\intercal + \mathbf{s}_t, \tag{4.7}$$

where $\mathbf{s}_t = \langle \mathbf{Z}, \mathbf{Z} \rangle_t'$, in view of (3.1), and (4.2). In this paper, we adopt the sparsity structure for $\mathbf{s}_t$, which is measured by

$$m_d = \sup_{0 \le t \le \mathcal{T}} \max_{1 \le i \le d} \sum_{1 \le j \le d} \left| \mathbf{s}_t^{(i,j)} \right|^{\nu} \text{ for some } \nu \in (0,1)\,,$$

and for $\nu = 0$, define $m_d = \sup_t \max_i \sum_j I\left(\mathbf{s}_t^{(i,j)} \neq 0 \right)$. This measure is widely used in existing literature, i.e., Bickel and Levina (2008) and Cai and Liu (2011) and as pointed out by Fan et al. (2013).

4.2. Least quadratic variation (LQV) optimization

In this case, the factors are observable. Thus, in order to get the estimates of factor loadings $\mathbf{B}_t$, we use the least quadratic variation (LQV) method:

$$(\mathbf{B}_t)_{0 \le t \le \mathcal{T}} = \underset{\mathbf{B}_t \in \mathbb{R}^{d \times q}, 0 \le t \le \mathcal{T}}{\arg\min} \operatorname{tr} \langle \mathbf{Z}, \mathbf{Z} \rangle_t \,.$$

Based on the similar derivation of (3.4), the LQV solution of factor loading can be expressed as:

$$\mathbf{B}_t^{\intercal} = \left(c_t^{\mathbf{X},\mathbf{X}} \right)^{-1} c_t^{\mathbf{X},\mathbf{Y}},$$

since we assume that that $\inf_{0 \le t \le \mathcal{T}} \lambda_{\min} \left(c_t^{\mathbf{X},\mathbf{X}} \right) > 0$ (Condition 4). Therefore, by the formula (4.7), the LQV solution for the spot idiosyncratic covariance matrix is:

$$\mathbf{s}_t = c_t^{\mathbf{Y},\mathbf{Y}} - c_t^{\mathbf{B} \bullet \mathbf{X}}, \tag{4.8}$$

where

$$c_t^{\mathbf{B} \bullet \mathbf{X}} = \left(c_t^{\mathbf{X},\mathbf{Y}} \right)^{\intercal} \left(c_t^{\mathbf{X},\mathbf{X}} \right)^{-1} c_t^{\mathbf{X},\mathbf{Y}}. \tag{4.9}$$

4.3. Estimators and convergence rates

We define the related estimators as follows:

$$
\begin{aligned}
\hat{\mathbf{B}}_t^{\intercal} &= \left(\hat{c}_t^{\mathbf{X},\mathbf{X}}\right)^{-1} \hat{c}_t^{\mathbf{X},\mathbf{Y}}, \\
\hat{c}_t^{\mathbf{B}\bullet\mathbf{X}} &= \hat{\mathbf{B}}_t \hat{c}_t^{\mathbf{X},\mathbf{X}} \hat{\mathbf{B}}_t^{\intercal} = \left(\hat{c}_t^{\mathbf{X},\mathbf{Y}}\right)^{\intercal} \left(\hat{c}_t^{\mathbf{X},\mathbf{X}}\right)^{-1} \hat{c}_t^{\mathbf{X},\mathbf{Y}}, \\
\hat{\mathbf{s}}_t &= \hat{c}_t^{\mathbf{Y},\mathbf{Y}} - \hat{c}_t^{\mathbf{B}\bullet\mathbf{X}},
\end{aligned}
\tag{4.10}
$$

where $\hat{c}_t^{\mathbf{X},\mathbf{X}}$, $\hat{c}_t^{\mathbf{X},\mathbf{Y}}$ and $\hat{c}_t^{\mathbf{Y},\mathbf{Y}}$ are defined in (4.3).

In the case of high dimensional factor models, we allow the number of common factors to diverge slowly, as the cross-sectional dimension d goes to infinity. The detailed technical assumption is stated as follows.

Condition 5. For the number of common factors q, we assume that $q = o(d)$ and $q^4 \Delta T_n \log d = o(1)$.

We now show the result for convergence rate of $\hat{c}_t^{\mathbf{B}\bullet\mathbf{X}}$ under elementwise max norm. Define:

$$
\omega_n = \left(q^4 \Delta T_n \log d\right)^{\frac{1}{2}}. \tag{4.11}
$$

Theorem 3. *Define* $\hat{c}_t = \left\{\hat{c}_{\Delta T_n,t}^{(r,s)}\right\}_{1 \le r,s \le q+d}$ *with* $\Delta T_n \asymp a_n$. *Assume Conditions 1-5. The following is then valid:*

$$
\left\|\hat{c}_t^{\mathbf{Y},\mathbf{Y}} - c_t^{\mathbf{Y},\mathbf{Y}}\right\|_{\max} = O_p\left((\Delta T_n \log d)^{\frac{1}{2}}\right), \text{ and} \tag{4.12}
$$

$$
\left\|\hat{c}_t^{\mathbf{B}\bullet\mathbf{X}} - c_t^{\mathbf{B}\bullet\mathbf{X}}\right\|_{\max} = O_p(\omega_n), \tag{4.13}
$$

where ω_n *is defined in (4.11). Consequently by the triangular inequality and formulas (4.8) and (4.10), we obtain:*

$$
\|\hat{\mathbf{s}}_t - \mathbf{s}_t\|_{\max} = O_p(\omega_n).
$$

Proof. The proof of this theorem is collected in the Appendix G. Specifically, (4.12) is a consequence of Lemma 3, while (4.13) follows from (G.6). □

Now we apply the adaptive thresholding on $\hat{\mathbf{s}}_t$. Denote the thresholding estimator by $\hat{\mathbf{s}}_t^*$, defined as follows:

$$\hat{\mathbf{s}}_t^* \triangleq \begin{cases} \hat{\mathbf{s}}_t^{(i,j)}, & i = j, \\ \phi_{ij}\left(\hat{\mathbf{s}}_t^{(i,j)}\right), & i \neq j, \end{cases}$$

where ϕ_{ij} is the adaptive thresholding rule, for $z \in \mathbb{R}$,

$$\phi_{ij}(z) = 0 \text{ when } |z| \leq \chi_{ij}, \text{ otherwise } \left|\phi_{ij}(z) - z\right| \leq \chi_{ij}.$$

The examples of adaptive thresholding rule include the hard thresholding $\phi_{ij}(z) = zI\left(|z| \geq \chi_{ij}\right)$, soft thresholding, SCAD and the adaptive lasso, i.e., see Rothman et al. (2009) and Fan et al. (2016b). Because of the absence of residuals, the standard error estimator of $\hat{\mathbf{s}}_t^{(i,j)}$ can not be easily obtained. Thus, in contrast to the settings of χ_{ij} in Fan et al. (2013), the thresholding parameter are set to be elementwise constant, i.e., defined as:

$$\chi_{ij} = C\omega_n, \tag{4.14}$$

with a sufficiently large $C > 0$. Before stating the results of the thresholding estimator, we first make one assumption about the spot residual covariance matrix.

Condition 6. For the spot residual covariance matrix $\mathbf{s}_t$, there exist constants $\vartheta_1', \vartheta_2' > 0$ such that $\vartheta_1' < \lambda_{\min}(\mathbf{s}_t) \leq \lambda_{\max}(\mathbf{s}_t) < \vartheta_2'$ for all $0 \leq t \leq \mathcal{T}$.

Based on the result in Theorem 3, we obtain the following proposition.

Proposition 1. *Assume Conditions 1-6. Then, for a sufficiently large $C > 0$ in the thresholding parameter (4.14), the estimator for the sparse residual covariance matrix satisfies:*

$$\left\|\hat{\mathbf{s}}_t^* - \mathbf{s}_t\right\| = O_p\left(\omega_n^{1-\nu} m_d\right).$$

If $\omega_n^{1-\nu} m_d = o_p(1)$ is assured, then the eigenvalues of $\hat{\mathbf{s}}_{\hat{q}_t,t}^*$ are all bounded away from 0 with probability approaching 1, and

$$\left\|\left(\hat{\mathbf{s}}_t^*\right)^{-1} - \mathbf{s}_t^{-1}\right\| = O_p\left(\omega_n^{1-\nu} m_d\right).$$

Proof. The proof of this proposition directly follows from the similar discussions in the proof of Theorem 5 of Fan et al. (2013). □

Next, let's define the spot covariance matrix estimator based on the thresholding estimator as follows:

$$\hat{c}_t^{\mathbf{Y},\mathbf{Y},*} := \left(\hat{c}_t^{\mathbf{X},\mathbf{Y}}\right)^\intercal \left(\hat{c}_t^{\mathbf{X},\mathbf{X}}\right)^{-1} \hat{c}_t^{\mathbf{X},\mathbf{Y}} + \hat{\mathbf{s}}_t^*.$$

Then we consider the estimation performance of precision matrix based on $\left(\hat{c}_t^{\mathbf{Y},\mathbf{Y},*}\right)^{-1}$. The theoretical development is based on the Sherman-Morrison-Woodbury formula, i.e., recall the formulas (4.8) and (4.9), we obtain:

$$\left(\hat{c}_t^{\mathbf{Y},\mathbf{Y},*}\right)^{-1} = (\hat{\mathbf{s}}_t^*)^{-1} - (\hat{\mathbf{s}}_t^*)^{-1} \left(\hat{c}_t^{\mathbf{X},\mathbf{Y}}\right)^\intercal \left[\hat{c}_t^{\mathbf{X},\mathbf{X}} + \hat{c}_t^{\mathbf{X},\mathbf{Y}} (\hat{\mathbf{s}}_t^*)^{-1} \left(\hat{c}_t^{\mathbf{X},\mathbf{Y}}\right)^\intercal\right]^{-1} \hat{c}_t^{\mathbf{X},\mathbf{Y}} (\hat{\mathbf{s}}_t^*)^{-1}.$$

We first assume the pervasiveness of the common factors by the following technical assumption, which is parallel to the Assumption 3.5 in Fan et al. (2011).

Condition 7. Assume

$$\left\| d^{-1} c_t^{\mathbf{X},\mathbf{Y}} \left(c_t^{\mathbf{X},\mathbf{Y}}\right)^\intercal - \mathbf{\Omega}_t \right\| = o(1)$$

for some $q \times q$ symmetric positive definite matrix $\mathbf{\Omega}_t$ and some constants $\vartheta_3', \vartheta_4' > 0$ such that

(i) $\inf_{0 \le t \le \mathcal{T}} \lambda_{\min}(\mathbf{\Omega}_t) > \vartheta_3'$ almost surely;

(ii) if $q \to \infty$ as $n, d \to \infty$, we further assume $\sup_{0 \le t \le \mathcal{T}} \lambda_{\max}(\mathbf{\Omega}_t) < \vartheta_4'$ almost surely.

Then we show the convergence rate for the estimator of the precision matrix as follows.

Theorem 4. *Assume Conditions 1-7. Also suppose that $\omega_n^{1-\nu} m_d = o_p(1)$. Then, for a sufficiently large $C > 0$ in thresholding parameter (4.14), $\left(\hat{c}_t^{\mathbf{Y},\mathbf{Y},*}\right)^{-1}$ is non-singular with probability approaching* 1*, and*

$$\left\| \left(\hat{c}_t^{\mathbf{Y},\mathbf{Y},*}\right)^{-1} - \left(c_t^{\mathbf{Y},\mathbf{Y}}\right)^{-1} \right\| = O_p\left(\omega_n^{1-\nu} m_d\right).$$

Proof. The proof of this theorem is collected in Appendix G. 1. □

5. Monte Carlo Evidence

We conduct Monte Carlo simulation to verify the validity of our methodology.

5.1. *Simulation settings*

Following the model setup in (4.4) - (4.5) in Section 4, we consider the log-price process $\mathbf{Y}_t = \left(Y_t^{(1)}, Y_t^{(2)}, \ldots, Y_t^{(d)}\right)$ of d stocks is generated from a factor model $d\mathbf{Y}_t = \mathbf{B}_t d\mathbf{X}_t + d\mathbf{Z}_t$, where the common factors $\mathbf{X}_t$ and factor loadings $\mathbf{B}_t$ are $q \times 1$ and $d \times q$ time-varying processes, respectively. And $\mathbf{Z}_t$ is a $d \times 1$ vector process of idiosyncratic noise components.

In the simulation, we specify

$$dX_t^{(j)} = \mu_j dt + \sigma_t^{(j)} d\mathcal{W}_t^{(j)} \text{ and } dZ_t^{(i)} = \nu_t d\mathcal{B}_t^{(i)},$$

where $q = 3$, $j = 1, 2, ..., q$. And $\left\{\mathcal{B}_t^{(i)}\right\}_{1 \le i \le d}$ are the independent standard Brownian motions.

The correlation matrix of $d\mathcal{W}$ is defined as $\rho^{\mathbf{X}}$. The volatility processes of $\mathbf{X}$ and $\mathbf{Z}$ are simulated as follows:

$$d\left(\sigma_t^{(j)}\right)^2 = \kappa_j \left(\theta_j - \left(\sigma_t^{(j)}\right)^2\right) dt + \eta_j \sigma_t^{(j)} d\tilde{\mathcal{W}}_t^{(j)} \text{ and } d\nu_t^2 = \kappa^\nu \left(\theta^\nu - \nu_t^2\right) dt + \eta^\nu \nu_t d\bar{\mathcal{B}}_t$$

where the correlation between $d\mathcal{W}^{(j)}$ and $d\tilde{\mathcal{W}}^{(j)}$ is ρ_j.

The first component of $\mathbf{X}$ in the simulation is set as the market factor. To guarantee its factor loadings $\mathbf{B}_{\bullet,1}$ are positive, we simulate the factor loading in the following scheme, for $i = 1, \cdots, d$,

$$d\mathbf{B}_t^{(i,j)} = \begin{cases} \tilde{\kappa}_1 \left(\tilde{\theta}_{i1} - \mathbf{B}_t^{(i,j)}\right) dt + \tilde{\xi}_1 \sqrt{\mathbf{B}_t^{(i,j)}} d\tilde{\mathcal{B}}_t^{(i,j)} & \text{if } j = 1, \\ \tilde{\kappa}_j \left(\tilde{\theta}_{ij} - \mathbf{B}_t^{(i,j)}\right) dt + \tilde{\xi}_j d\tilde{\mathcal{B}}_t^{(i,j)} & \text{if } j \ge 2. \end{cases}$$

The parameters are set as follows[8]: $\mu = (0.05, 0.03, 0.02)$, $\tilde{\kappa} = (1, 2, 3)$, $\tilde{\xi} = (0.5, 0.6, 0.7)$, $\tilde{\theta}_1 \sim U[0.25, 1.75]$, $\tilde{\theta}_2, \tilde{\theta}_3 \sim N\left(0, 0.5^2\right)$, $\kappa = (3, 4, 5)$, $\theta = (0.05, 0.04, 0.03)$, $\eta = (0.3, 0.4, 0.3)$, $\rho = (-0.6, -0.4, -0.25)$, $\rho_{12}^{\mathbf{X}} = 0.05$, $\rho_{13}^{\mathbf{X}} = 0.1$, $\rho_{23}^{\mathbf{X}} = 0.15$, $\kappa^\nu = 4$, $\theta^\nu = 0.06$ and $\eta^\nu = 0.3$.

[8] this is similar to Aït-Sahalia and Xiu (2019) with $\theta^\nu = 0.06$ and $\eta^\nu = 0.3$

The processes are simulated in the equidistant grid with $\Delta t_n = 1$ second. The observed processes are contaminated by microstructure noise:

$$\Xi^*_{t_j} = \Xi_{t_j} + \epsilon_{t_j}, \tag{5.1}$$

where $\Xi = \left(X^{(1)}, X^{(2)}, \ldots, X^{(q)}, Y^{(1)}, Y^{(2)}, \ldots, Y^{(d)}\right)$ and ϵ_{t_j} are i.i.d. $(q+d)$-dimensional random vectors, sampled from $\mathbf{N}_{q+d}\left(0, \boldsymbol{\Sigma}^{\epsilon}\right)$, with $\boldsymbol{\Sigma}^{\epsilon} = \boldsymbol{\Phi}\boldsymbol{\Phi}^{\intercal}$ and $\boldsymbol{\Phi} = (\Phi_1, \Phi_2, \ldots, \Phi_{q+d})^{\intercal}$ and $\Phi_1, \Phi_2, \ldots, \Phi_{q+d}$ are i.i.d. random variables from $N\left(0, (0.0005)^2\right)$.

The time horizon in the simulation experiment is set as: $\mathcal{T} = 1$ day. We assume that a trading day consists of 6.5 hours for open trading.

5.2. *Simulation results for $d = 1$*

We note that for $d = 1$, the factoring loading $\mathbf{B}^{(1,j)}_t$ is the same as $\beta^{(j)}_t$ in model (3.2).

We apply the realized regression procedure by estimating $\tilde{\boldsymbol{\theta}}_n$, defined in (3.9). To illustrate the effect of market microstructure noise, we also construct the estimator $\tilde{\boldsymbol{\theta}}_n$ by replacing the spot covariance matrix estimator $\hat{c}^{(r,s)}_{\Delta T_n, t}$ with simple CV: $\hat{c}_t = \frac{1}{k_n \Delta\tau_n} \sum_{j=N^*(t)+1}^{N^*(t)+k_n} \Delta\Xi_{\tau_j} \Delta\Xi^{\intercal}_{\tau_j}$ (without noise) and $\hat{c}_t = \frac{1}{k_n \Delta\tau_n} \sum_{j=N^*(t)+1}^{N^*(t)+k_n} \Delta\Xi^*_{\tau_j} \left(\Delta\Xi^*_{\tau_j}\right)^{\intercal}$ (with noise). The number of simulation trials is 10000. The examination is conducted under different settings of sampling frequency. The sampling frequency is set in two scenarios:

1. $\Delta\tau_n = 5$ seconds and $\Delta T_n = 468\Delta\tau_n$, with $K = 20, J = 3$.
2. $\Delta\tau_n = 15$ seconds and $\Delta T_n = 156\Delta\tau_n$, with $K = 10, J = 3$.

Table 5.1 shows that in the presence of microstructure noise, the estimator based on Simple CV becomes inconsistent: it tends to under-estimate the market beta $\int_0^{\mathcal{T}} \beta^{(1)}_t dt$, and over-estimate the other non-market betas $\int_0^{\mathcal{T}} \beta^{(2)}_t dt$ and $\int_0^{\mathcal{T}} \beta^{(3)}_t dt$. When $\Delta\tau_n = 5$ seconds, the magnitude of the estimation bias for $\int_0^{\mathcal{T}} \beta^{(1)}_t dt$ is 26.8% of the averaged true value and the bias magnitude for $\int_0^{\mathcal{T}} \beta^{(2)}_t dt$ and $\int_0^{\mathcal{T}} \beta^{(3)}_t dt$ are 81.6% and 230.2% comparing to their averaged true values. It also appears that the estimation bias (under the market microstructure noise) becomes more severe as the length of the sampling interval $\Delta\tau_n$ decreases from 15 to 5 seconds. On the other hand, our proposed estimator (based Smoothed TSRV) is well behaved, regardless of the sampling interval.

Table 5.1. Simulation Results for Integrated Beta Estimates

True Value: $\int_0^{\mathcal{T}} \beta_t^{(1)} dt$	$\Delta\tau_n = 5$ seconds		$\Delta\tau_n = 15$ seconds	
Averaged Mean: 1.002307	Bias	Stdev	Bias	Stdev
Simple CV without Noise (unobservable)	0.000047	0.017227	0.000256	0.031344
Simple CV with Noise	−0.268700	0.349729	−0.248314	0.326741
Smoothed TSRV	0.002764	0.076635	0.002519	0.112432
True Value: $\int_0^{\mathcal{T}} \beta_t^{(2)} dt$	$\Delta\tau_n = 5$ seconds		$\Delta\tau_n = 15$ seconds	
Averaged Mean: −0.006275	Bias	Stdev	Bias	Stdev
Simple CV without Noise (unobservable)	−0.000238	0.019537	−0.000471	0.035373
Simple CV with Noise	0.005119	0.374072	0.004309	0.350225
Smoothed TSRV	0.000011	0.083769	−0.000045	0.126580
True Value: $\int_0^{\mathcal{T}} \beta_t^{(3)} dt$	$\Delta\tau_n = 5$ seconds		$\Delta\tau_n = 15$ seconds	
Averaged Mean: −0.007281	Bias	Stdev	Bias	Stdev
Simple CV without Noise (unobservable)	0.000118	0.022624	0.000358	0.040619
Simple CV with Noise	0.016762	0.460584	0.016331	0.433926
Smoothed TSRV	0.000568	0.096653	0.000924	0.146875

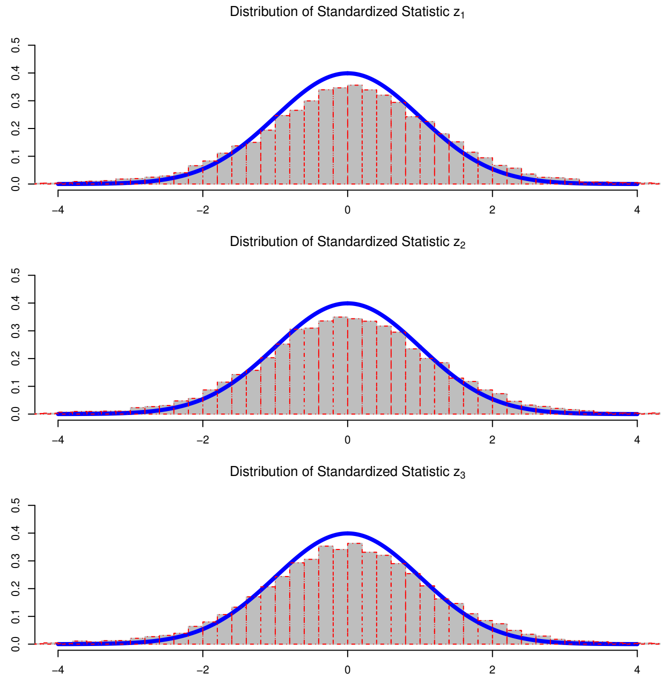
This table reports the summary statistics for the estimation of the three integrated betas, i.e., for $p = 1, 2$ and 3, $\int_0^{\mathcal{T}} \beta_t^{(p)} dt$ denotes the integrated p-th beta. The Monte Carlo simulation consists of 10000 trials and $\Delta\tau_n = 5$ and 15 seconds. The Column "Bias" denotes the mean of estimation error; Column "Stdev" denotes the standard deviation of the estimation error.

To validate the asymptotic behavior of the bias corrected estimator, the finite sample distribution of the standardized statistics are reported in Figure 5.1, where $\Delta\tau_n = 5$ seconds. Note that the standardized statistics are calculated by the following formulas

$$z_k = \frac{\tilde{\boldsymbol{\theta}}_n^{(k)} - \int_0^{\mathcal{T}} \beta_t^{(k)} dt}{\left(\hat{\boldsymbol{\Sigma}}_{\mathcal{T}}^{(k,k)}\right)^{1/2}}, \text{ for } k = 1, 2, \ldots, q, \tag{5.2}$$

where $\tilde{\boldsymbol{\theta}}_n^{(k)}$ is defined in (3.9) and $\hat{\boldsymbol{\Sigma}}_{\mathcal{T}}^{(k,k)}$ is defined in (3.13). In Figure 5.1, the finite sample distributions are approximately normal, with slight fat-tailed. It is worth to emphasize that the edge effects can

Figure 5.1. Finite Sample Distributions of Standardized Statistics



Notes. This figure reports the histogram of the 10000 trials simulation for estimating the three integrated betas with $\Delta\tau_n = 5$ seconds over 1 day. The solid blue lines are the standard normal density; the gray histograms with bars of red dashed border are the distributions of the bias corrected estimator. The standardized statistic z_k is defined in formula (5.2), for $k = 1, 2, ..., q$.

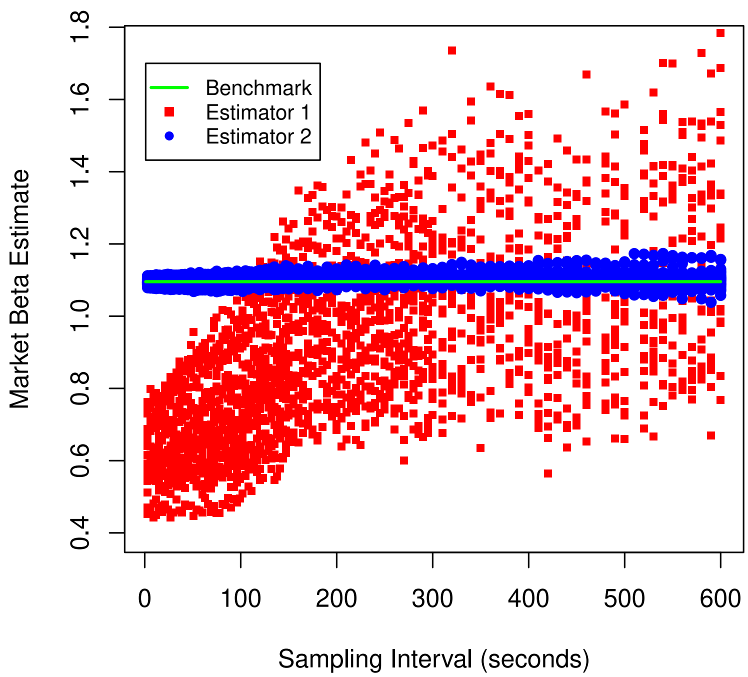
greatly affect the validity of the asymptotic normality in this scenario (i.e., in practice, $\mathcal{T}/\Delta T_n$ should be a positive integer exactly).

As sampling interval increases, say, to 5 minutes or 10 minutes, one could expect the bias of the integrated beta estimate using simple CV goes down. However, its variance increases at the same time. This phenomenon is demonstrated in the signature plot 5.2. So, even when one samples very 10 minutes, we still recommend our proposed estimator because of its precision.

5.3. Simulation results for high dimension

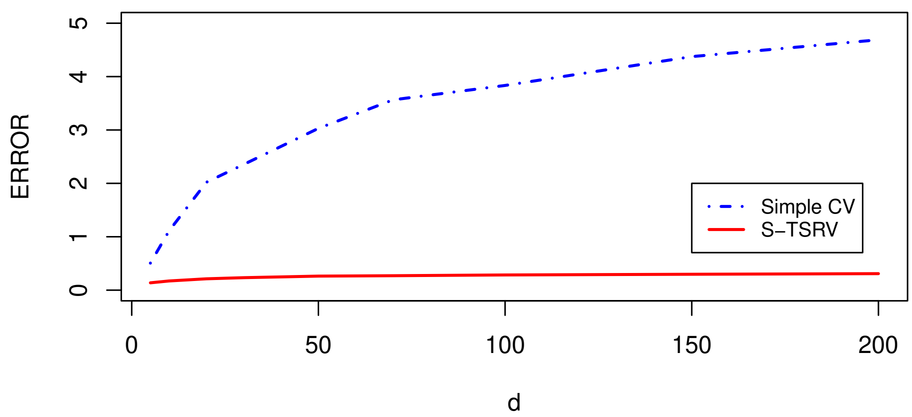
For d large, we next show the performance of the estimator of the precision matrix $\left(\hat{c}_t^{\mathbf{Y},\mathbf{Y},*}\right)^{-1}$, as d gets large. The simulation setting remains the same as in Section 5.1. For ease of demonstration, we fix $\Delta\tau_n = 5$ seconds.

Figure 5.2. Signature Plot of Market Beta Estimate



This figure presents the signature plot for the estimates of integrated market beta $\int_0^{\mathcal{T}} \beta_t^{(1)} dt$ in the presence of market microstructure noise. "Estimator 1" denotes the integrated beta estimate based on the Simple CV estimator with subsampled data. "Estimator 2" denotes integrated beta estimate proposed in this paper which is based on Smoothed TSRV.

Figure 5.3. Estimation Performance of the Large Precision Matrix



This figure compares the estimation performance of the large precision matrix in the presence of microstructure noise. The precision matrix using Smooth TSRV is indicated by red solid line, while the one using simple CV is in blue dashed line. The error measure on y-xis is as defined in (5.3).

Consider the spectral norm of the estimation error of the precision matrix, as defined in Theorem 4, as the error measure, i.e.

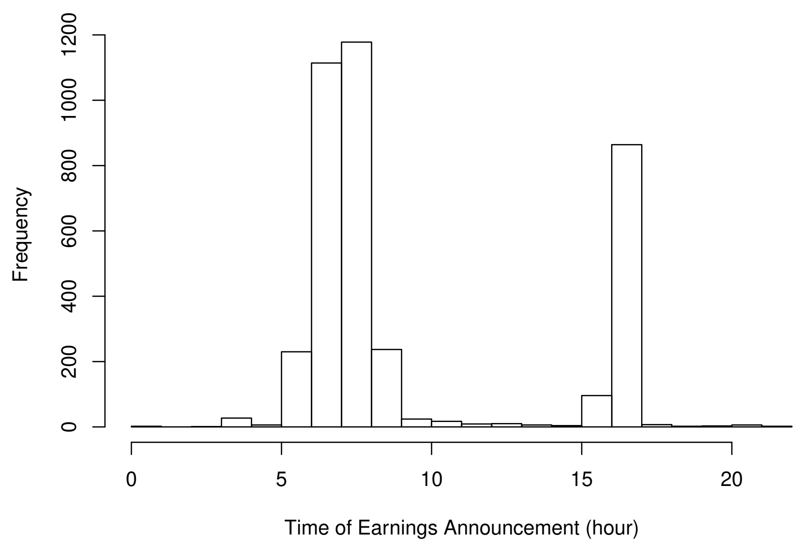
$$\text{ERROR} = \left\| \left(\hat{c}_t^{\mathbf{Y},\mathbf{Y},*}\right)^{-1} - \left(c_t^{\mathbf{Y},\mathbf{Y}}\right)^{-1} \right\|. \tag{5.3}$$

As in Figure 5.3, we see that in the presence of microstructure noise, the precision matrix using Smoothed TSRV performs satisfactorily, with contained error (red line) even at high dimensionality situation. However, the precision matrix using simple CV gets worse as d increases from 5 to 200, with error increasing in logarithmic shape.

6. Empirical Study

In this section, we apply high frequency beta estimation to study the variation of stock betas on earnings announcement days. We implement the high frequency regression of the intraday returns of the S&P 100 constituent stocks on the returns of exchange-traded fund OEF. The latter serves as a proxy for the large-cap market returns. The trade prices are extracted from the Trade and Quote (TAQ) database of

Figure 6.1. Distribution of Earnings Announcements' Arrival Times



This figure shows the distribution of the arrival times of the earnings announcements for the S&P 100 constituents during the sampling period between January 2007 and December 2017.

the New York Stock Exchange (NYSE). In particular, we collect the intraday trade prices of OEF as well as those of the S&P 100 Index constituents, between 9:35 a.m. EST and 3:55 p.m. EST of each trading day, from January 2007 to December 2017 (2769 trading days in total). Our spot beta estimate is then applied to explore the change in betas around the earning announcements.

For the earning data, the dates and times of quarterly earnings announcements are downloaded from the Thomson Reuters I/B/E/S database for the components of S&P 100 Index ranging from January 2007 to December 2017. The earnings announced at non-trading days are deleted. At the end, 3845 earnings announcements are collected during this sampling period. We can see from Figure 6.1 that for the stocks in our sample, most earnings announcements arrived right before the market open (6-8 AM) or right after market close (4-5 PM).

To investigate the beta changes on earnings announcement days, we extended the model in Patton and Verardo (2012)[9] by adding the hourly effects. Specifically, we regress the market beta estimates

[9]We should note that we deviate from Patton and Verado (2012) in how to define event day. The former paper relabeled

$\beta_{i,t}^{\mathrm{OEF}}$ on event time dummies using the following model:

$$\beta_{i,t}^{\mathrm{OEF}} \quad = \quad \sum_{j=-2}^{2}\sum_{k=10}^{16}\delta_{j,k}I_{i,j,k,t}+\gamma_{i,1}D_{1,t}+\gamma_{i,2}D_{2,t}+\cdots\gamma_{i,10}D_{10,t}+\varepsilon_{i,t},$$

where $\beta_{i,t}^{\mathrm{OEF}}$ is the spot beta estimates of stock i on time t by using the following formula, with $\hat{c}^{\text{OEF, stock i}}$ and $\hat{c}^{\mathrm{OEF,OEF}}$ being the Smoothed TSCV and Smoothed TSRV estimates,

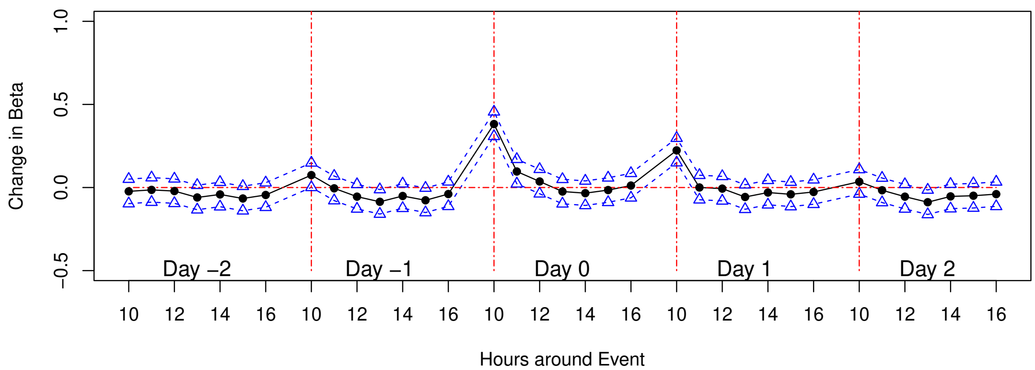
$$\beta_{i,t}^{\mathrm{OEF}}=\frac{\hat{c}_{\Delta T_n,t}^{\text{OEF, stock i}}}{\hat{c}_{\Delta T_n,t}^{\mathrm{OEF,OEF}}},$$

and $I_{i,\mathrm{day},\mathrm{hour},t}$ are dummy variables defined over a 5-day time window around the earnings announcements, with day$=0$ representing the earnings announcement date, and hour$=10,11,\ldots,16$ representing the hour in each trading day. For each trading day, the spot beta estimates $\beta_{i,t}^{\mathrm{OEF}}$ are computed with the 5-second returns over the following 7 time intervals: $[9:30,10:00]$, $(10:00,11:00],\cdots,$ $(15:00,16:00]$. The dummy variables $D_{1,t}$ to $D_{10,t}$ are the year fixed effects, corresponding to the 10 years from 2007 to 2016. D_{11t} for year 2017 is excluded for the identification purpose.

In order to get an impression on the hourly behavior of beta, we first conduct aggregating regression on the entire sample. Figure 6.2 and Table 6.1 suggest that the stock betas stay at the non-announcement level during most hours over the 5-day window around earnings release. The exceptions occur at the early hours of market open. In particular, we observe large beta increase at the first hour (i.e. 10am) on both Day 0 and Day 1. The first-hour beta increase (0.38 with a $t-$statistic of 10.15) in Day 0 seems to reflect the incorporation of the earnings news which arrive before the market opens that day; on the other hand, the first-hour beta increase (0.22 with a $t-$statistic of 5.94) in Day 1 suggest the impact of the earnings news which are announced post-4pm from the preceding day. This interpretation is further confirmed when we zoom in two subsamples, those with earnings announced prior to market opens at 9:30 and those announced after 16:00. Figure 6.3 displays the separation of before- and after-market earnings announcement impact on beta, with before-market effect on panel (a) and post-market effect on

the announcement at or after 4:00 p.m. on a given date to have the following trading day's date. In contrast, we follow the exact calendar day when labeling the announcement day. In other words, "Day 0" in our paper is the day when the earnings are announced, no matter the announcement time is pre-market, during market open, or post-4pm.

Figure 6.2. Changes in Market Beta around Earnings Announcements



This figure shows the estimated changes in market beta for the five-day window around quarterly earnings announcements of the components in S&P 100 Index. Black solid line denotes the beta estimates, while the blue dashed line denotes the 95% confidence intervals.

Table 6.1. Changes in Market Beta around Earnings Announcements

Day −2		Day −1		Day 0		Day 1		Day 2	
Hour	Beta	Hour	Beta	Hour	Beta	Hour	Beta	Hour	Beta
10:00	−0.023	10:00	0.075	10:00	**0.382**	10:00	**0.223**	10:00	0.035
	(−0.615)		(1.985)		(**10.149**)		(**5.939**)		(0.925)
11:00	−0.014	11:00	−0.005	11:00	0.096	11:00	2×10^{-4}	11:00	−0.016
	(−0.377)		(−0.129)		(2.560)		(0.007)		(−0.438)
12:00	−0.021	12:00	−0.055	12:00	0.036	12:00	−0.007	12:00	−0.055
	(−0.567)		(−1.460)		(0.964)		(−0.174)		(−1.462)
13:00	−0.060	13:00	−0.086	13:00	−0.024	13:00	−0.057	13:00	−0.088
	(−1.592)		(−2.280)		(−0.642)		(−1.518)		(−2.350)
14:00	−0.042	14:00	−0.051	14:00	−0.034	14:00	−0.030	14:00	−0.053
	(−1.108)		(−1.355)		(−0.903)		(−0.799)		(−1.420)
15:00	−0.066	15:00	−0.077	15:00	−0.015	15:00	−0.041	15:00	−0.049
	(−1.755)		(−2.040)		(−0.406)		(−1.089)		(−1.313)
16:00	−0.045	16:00	−0.039	16:00	0.012	16:00	−0.027	16:00	−0.040
	(−1.192)		(−1.045)		(0.314)		(−0.714)		(−1.057)

This table reports the beta estimates and related t-statistics over the five days around each earnings announcement during 2007-2017 for the components of S&P 100 Index. The Day 0 denotes the earnings announcement date. The Day −1 and Day −2 denotes the two days before the earnings announcement date, and the Day 1 and Day 2 indicate the two days after the earnings announcement date. The *t*-statistics are shown in parentheses.

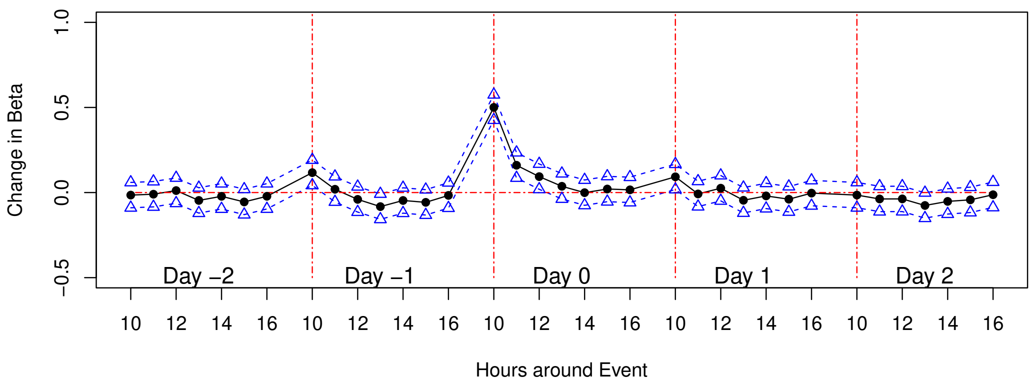
Table 6.2. Distribution of Two Types of News

	Before Market	After Market	Market Open Hours
Good News	2707	945	145
Bad News	68	18	6

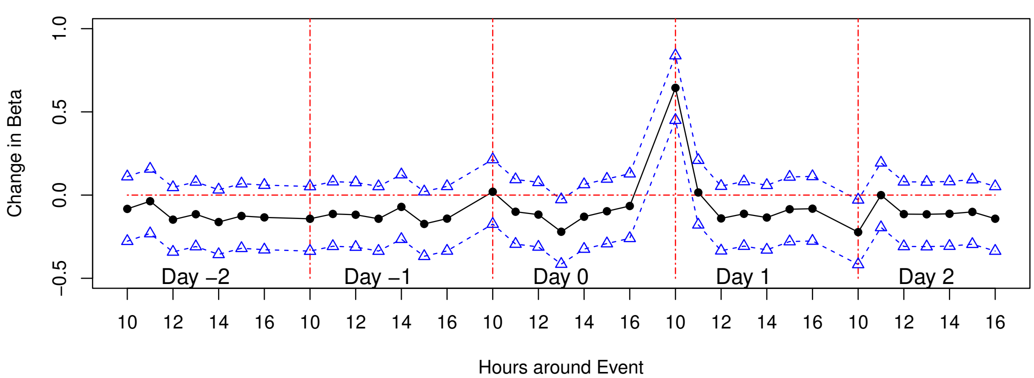
panel (b). We should mention that panel (a) also shows a small increase in beta at 10am on Day -1 and Day $+1$, when earnings were announced in the morning prior to market open. Since the magnitude of the latter beta changes is small, we do not put emphasis on its economic implication. Though, one could not rule out the possibility of overnight information (earnings as well as non-earnings) accumulation and its impact on first hour beta.

The change in stock betas around different announcement times cannot be explained by good versus bad news. We can see in Table 6.2 that in our sample from 2007-2017, most of earnings announcements belong to good news and their arrival times do not follow systematic pattern. We also looked into the pattern of announcement arrivals during market open hours in Figure 6.4. Among the relatively small number of announcements arriving over the market open hours, the announcement seems to evenly spread out from 9:30 to 3pm and then there is an increase in the final hour (3-4PM) of market open time. The news in the final hour of trading period may also contribute to the beta increase in the next morning.

Figure 6.3. Changes in Market Beta around Earnings Announcements by Separating Data

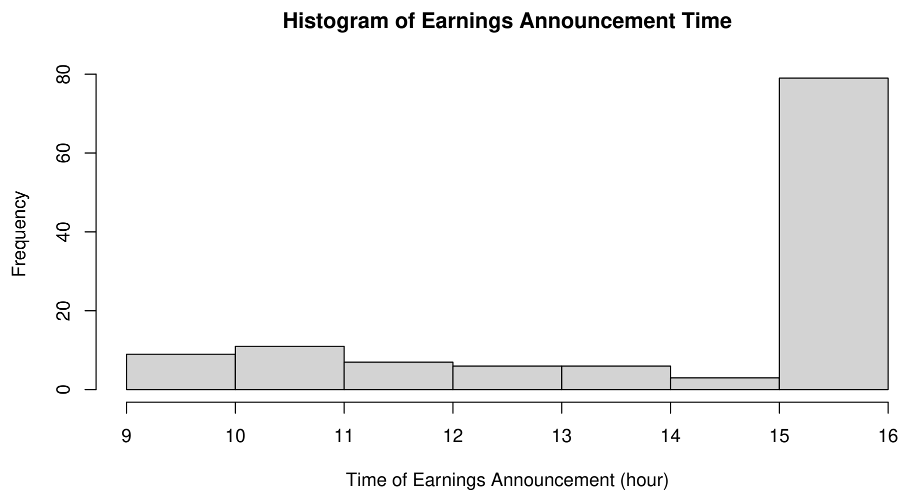


(a) beta increase at 10 same day when earnings announced before market open on Day 0



(b) beta increase at 10 next day when earnings announced after market close on Day 0

Figure 6.4. Distribution of Earnings Announcements' Arrival Times between 9:30 and 16:00



Notes. This figure reports the distribution of the arrival times of the earnings announcements between 9:30 and 16:00.

7. Conclusion

The central contribution of this paper is a feasible estimator of spot beta, which is robust to noise and asynchronicity. With the help of the spot-version of the Smoothed TSRV estimator, spot beta can be consistently estimated. There are two direct applications of the spot beta estimates in the current paper. In the first application, the integrated beta can be consistently estimated by aggregating the spot beta estimates. After a bias-correction procedure, a fixed dimension central limit theorem is established for the bias-corrected estimator, with convergence rate which may be arbitrarily close to $O_p(n^{-1/4})$. In the second application we assume time-varying factor structure and conditional sparsity. The spot beta matrix estimator enables the estimation of high dimensional spot covariance and precision matrices. Simulation results show that our proposed estimators perform well.

As an empirical application, this paper explores the hourly change in beta around earnings announcements of the S&P 100 constituents. The hourly beta was constructed with the help of Smooth TSRV using 5-second pre-averaged returns from 2007 to 2017. We separate the impact of pre- and post-market announcement on beta change, and find that significant beta change takes place in the first hour of market open.

References

Aït-Sahalia, Y., Jacod, J., Xiu, D., 2021. Inference on risk premia in continuous-time asset pricing models, technical report.

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

Aït-Sahalia, Y., Mykland, P. A., Zhang, L., 2011. Ultra high frequency volatility estimation with dependent microstructure noise. Journal of Econometrics 160, 160–175.

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

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

Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P., 2000. Great realizations. Risk 13, 105–108.

Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P., 2001. The distribution of realized exchange rate volatility. Journal of the American Statistical Association 96, 42–55.

Andersen, T. G., Bollerslev, T., Diebold, F. X., Wu, G., 2006. Realized beta: Persistence and predictability. In: Econometric Analysis of Financial and Economic Time Series. Emerald Group Publishing Limited, pp. 1–39.

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

Ball, R., Kothari, S. P., 1991. Security Returns around Earnings Announcements. The Accounting Review 66 (4), 718–738.

Barndorff-Nielsen, O. E., Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. Journal of the Royal Statistical Society, B 64, 253–280.

Barndorff-Nielsen, O. E., Shephard, N., 2004. Econometric analysis of realised covariation: high frequency based covariance, regression and correlation in financial economics. Econometrica 72, 885–925.

Bibinger, M., Hautsch, N., Malec, P., Reiss, M., 2017. Estimating the spot covariation of asset prices: Statistical theory and empirical evidence. Journal of Business and Economic Statistics , 1–17.

Bibinger, M., Mykland, P. A., 2016. Inference for multi-dimensional high-frequency data with an application to conditional independence testing. Scandinavian Journal of Statistics 43 (4), 1078–1102, 10.1111/sjos.12230.
URL `http://dx.doi.org/10.1111/sjos.12230`

Bickel, P. J., Levina, E., 2008. Covariance regularization by thresholding. The Annals of Statistics , 2577–2604.

Black, F., 1972. Capital market equilibrium with restricted borrowing. The journal of business 45 (3), 444–455.

Boguth, O., Carlson, M., Fisher, A., Simutin, M., 2011. Conditional risk and performance evaluation: Volatility timing, overconditioning, and new estimates of momentum alphas. Journal of Financial Economics 102 (2), 363–389.

Bollerslev, T., Engle, R. F., Wooldridge, J. M., 1988. A capital asset pricing model with time-varying covariances. Journal of Political Economy 96 (1), 116–131.

Cai, T., Liu, W., 2011. Adaptive thresholding for sparse covariance matrix estimation. Journal of the American Statistical Association 106 (494), 672–684.

Campbell, J. Y., Lo, A. W., MacKinlay, A. C., 1997. The Econometrics of Financial Markets. Princeton University Press, Princeton, NJ.

Chen, D., Mykland, P. A., Zhang, L., 2020. The five trolls under the bridge: Principal component analysis with asynchronous and noisy high frequency data. Journal of the American Statistical Association 115, 1960–1977.

Cochrane, J. H., 2005. Asset pricing, 2nd Edition. Princeton.

Dai, C., Lu, K., Xiu, D., 2019. Knowing factors or factor loadings, or neither? evaluating estimators of large covariance matrices with noisy and asynchronous data. Journal of Econometrics 208.

Engle, R. F., 2000. The econometrics of ultra-high frequency data. Econometrica 68, 1–22.

Engle, R. F., 2016. Dynamic conditional beta. Journal of Financial Econometrics 14 (4), 643–667.

Fama, E. F., MacBeth, J. D., 1973. Risk, return, and equilibrium: Empirical tests. Journal of Political Economy 81 (3), 607–636.

Fan, J., Furger, A., Xiu, D., 2016a. Incorporating global industrial classification standard into portfolio allocation: A simple factor-based large covariance matrix estimator with high-frequency data. Journal of Business and Economic Statistics 34 (4), 489–503.

Fan, J., Liao, Y., Liu, H., 2016b. An overview of the estimation of large covariance and precision matrices. The Econometrics Journal 19 (1).

Fan, J., Liao, Y., Mincheva, M., 2011. High dimensional covariance matrix estimation in approximate factor models. Annals of Statistics 39 (6), 3320.

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

Gagliardini, P., Ossola, E., Scaillet, O., 2016. Time-varying risk premium in large cross-sectional equity data sets. Econometrica 84 (3), 985–1046.

Hall, P., Heyde, C. C., 1980. Martingale Limit Theory and Its Application. Academic Press, Boston.

Hansen, L. P., Richard, S. F., 1987. The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. Econometrica: Journal of the Econometric Society , 587–613.

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

Jacod, J., Protter, P., 1998. Asymptotic error distributions for the euler method for stochastic differential equations. Annals of Probability 26, 267–307.

Jacod, J., Protter, P., 2012. Discretization of Processes, 1st Edition. Springer-Verlag, New York.

Jacod, J., Rosenbaum, M., 2013. Quarticity and other functionals of volatility: efficient estimation. The Annals of Statistics 41 (3), 1462–1484.

Jacod, J., Shiryaev, A. N., 2003. Limit Theorems for Stochastic Processes, 2nd Edition. Springer-Verlag, New York.

Jagannathan, R., Wang, Z., 1996. The conditional capm and the cross-section of expected returns. The Journal of finance 51 (1), 3–53.

Kalnina, I., 2012. Nonparametric tests of time variation in betas. Université de Montréal .

Kong, X.-B., Liu, C., 2018. Testing against constant factor loading matrix with large panel high-frequency data. Journal of Econometrics .

Lintner, J., 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. Review of Economics and Statistics 47, 13–37.

Markowitz, H., 1952. Portfolio selection. The journal of finance 7 (1), 77–91.

Markowitz, H., 1959. Portfolio Selection. New York.

McLeish, D. L., 1975. A maximal inequality and dependent strong laws. Annals of Probability 3 (5), 829–839.

Mykland, P. A., Zhang, L., 2006. Anova for diffusions and ito processes. The Annals of Statistics 34 (4), 1931–1963.

Mykland, P. A., Zhang, L., 2008. Inference for volatility-type objects and implications for hedging. Statistics and its Interface 1, 255–278.

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

Mykland, P. A., Zhang, L., 2012. The econometrics of high frequency data. In: Kessler, M., Lindner, A., Sørensen, M. (Eds.), Statistical Methods for Stochastic Differential Equations. Chapman and Hall/CRC Press, New York, pp. 109–190.

Mykland, P. A., Zhang, L., 2017. Assessment of uncertainty in high frequency data: The observed asymptotic variance. Econometrica 85 (1), 197–231.

Mykland, P. A., Zhang, L., Chen, D., 2019. The algebra of two scales estimation, and the s-tsrv: high frequency estimation that is robust to sampling times. Journal of Econometrics 208 (1), 101–119.

Patton, A. J., Verardo, M., 07 2012. Does Beta Move with News? Firm-Specific Information Flows and Learning about Profitability. The Review of Financial Studies 25 (9), 2789–2839.
URL https://doi.org/10.1093/rfs/hhs073

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

Ross, S. A., 1976. The arbitrage theory of capital asset pricing. Journal of Economic Theory 13 (3), 341–360.

Rothman, A. J., Levina, E., Zhu, J., 2009. Generalized thresholding of large covariance matrices. Journal of the American Statistical Association 104 (485), 177–186.

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

Todorov, V., Bollerslev, T., 2010. Jumps and betas: A new framework for disentangling and estimating systematic risks. Journal of Econometrics 157 (2), 220–235.

Vijh, A. M., 1994. S&P 500 Trading Strategies and Stock Betas. The Review of Financial Studies 7 (1), 215–251.

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

Zhang, L., 2012. Implied and realized volatility: Empirical model selection. Annals of Finance 8, 259–275.

Zhang, L., Mykland, P. A., Aït-Sahalia, Y., 2005. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. Journal of the American Statistical Association 100, 1394–1411.

Appendices

A. Proof of Lemma 1

(i). Plugging (2.13) into (2.19), we obtain:

$$\breve{\pi}^{(r,s)}_{\Delta T_n,t} = \frac{1}{\Delta T_n} \sum_{i=N^*(t)+1}^{N^*(t+\Delta T_n)} \tilde{B}^{(r,s)}_i$$

with

$$\tilde{B}^{(r,s)}_i = \left(\sum_{p=1}^{K-J-1} \left(\frac{K-J-p}{K-J} \right) \Delta X^{(r)}_{\tau_{i-p}} \right) \Delta X^{(s)}_{\tau_i} + \frac{1}{(K-J)} \left(\bar{\epsilon}^{(r)}_{i-J} - \bar{\epsilon}^{(r)}_{i-K} \right) \bar{\epsilon}^{(s)}_i .$$

Then we have:

$$\breve{\pi}^{(r_1,s_1)}_{\Delta T_n,t} \breve{\pi}^{(r_2,s_2)}_{\Delta T_n,t} = \frac{1}{\Delta T_n^2} \sum_{i=N^*(t)+1}^{N^*(t+\Delta T_n)} \tilde{B}^{(r_1,s_1)}_i \tilde{B}^{(r_2,s_2)}_i + \varpi_{\Delta T_n,t} \text{ with } \varpi_{\Delta T_n,t} = \frac{1}{\Delta T_n^2} \sum_{i=N^*(t)+2}^{N^*(t+\Delta T_n)} \left(\sum_{l=1}^{i-N^*(t)-1} \tilde{B}^{(r_1,s_1)}_{i-l} \right) \tilde{B}^{(r_2,s_2)}_i [2]. \tag{A.1}$$

Therefore, $E\left(\breve{\pi}^{(r_1,s_1)}_{\Delta T_n,t} \breve{\pi}^{(r_2,s_2)}_{\Delta T_n,t} \right) = \frac{1}{\Delta T_n^2} \sum_{i=N^*(t)+1}^{N^*(t+\Delta T_n)} E\left[\tilde{B}^{(r_1,s_1)}_i \tilde{B}^{(r_2,s_2)}_i \right]$, where

$$E\left[\tilde{B}^{(r_1,s_1)}_i \tilde{B}^{(r_2,s_2)}_i \right] = O_p\left(\Delta\tau_{n,i} (K-J) \Delta\tau_n^+ + \frac{1}{(K-J)^2 \left(\mathcal{M}_n^-\right)^2} \right), \tag{A.2}$$

and finally, we obtain:

$$E\left(\breve{\pi}^{(r_1,s_1)}_{\Delta T_n,t} \breve{\pi}^{(r_2,s_2)}_{\Delta T_n,t} \right) = O_p\left(\Delta T_n^{-1} (K-J) \Delta\tau_n^+ + \frac{N^*(t+\Delta T_n) - N^*(t)}{\Delta T_n^2 (K-J)^2 \left(\mathcal{M}_n^-\right)^2} \right)$$

uniformly. By Condition 2, we know that $N^*(t+\Delta T_n) - N^*(t) \sim N \Delta T_n^{-1}$ which implies that

$$\frac{N^*(t+\Delta T_n) - N^*(t)}{\Delta T_n^2 (K-J)^2 \left(\mathcal{M}_n^-\right)^2} \sim \frac{N}{\Delta T_n (K-J)^2 \left(\mathcal{M}_n^-\right)^2} = O\left(a_n^2 \Delta T_n^{-1} \right).$$

Therefore, we obtain (2.20).

On the other hand, we know that $\breve{\pi}^{(r_1,s_1)}_{\Delta T_n,t} \breve{\pi}^{(r_2,s_2)}_{\Delta T_n,t} - E\left(\breve{\pi}^{(r_1,s_1)}_{\Delta T_n,t} \breve{\pi}^{(r_2,s_2)}_{\Delta T_n,t} \right)$ has the same order as $\varpi_{\Delta T_n,t}$, which is defined in (A.1). In what follows, we prove $\|\varpi_{\Delta T_n,t}\|_2 = O_p\left(a_n^2 \Delta T_n^{-1}\right)$. Note that

$$
\begin{aligned}
E\left[\varpi^2_{\Delta T_n,t}\right] &= \frac{1}{\Delta T_n^4} E\left[\left(\sum_{i=N^*(t)+2}^{N^*(t+\Delta T_n)} \left(\sum_{l=1}^{i-N^*(t)-1} \tilde{B}_{i-l}^{(r_1,s_1)}\right) \tilde{B}_i^{(r_2,s_2)}[2]\right)^2\right] \\
&= \frac{1}{\Delta T_n^4} \sum_{i=N^*(t)+2}^{N^*(t+\Delta T_n)} E\left[\left(\sum_{l=1}^{i-N^*(t)-1} \tilde{B}_{i-l}^{(r_1,s_1)}\right)^2 \left(\tilde{B}_i^{(r_2,s_2)}\right)^2 [2]\right] \\
&= \frac{1}{\Delta T_n^4} \sum_{i=N^*(t)+2}^{N^*(t+\Delta T_n)} \sum_{l=1}^{i-N^*(t)-1} E\left[\left(\tilde{B}_{i-l}^{(r_1,s_1)}\right)^2 \left(\tilde{B}_i^{(r_2,s_2)}\right)^2\right] [2], \qquad \text{(A.3)}
\end{aligned}
$$

where by (A.2),

$$
E\left[\left(\tilde{B}_{i-l}^{(r_1,s_1)}\right)^2 \left(\tilde{B}_i^{(r_2,s_2)}\right)^2\right][2] = O_p\left(\left(\Delta\tau_{n,i}\left(K-J\right)\Delta\tau_n^+ + \frac{1}{\left(K-J\right)^2\left(\mathcal{M}_n^-\right)^2}\right)^2\right). \qquad \text{(A.4)}
$$

Then it is straightforward to see that $\sup_t E\left[\varpi^2_{\Delta T_n,t}\right] = O_p\left(a_n^4 \Delta T_n^{-2}\right)$. Therefore, we obtain (2.21).

(ii). Recall the formulas (2.16), (2.18), (2.19) and (2.12), we have:

$$
\tilde{\pi}^{(r,s)}_{\Delta T_n,t} = \breve{\pi}^{(r,s)}_{\Delta T_n,t} + \frac{1}{\Delta T_n}\left(\tilde{e}^{(r,s)}_{t+\Delta T_n} - \tilde{e}^{(r,s)}_t\right).
$$

Moreover, by direct calculation, we have:

$$
\eta^{(r_1,s_1,r_2,s_2)}_{\Delta T_n,t} = \frac{1}{\Delta T_n}\breve{\pi}^{(r_1,s_1)}_{\Delta T_n,t}\left(\tilde{e}^{(r_2,s_2)}_{t+\Delta T_n} - \tilde{e}^{(r_2,s_2)}_t\right)[2] + \frac{1}{\Delta T_n^2}\left(\tilde{e}^{(r_1,s_1)}_{t+\Delta T_n} - \tilde{e}^{(r_1,s_1)}_t\right)\left(\tilde{e}^{(r_2,s_2)}_{t+\Delta T_n} - \tilde{e}^{(r_2,s_2)}_t\right), \qquad \text{(A.5)}
$$

where [2] denotes the summation by switching (r_1,s_1) and (r_2,s_2). Because $\sup_{t,r,s}\left\|\tilde{e}^{(r,s)}_t\right\|_2 = O_p\left(a_n^2\right)$, then we obtain (2.22). Based on (2.20), we have $\sup_{t,r,s}\left\|\breve{\pi}^{(r,s)}_{\Delta T_n,t}\right\|_2 = O_p\left(a_n \Delta T_n^{-1/2}\right)$. Applying Cauchy-Schwarz inequality on (A.5), we obtain: $\sup_{t,r_1,r_2,s_1,s_2}\left\|\eta^{(r_1,s_1,r_2,s_2)}_{\Delta T_n,t}\right\|_2 = O_p\left(a_n^3 \Delta T_n^{-3/2}\right)$. By Minkowski inequality, we have:

$$
\left\|\eta^{(r_1,s_1,r_2,s_2)}_{\Delta T_n,t} - E\left(\eta^{(r_1,s_1,r_2,s_2)}_{\Delta T_n,t}\right)\right\|_2 \le \left\|\eta^{(r_1,s_1,r_2,s_2)}_{\Delta T_n,t}\right\|_2 + \left\|E\left(\eta^{(r_1,s_1,r_2,s_2)}_{\Delta T_n,t}\right)\right\|_2,
$$

thus (2.23) got proved. □

B. Proof of Lemma 2

The estimation error can be decomposed as in (3.7). Using integration by parts, we obtain

$$
\int_{T_{i-1}}^{T_i}\left(\beta_s - \beta_{T_{i-1}}\right) ds = \int_{T_{i-1}}^{T_i}\left(T_i - s\right) d\beta_s.
$$

Thus, $\left\|\int_{T_{i-1}}^{T_i}\left(\beta_s-\beta_{T_{i-1}}\right)ds\right\|_2=O_p\left(\Delta T_n^{3/2}\right)$ which implies that

$$\mathbf{R}^{\text{Discrete}}=O_p\left(\Delta T_n\right). \tag{B.1}$$

On the other hand, because the estimation error of spot beta can be further decomposed as:

$$\begin{aligned}\bar{\boldsymbol{\beta}}_{\Delta T_n,T_{i-1}}-\boldsymbol{\beta}_{T_{i-1}} &= \left(\bar{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\left(\bar{\pi}^{\mathbf{X},Y}_{\Delta T_n,T_{i-1}}-\bar{\pi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\boldsymbol{\beta}_{T_{i-1}}\right), &\text{(B.2)}\\ \hat{\boldsymbol{\beta}}_{\Delta T_n,T_{i-1}}-\bar{\boldsymbol{\beta}}_{\Delta T_n,T_{i-1}} &= \left(\hat{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\left(\tilde{\pi}^{\mathbf{X},Y}_{\Delta T_n,T_{i-1}}-\tilde{\pi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\bar{\boldsymbol{\beta}}_{\Delta T_n,T_{i-1}}\right), &\text{(B.3)}\end{aligned}$$

where $\bar{c},\hat{c},\bar{\pi},\tilde{\pi}$ and $\bar{\beta}$ are defined in (2.16), (2.18), and (3.6).

Again from integration by parts, we have:

$$\bar{\pi}^{\mathbf{X},Y}_{\Delta T_n,t}=\bar{c}^{\mathbf{X},Y}_{\Delta T_n,t}-c^{\mathbf{X},Y}_t=\frac{1}{\Delta T_n}\int_t^{t+\Delta T_n}\left(\left(t+\Delta T_n\right)-s\right)dc^{\mathbf{X},Y}_s,$$

which still holds for $\bar{\pi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,t}$ by switching the superscript from "X,Y" to "X,X" in above equation. Following the same techniques in Lemma 4 of Mykland and Zhang (2006), we obtain that $\sup_t\left\|\bar{\pi}^{\mathbf{X},Y}_{\Delta T_n,t}\right\|_2=O_p\left(\Delta T_n^{1/2}\right)$ and $\sup_t\left\|\bar{\pi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,t}\right\|_2=O_p\left(\Delta T_n^{1/2}\right)$, and moreover, for any $\varepsilon>0$,

$$\sup_t\left\|\bar{c}^{\mathbf{X},Y}_{\Delta T_n,t}-c^{\mathbf{X},Y}_t\right\|_2=O_p\left(\Delta T_n^{1/2-\varepsilon}\right)\ \text{ and }\ \sup_t\left\|\bar{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,t}-c^{\mathbf{X},\mathbf{X}}_t\right\|_2=O_p\left(\Delta T_n^{1/2-\varepsilon}\right). \tag{B.4}$$

Similar to Corollary 1 in Mykland and Zhang (2006), since $\inf_t\lambda_{\min}\left(c^{\mathbf{X},\mathbf{X}}_t\right)>0$ (Condition 4), we obtain:

$$\sup_{n,i}\left\|\bar{\boldsymbol{\beta}}_{\Delta T_n,T_{i-1}}-\boldsymbol{\beta}_{T_{i-1}}\right\|=O_p\left(\Delta T_n^{1/2-\varepsilon}\right) \tag{B.5}$$

since the finite dimensional vector norms are equivalent. – On the other hand, by formula (2.20) in Lemma 1, we know that $\sup_t\left\|\tilde{\pi}^{\mathbf{X},Y}_{\Delta T_n,t}\right\|_2=O_p\left(a_n\Delta T_n^{-1/2}\right)$ and $\sup_t\left\|\tilde{\pi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,t}\right\|_2=O_p\left(a_n\Delta T_n^{-1/2}\right)$. Applying Lemma 1 in Mykland and Zhang (2006), we obtain:

$$\sup_t\left\|\hat{c}^{\mathbf{X},Y}_{\Delta T_n,t}-\bar{c}^{\mathbf{X},Y}_t\right\|_2=O_p\left(\left(a_n^2\Delta T_n^{-1}\right)^{1/2-\varepsilon}\right)\ \text{ and }\ \sup_t\left\|\hat{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,t}-\bar{c}^{\mathbf{X},\mathbf{X}}_t\right\|_2=O_p\left(\left(a_n^2\Delta T_n^{-1}\right)^{1/2-\varepsilon}\right). \tag{B.6}$$

Then combining (B.4)-(B.6) and B.3, by Taylor expansion, we obtain:

$$\sup_{n,i}\left\|\hat{\boldsymbol{\beta}}_{\Delta T_n,T_{i-1}}-\bar{\boldsymbol{\beta}}_{\Delta T_n,T_{i-1}}\right\|=O_p\left(\left(a_n^2\Delta T_n^{-1}\right)^{1/2-\varepsilon}\right). \tag{B.7}$$

This implies that $\sup_{n,i}\left\|\hat{\boldsymbol{\beta}}_{\Delta T_n,T_{i-1}}-\boldsymbol{\beta}_{T_{i-1}}\right\|=o_p(1)$. Note that $\left\|\mathbf{R}^{\text{Spot}}\right\|_2\le\sup_{n,i}\left\|\hat{\boldsymbol{\beta}}_{\Delta T_n,T_{i-1}}-\boldsymbol{\beta}_{T_{i-1}}\right\|_2=o_p(1)$. Thus, this lemma has been proved.

C. Proof of Theorem 1

Before stating the second order behavior of $\hat{\theta}_n$, let's define a new type of observed covariation for processes Θ_t and Ξ_t:

$$[\Theta, \Phi]_t^{(B)} \triangleq \sum_{T_i \leq t} \left(\Theta_{T_i} - \Theta_{T_{i-1}}\right)\left(\Phi_{T_i} - \Phi_{T_{i-1}}\right). \tag{C.1}$$

The statement in Remark 3 applies equally to this quadratic variation.

The estimation error may be decomposed as follows:

$$a_n^{-2}\Delta T_n\left(\hat{\theta}_n - \theta + a_n^2 \Delta T_n^{-1}\varphi_{\mathcal{T}}\right) = a_n^{-2}\Delta T_n\left(R^{\text{Spot-V}} + R^{\text{Spot-B}} - R^{\text{Discrete}}\right), \tag{C.2}$$

where R^{Discrete} is defined in (3.7), and

$$\begin{aligned} R^{\text{Spot-V}} &= R^{\text{Spot-V-I}} + R^{\text{Spot-V-II}}, & \text{(C.3)}\\ R^{\text{Spot-B}} &= R^{\text{Spot-B-I}} + R^{\text{Spot-B-II}} + R^{\text{Spot-B-III}}, & \text{(C.4)} \end{aligned}$$

and

$$\begin{aligned} R^{\text{Spot-V-I}} &= \sum_{i=1}^{B} \xi^{\text{Spot-V-I}}_{\Delta T_n, T_{i-1}}\Delta T_n, & \text{(C.5)}\\ R^{\text{Spot-V-II}} &= \sum_{i=1}^{B} \xi^{\text{Spot-V-II}}_{\Delta T_n, T_{i-1}}\Delta T_n, & \text{(C.6)}\\ R^{\text{Spot-B-I}} &= a_n^2\Delta T_n^{-1}\tilde{\varphi}_{\mathcal{T}} - \sum_{i=1}^{B}\left(\hat{c}^{X,X}_{\Delta T_n,T_{i-1}}\bar{c}^{X,X}_{\Delta T_n,T_{i-1}}\right)^{-1}\tilde{\pi}^{X,X}_{\Delta T_n,T_{i-1}}\left(\tilde{\pi}^{X,Y}_{\Delta T_n,T_{i-1}} - \bar{\beta}_{\Delta T_n,T_{i-1}}\tilde{\pi}^{X,X}_{\Delta T_n,T_{i-1}}\right)\Delta T_n, &\\ R^{\text{Spot-B-II}} &= a_n^2\Delta T_n^{-1}\varphi_{\mathcal{T}} - a_n^2\Delta T_n^{-1}\tilde{\varphi}_{\mathcal{T}}, & \text{(C.7)}\\ R^{\text{Spot-B-III}} &= \sum_{i=1}^{B}\left(\bar{\beta}_{\Delta T_n,T_{i-1}} - \beta_{T_{i-1}}\right)\Delta T_n, & \text{(C.8)} \end{aligned}$$

with

$$\begin{aligned} \xi^{\text{Spot-V-I}}_{\Delta T_n,T_{i-1}} &= \left(c^{X,X}_{T_{i-1}}\right)^{-1}\left(\breve{\pi}^{X,Y}_{\Delta T_n,T_{i-1}} - \beta_{T_{i-1}}\breve{\pi}^{X,X}_{\Delta T_n,T_{i-1}}\right),\\ \xi^{\text{Spot-V-II}}_{\Delta T_n,T_{i-1}} &= \hat{\beta}_{\Delta T_n,T_{i-1}} - \bar{\beta}_{\Delta T_n,T_{i-1}} + \left(\hat{c}^{X,X}_{\Delta T_n,T_{i-1}}\bar{c}^{X,X}_{\Delta T_n,T_{i-1}}\right)^{-1}\tilde{\pi}^{X,X}_{\Delta T_n,T_{i-1}}\left(\tilde{\pi}^{X,Y}_{\Delta T_n,T_{i-1}} - \bar{\beta}_{\Delta T_n,T_{i-1}}\tilde{\pi}^{X,X}_{\Delta T_n,T_{i-1}}\right) - \xi^{\text{Spot-V-I}}_{\Delta T_n,T_{i-1}},\\ \tilde{\varphi}_t &= n^{2\alpha}\left[\int_0^t \left(c^{X,X}_u\right)^{-2} d\left[M^{X,X}.M^{X,Y}\right]^{(B)}_u - \int_0^t \left(c^{X,X}_u\right)^{-2}\beta_u d\left[M^{X,X}.M^{X,X}\right]^{(B)}_u\right]. \end{aligned}$$

First of all, by Lemma 1 and definition (2.19), it is easy to see that

$$R^{\text{Spot-V-I}} = O_p\left(a_n\right). \tag{C.9}$$

Second, recall the definitions (2.18) and (3.5), by direct calculation, we obtain:

$$\begin{aligned}\hat{\beta}_{\Delta T_n,T_{i-1}}-\bar{\beta}_{\Delta T_n,T_{i-1}} &= \left(\bar{c}^{X,X}_{\Delta T_n,T_{i-1}}\right)^{-1}\left(\tilde{\pi}^{X,Y}_{\Delta T_n,T_{i-1}}-\bar{\beta}_{\Delta T_n,T_{i-1}}\tilde{\pi}^{X,X}_{\Delta T_n,T_{i-1}}\right)\\ &\quad -\left(\hat{c}^{X,X}_{\Delta T_n,T_{i-1}}\bar{c}^{X,X}_{\Delta T_n,T_{i-1}}\right)^{-1}\left(\tilde{\pi}^{X,X}_{\Delta T_n,T_{i-1}}\right)\left(\tilde{\pi}^{X,Y}_{\Delta T_n,T_{i-1}}-\bar{\beta}_{\Delta T_n,T_{i-1}}\tilde{\pi}^{X,X}_{\Delta T_n,T_{i-1}}\right),\end{aligned}$$

and thus, $\xi^{\text{Spot-V-II}}_{\Delta T_n,T_{i-1}}$ could be further simplied as:

$$\xi^{\text{Spot-V-II}}_{\Delta T_n,T_{i-1}}=\bar{\xi}^{\text{Spot-V-II}}_{\Delta T_n,T_{i-1}}-\tilde{\xi}^{\text{Spot-V-II}}_{\Delta T_n,T_{i-1}}.$$

where

$$\begin{aligned}\tilde{\xi}^{\text{Spot-V-II}}_{\Delta T_n,T_{i-1}} &= \left(\bar{c}^{X,X}_{\Delta T_n,T_{i-1}}c^{X,X}_{T_{i-1}}\right)^{-1}\bar{\pi}^{X,X}_{\Delta T_n,T_{i-1}}\left(\check{\pi}^{X,Y}_{\Delta T_n,T_{i-1}}-\beta_{T_{i-1}}\check{\pi}^{X,X}_{\Delta T_n,T_{i-1}}\right)\\ &\quad+\left(\bar{c}^{X,X}_{\Delta T_n,T_{i-1}}\right)^{-1}\left(\bar{\beta}_{\Delta T_n,T_{i-1}}-\beta_{T_{i-1}}\right)\check{\pi}^{X,X}_{\Delta T_n,T_{i-1}},\\ \bar{\xi}^{\text{Spot-V-II}}_{\Delta T_n,T_{i-1}} &= \frac{1}{\Delta T_n}\left(\bar{c}^{X,X}_{\Delta T_n,T_{i-1}}\right)^{-1}\left[\left(\tilde{e}^{X,Y}_{T_i}-\tilde{e}^{X,Y}_{T_{i-1}}\right)-\bar{\beta}_{\Delta T_n,T_{i-1}}\left(\tilde{e}^{X,X}_{T_i}-\tilde{e}^{X,X}_{T_{i-1}}\right)\right].\end{aligned}$$

Denote by $\tilde{R}^{\text{Spot-V-II}}=\sum_{i=1}^{B}\tilde{\xi}^{\text{Spot-V-II}}_{\Delta T_n,T_{i-1}}\Delta T_n$ and $\bar{R}^{\text{Spot-V-II}}=\sum_{i=1}^{B}\bar{\xi}^{\text{Spot-V-II}}_{\Delta T_n,T_{i-1}}\Delta T_n$, then we know that

$$R^{\text{Spot-V-II}}=\bar{R}^{\text{Spot-V-II}}-\tilde{R}^{\text{Spot-V-II}}. \tag{C.10}$$

On one hand, by the results of Lemma 1 and eq. (B.5)-(B.7) we know that $\sup_i\left|E\left(\tilde{\xi}^{\text{Spot-V-II}}_{\Delta T_n,T_{i-1}}\right)\right|=o_p\left(a_n\right)$ and by Cauchy-Swartz inequality, we have $\sup_i\text{Var}\left(\tilde{\xi}^{\text{Spot-V-II}}_{\Delta T_n,T_{i-1}}\right)=O_p\left(a_n^2\right)$. Based on this fact, we know that

$$\tilde{R}^{\text{Spot-V-II}}=o_p\left(a_n\right). \tag{C.11}$$

On the other hand, note that $\bar{R}^{\text{Spot-V-II}}$ could be rewrite as follows:

$$\begin{aligned}\bar{R}^{\text{Spot-V-II}} &= \sum_{i=1}^{B-1}\left(\bar{c}^{X,X}_{\Delta T_n,T_i}\bar{c}^{X,X}_{\Delta T_n,T_{i-1}}\right)^{-1}\left(\bar{c}^{X,X}_{\Delta T_n,T_i}-\bar{c}^{X,X}_{\Delta T_n,T_{i-1}}\right)\left(\tilde{e}^{X,Y}_{T_i}-\bar{\beta}_{\Delta T_n,T_i}\tilde{e}^{X,X}_{T_i}\right)\\ &\quad+\sum_{i=1}^{B-1}\left(\bar{c}^{X,X}_{\Delta T_n,T_{i-1}}\right)^{-1}\left(\bar{\beta}_{\Delta T_n,T_i}-\bar{\beta}_{\Delta T_n,T_{i-1}}\right)\tilde{e}^{X,X}_{T_i}+O_p\left(a_n^2\right).\end{aligned} \tag{C.12}$$

Because

$$\bar{c}^{X,X}_{\Delta T_n,T_i}-\bar{c}^{X,X}_{\Delta T_n,T_{i-1}}=\int_{i\Delta T_n}^{(i+1)\Delta T_n}\left(\frac{T_{i+1}-u}{\Delta T_n}\right)dc^{X,X}_u+\int_{(i-1)\Delta T_n}^{i\Delta T_n}\left(\frac{u-T_{i-1}}{\Delta T_n}\right)dc^{X,X}_u,$$

we know that $\left\|\bar{c}^{X,X}_{\Delta T_n,T_i}-\bar{c}^{X,X}_{\Delta T_n,T_{i-1}}\right\|_2=O_p\left(\Delta T_n^{1/2}\right)$ and by Tayor expansion, we have $\left\|\bar{\beta}_{\Delta T_n,T_i}-\bar{\beta}_{\Delta T_n,T_{i-1}}\right\|_2=O_p\left(\Delta T_n^{1/2}\right)$. Based on Cauchy-Swartz inequality and formula (C.12), we obtain:

$$\bar{R}^{\text{Spot-V-II}}=O_p\left(\Delta T_n^{-1/2}n^{-2\alpha}\right)=o_p\left(a_n\right). \tag{C.13}$$

Plugging (C.11) and (C.13) into (C.10),

$$R^{\text{Spot-V-II}} = o_p\left(a_n\right), \tag{C.14}$$

and by plugging (C.9) and (C.14) into (C.3), we obtain:

$$R^{\text{Spot-V}} = R^{\text{Spot-V-I}} + o_p\left(a_n\right) = O_p\left(a_n\right). \tag{C.15}$$

Next, we calculate the order of $R^{\text{Spot-B-I}}$. Recall the definitions (2.19) and (C.1), we could rewrite $\tilde{\varphi}_{\mathcal{T}}$ as follows:

$$\tilde{\varphi}_{\mathcal{T}} = a_n^{-2}\Delta T_n \sum_{i=1}^{B}\left(c_{T_{i-1}}^{X,X}\right)^{-2}\left(\breve{\pi}_{\Delta T_n,T_{i-1}}^{X,X}\breve{\pi}_{\Delta T_n,T_{i-1}}^{X,Y} - \beta_{T_{i-1}}\left(\breve{\pi}_{\Delta T_n,T_{i-1}}^{X,X}\right)^2\right)\Delta T_n,$$

thus, we know that

$$R^{\text{Spot-B-I}} = \sum_{i=1}^{B}\xi_{\Delta T_n,T_{i-1}}^{\text{Spot-B-I}}\Delta T_n,$$

with

$$\begin{aligned}\xi_{\Delta T_n,T_{i-1}}^{\text{Spot-B-I}} \quad = \quad & \left(c_{T_{i-1}}^{X,X}\right)^{-2}\left(\breve{\pi}_{\Delta T_n,T_{i-1}}^{X,X}\breve{\pi}_{\Delta T_n,T_{i-1}}^{X,Y} - \beta_{T_{i-1}}\left(\breve{\pi}_{\Delta T_n,T_{i-1}}^{X,X}\right)^2\right)\\ & -\left(\hat{c}_{\Delta T_n,T_{i-1}}^{X,X}\bar{c}_{\Delta T_n,T_{i-1}}^{X,X}\right)^{-1}\tilde{\pi}_{\Delta T_n,T_{i-1}}^{X,X}\left(\tilde{\pi}_{\Delta T_n,T_{i-1}}^{X,Y} - \bar{\beta}_{\Delta T_n,T_{i-1}}\tilde{\pi}_{\Delta T_n,T_{i-1}}^{X,X}\right).\end{aligned}$$

By algebraic calculation, $\xi_{\Delta T_n,T_{i-1}}^{\text{Spot-B-I}}$ could be further simplified as follows:

$$\begin{aligned}\xi_{\Delta T_n,T_{i-1}}^{\text{Spot-B-I}} \quad = \quad & \left(c_{T_{i-1}}^{X,X}\right)^{-2}\left(\bar{\beta}_{\Delta T_n,T_{i-1}} - \beta_{T_{i-1}}\right)\left(\tilde{\pi}_{\Delta T_n,T_{i-1}}^{X,X}\right)^2 - \left(c_{T_{i-1}}^{X,X}\right)^{-2}\left(\eta_{\Delta T_n,T_{i-1}}^{(1,1,1,2)} - \beta_{T_{i-1}}\eta_{\Delta T_n,T_{i-1}}^{(1,1,1,1)}\right)\\ & +\left[\left(c_{T_{i-1}}^{X,X}\right)^{-2} - \left(\hat{c}_{\Delta T_n,T_{i-1}}^{X,X}\bar{c}_{\Delta T_n,T_{i-1}}^{X,X}\right)^{-1}\right]\left(\tilde{\pi}_{\Delta T_n,T_{i-1}}^{X,X}\tilde{\pi}_{\Delta T_n,T_{i-1}}^{X,Y} - \bar{\beta}_{\Delta T_n,T_{i-1}}\left(\tilde{\pi}_{\Delta T_n,T_{i-1}}^{X,X}\right)^2\right),\end{aligned}$$

where

$$\begin{aligned}\left(c_{T_{i-1}}^{X,X}\right)^{-2} - \left(\hat{c}_{\Delta T_n,T_{i-1}}^{X,X}\bar{c}_{\Delta T_n,T_{i-1}}^{X,X}\right)^{-1} \quad = \quad & \left(c_{T_{i-1}}^{X,X}\right)^{-2}\left(\bar{c}_{\Delta T_n,T_{i-1}}^{X,X}\right)^{-1}\bar{\pi}_{\Delta T_n,T_{i-1}}^{X,X}\\ & +\left(c_{T_{i-1}}^{X,X}\bar{c}_{\Delta T_n,T_{i-1}}^{X,X}\hat{c}_{\Delta T_n,T_{i-1}}^{X,X}\right)^{-1}\left(\tilde{\pi}_{\Delta T_n,T_{i-1}}^{X,X} + \bar{\pi}_{\Delta T_n,T_{i-1}}^{X,X}\right).\end{aligned}$$

By the results of Lemma 1 and eq. (B.5)-(B.7), we know that $\sup_i\left|E\left(\tilde{\xi}_{\Delta T_n,T_{i-1}}^{\text{Spot-B-I}}\right)\right| = O_p\left(a_n^4\Delta T_n^{-2}\right)$ and by Cauchy-Swartz inequality, we have $\sup_i \mathrm{Var}\left(\tilde{\xi}_{\Delta T_n,T_{i-1}}^{\text{Spot-B-I}}\right) = O_p\left(a_n^6\Delta T_n^{-3}\right)$. Then it is easy to see that

$$R^{\text{Spot-B-I}} = O_p\left(a_n^4\Delta T_n^{-2}\right) + O_p\left(a_n^3\Delta T_n^{-1}\right). \tag{C.16}$$

Next, for $R^{\text{Spot-B-II}}$, because of Remark 3 in Section 2.2, we know that $a_n^{-2}\left[M_n^{X,X}, M_n^{X,Y}\right]_u^{(B)} \xrightarrow{p} \left[M_\infty^{X,X}, M_\infty^{X,Y}\right]_u$

and $a_n^{-2}\left[M_n^{X,X},M_n^{X,X}\right]_u^{(B)} \xrightarrow{p} \left[M_\infty^{X,X}.M_\infty^{X,X}\right]_u$. Recall the definition (C.7), we have: $\varphi_{\mathcal{T}} \xrightarrow{p} \tilde{\varphi}_{\mathcal{T}}$, that is:

$$R^{\text{Spot-B-II}} = o_p\left(a_n^2 \Delta T_n^{-1}\right). \tag{C.17}$$

Note that

$$\bar{\beta}_{\Delta T_n,T_{i-1}} - \beta_{\Delta T_n,T_{i-1}} = \left(\bar{c}_{\Delta T_n,T_{i-1}}^{X,X}\right)^{-1}\left(\bar{\pi}_{\Delta T_n,T_{i-1}}^{X,Y} - \beta_{T_{i-1}}\bar{\pi}_{\Delta T_n,T_{i-1}}^{X,X}\right),$$

and recall the definition (C.8) and apply Lemma 1 and eq. (B.5)-(B.7), it is easy to see that

$$R^{\text{Spot-B-III}} = O_p\left(\Delta T_n\right). \tag{C.18}$$

Then plugging (C.16), (C.17) and (C.18) into (C.4), we obtain:

$$R^{\text{Spot-B}} = O_p\left(\Delta T_n\right) + o_p\left(a_n^2 \Delta T_n^{-1}\right). \tag{C.19}$$

Finally, plugging (B.1), (C.15) and (C.19) into (C.2), and when $a_n^{-2}\Delta T_n \to \infty$ and $a_n^{-1}\Delta T_n \to 0$ we have:

$$a_n^{-2}\Delta T_n\left(\hat{\theta}_n - \theta + a_n^2 \Delta T_n^{-1}\varphi_{\mathcal{T}}\right) = o_p(1).$$

The theorem is thus proved.

D. The Central Limit Theorem in the Scalar Case.

We here display the CLT in the case of one-dimensional X and Y, along with the proof of the result. The purpose of this section is to facilitate the passage to the multi-regressor case, particularly on the level of the proofs.

Theorem 5. *(Central Limit Theorem for $\tilde{\theta}_n$) Assume Conditions 1-3, that* $\inf_{0\le t\le \mathcal{T}} c_t^{\mathbf{X},\mathbf{X}} > 0$*, and also that* $a_n^{-1}\Delta T_n \to 0$ *and* $a_n^{-3/2}\Delta T_n \to \infty$*. We then have:*

$$a_n^{-1}\left(\tilde{\theta}_n - \theta\right) \xrightarrow{\mathscr{L}} N\left(0, \Sigma_{\mathcal{T}}\right),$$

stably, with

$$\Sigma_t \triangleq \int_0^t \left(c_u^{X,X}\right)^{-2}\left(d\left[M_\infty^{X,Y},M_\infty^{X,Y}\right]_u - 2\beta_u d\left[M_\infty^{X,Y},M_\infty^{X,X}\right]_u + \beta_u^2 d\left[M_\infty^{X,X},M_\infty^{X,X}\right]_u\right). \tag{D.1}$$

Proof of Theorem 5.

$$\tilde{\theta}_n - \theta = R^{\text{Spot-V}} + R^{\text{Adjusted-Spot-B}} - R^{\text{Discrete}}, \tag{D.2}$$

where $R^{\text{Spot-V}}$ and R^{Discrete} are defined in (C.3) and (3.7), respectively, and

$$R^{\text{Adjusted-Spot-B}} = R^{\text{Adjusted-Spot-B-I}} + R^{\text{Adjusted-Spot-B-II}} + R^{\text{Adjusted-Spot-B-III}}, \tag{D.3}$$

and

$$
\begin{aligned}
R^{\text{Adjusted-Spot-B-I}} &= \sum_{i=1}^{B} \xi^{\text{Adj-Spot-B-I}}_{\Delta T_n, T_{i-1}} \Delta T_n, \\
R^{\text{Adjusted-Spot-B-II}} &= \sum_{i=1}^{B} \xi^{\text{Adj-Spot-B-II}}_{\Delta T_n, T_{i-1}} \Delta T_n, \\
R^{\text{Adjusted-Spot-B-III}} &= R^{\text{Spot-B-III}},
\end{aligned} \tag{D.4}
$$

with

$$
\begin{aligned}
\xi^{\text{Adj-Spot-B-I}}_{\Delta T_n, T_{i-1}} &= \left(\hat{c}^{X,X}_{\Delta T_n, T_{n,i-1}}\right)^{-2} \tilde{\pi}^{X,X}_{\Delta T_n, T_{i-1}} \left(\tilde{\pi}^{X,Y}_{\Delta T_n, T_{i-1}} - \hat{\beta}_{\Delta T_n, T_{i-1}} \tilde{\pi}^{X,X}_{\Delta T_n, T_{i-1}}\right) \\
&\quad - \left(\hat{c}^{X,X}_{\Delta T_n, T_{i-1}} \bar{c}^{X,X}_{\Delta T_n, T_{i-1}}\right)^{-1} \tilde{\pi}^{X,X}_{\Delta T_n, T_{i-1}} \left(\tilde{\pi}^{X,Y}_{\Delta T_n, T_{i-1}} - \bar{\beta}_{\Delta T_n, T_{i-1}} \tilde{\pi}^{X,X}_{\Delta T_n, T_{i-1}}\right), \\
\xi^{\text{Adj-Spot-B-II}}_{\Delta T_n, T_{i-1}} &= \left(\hat{c}^{X,X}_{\Delta T_n, T_{n,i-1}}\right)^{-1} \left(\hat{\phi}^{X,X,X,Y}_{\Delta T_n, T_{n,i-1}} - \hat{\beta}_{\Delta T_n, T_{n,i-1}} \hat{\phi}^{X,X,X,X}_{\Delta T_n, T_{n,i-1}}\right) \\
&\quad - \left(\hat{c}^{X,X}_{\Delta T_n, T_{n,i-1}}\right)^{-2} \tilde{\pi}^{X,X}_{\Delta T_n, T_{i-1}} \left(\tilde{\pi}^{X,Y}_{\Delta T_n, T_{i-1}} - \hat{\beta}_{\Delta T_n, T_{i-1}} \tilde{\pi}^{X,X}_{\Delta T_n, T_{i-1}}\right),
\end{aligned}
$$

and $R^{\text{Spot-B-III}}$ being defined in (C.8).

Note that by direct calculation,

$$
\begin{aligned}
\xi^{\text{Adj-Spot-B-I}}_{\Delta T_n, T_{i-1}} &= -\left(\hat{c}^{X,X}_{\Delta T_n, T_{n,i-1}}\right)^{-2} \left(\bar{c}^{X,X}_{\Delta T_n, T_{i-1}}\right)^{-1} \left(\tilde{\pi}^{X,X}_{\Delta T_n, T_{i-1}}\right)^{2} \left(\tilde{\pi}^{X,Y}_{\Delta T_n, T_{i-1}} - \hat{\beta}_{\Delta T_n, T_{i-1}} \tilde{\pi}^{X,X}_{\Delta T_n, T_{i-1}}\right) \\
&\quad - \left(\hat{c}^{X,X}_{\Delta T_n, T_{i-1}} \bar{c}^{X,X}_{\Delta T_n, T_{i-1}}\right)^{-1} \left(\tilde{\pi}^{X,X}_{\Delta T_n, T_{i-1}}\right)^{2} \left(\hat{\beta}_{\Delta T_n, T_{i-1}} - \bar{\beta}_{\Delta T_n, T_{i-1}}\right).
\end{aligned}
$$

Based on (B.5)-(B.7), we know that $\sup_i \left|\xi^{\text{Adj-Spot-B-I}}_{\Delta T_n, T_{i-1}}\right| = O_p\left(a_n^4 \Delta T_n^{-2}\right)$ and $\sup_i \text{Var}\left(\xi^{\text{Adj-Spot-B-I}}_{\Delta T_n, T_{i-1}}\right) = O_p\left(a_n^6 \Delta T_n^{-3}\right)$. Then it is easy to see that

$$
R^{\text{Adjusted-Spot-B-I}} = O_p\left(a_n^4 \Delta T_n^{-2}\right) + O_p\left(a_n^3 \Delta T_n^{-1}\right). \tag{D.5}
$$

Next, we consider the size of $R^{\text{Adjusted-Spot-B-II}}$. Note that it could be simplified as:

$$
\xi^{\text{Adj-Spot-B-II}}_{\Delta T_n, T_{i-1}} = \left(\hat{c}^{X,X}_{\Delta T_n, T_{n,i-1}}\right)^{-2} \left(\tilde{\eta}^{X,X,X,Y}_{\Delta T_n, T_{n,i-1}} - \hat{\beta}_{\Delta T_n, T_{n,i-1}} \tilde{\eta}^{X,X,X,X}_{\Delta T_n, T_{n,i-1}}\right),
$$

where $\tilde{\eta}^{X,X,X,X}_{\Delta T_n, T_{n,i-1}} = \tilde{\eta}^{(1,1,1,1)}_{\Delta T_n, T_{n,i-1}}$ and $\tilde{\eta}^{X,X,X,Y}_{\Delta T_n, T_{n,i-1}} = \tilde{\eta}^{(1,1,1,2)}_{\Delta T_n, T_{n,i-1}}$ with

$$
\tilde{\eta}^{(r_1,s_1,r_2,s_2)}_{\Delta T_n, T_{n,i-1}} \triangleq \hat{\varphi}^{(r_1,s_1,r_2,s_2)}_{\Delta T_n, T_{n,i-1}} - \tilde{\pi}^{(r_1,s_1)}_{\Delta T_n, T_{i-1}} \tilde{\pi}^{(r_2,s_2)}_{\Delta T_n, T_{i-1}}, \tag{D.6}
$$

and

$$
\hat{\varphi}^{(r_1,s_1,r_2,s_2)}_{\Delta T_n, T_{n,i-1}} \triangleq \check{\varphi}^{(r_1,s_1)}_{\Delta T_n, T_{n,i-1}} \check{\varphi}^{(r_2,s_2)}_{\Delta T_n, T_{n,i-1}}.
$$

To find the size of $\tilde{\eta}^{(r_1,s_1,r_2,s_2)}_{\Delta T_n,T_{n,i-1}}$, let's define two other quantities:

$$\begin{aligned}
\bar{\varphi}^{(r_1,s_1,r_2,s_2)}_{\Delta T_n,T_{n,i-1}} &= \frac{1}{4}\tilde{\pi}^{(r_1,s_1)}_{\Delta T_n/2,(i-1/2)\Delta T_n}\tilde{\pi}^{(r_2,s_2)}_{\Delta T_n/2,(i-1/2)\Delta T_n} + \frac{1}{4}\tilde{\pi}^{(r_1,s_1)}_{\Delta T_n/2,(i-1)\Delta T_n}\tilde{\pi}^{(r_2,s_2)}_{\Delta T_n/2,(i-1)\Delta T_n}, && \text{(D.7)}\\
\psi^{(r,s)}_{\Delta T_n,T_{n,i-1}} &= \bar{c}^{(r,s)}_{\Delta T_n/2,(i-1/2)\Delta T_n} - \bar{c}^{(r,s)}_{\Delta T_n/2,(i-1)\Delta T_n}. && \text{(D.8)}
\end{aligned}$$

Therefore, note that

$$\hat{c}^{(r,s)}_{\Delta T_n/2,(i-1/2)\Delta T_n} - \hat{c}^{(r,s)}_{\Delta T_n/2,(i-1)\Delta T_n} = \psi^{(r,s)}_{\Delta T_n,T_{n,i-1}} + \tilde{\pi}^{(r,s)}_{\Delta T_n/2,(i-1/2)\Delta T_n} - \tilde{\pi}^{(r,s)}_{\Delta T_n/2,(i-1)\Delta T_n},$$

then we could decompose $\tilde{\eta}^{(r_1,s_1,r_2,s_2)}_{\Delta T_n,T_{n,i-1}}$ as follows:

$$\tilde{\eta}^{(r_1,s_1,r_2,s_2)}_{\Delta T_n,T_{n,i-1}} = \left(\hat{\varphi}^{(r_1,s_1,r_2,s_2)}_{\Delta T_n,T_{n,i-1}} - \bar{\varphi}^{(r_1,s_1,r_2,s_2)}_{\Delta T_n,T_{n,i-1}}\right) + \left(\bar{\varphi}^{(r_1,s_1,r_2,s_2)}_{\Delta T_n,T_{n,i-1}} - \tilde{\pi}^{(r_1,s_1)}_{\Delta T_n,T_{i-1}}\tilde{\pi}^{(r_2,s_2)}_{\Delta T_n,T_{i-1}}\right),$$

where $\bar{\varphi}^{(r_1,s_1,r_2,s_2)}_{\Delta T_n,T_{n,i-1}}$ is defined in (D.7), and it is straightforward to obtain:

$$\begin{aligned}
&\tilde{\pi}^{(r_1,s_1)}_{\Delta T_n,T_{i-1}}\tilde{\pi}^{(r_2,s_2)}_{\Delta T_n,T_{i-1}} - \bar{\varphi}^{(r_1,s_1,r_2,s_2)}_{\Delta T_n,T_{n,i-1}}\\
=\ & \frac{1}{\Delta T_n^2}\left(M^{(r_1,s_1)}_{i\Delta T_n} - M^{(r_1,s_1)}_{(i-1/2)\Delta T_n}\right)\left(M^{(r_2,s_2)}_{(i-1/2)\Delta T_n} - M^{(r_2,s_2)}_{(i-1)\Delta T_n}\right)[2]\\
&+\frac{2}{\Delta T_n^2}\left(M^{(r_1,s_1)}_{i\Delta T_n} - M^{(r_1,s_1)}_{(i-1/2)\Delta T_n}\right)\left(\tilde{e}^{(r_2,s_2)}_{(i-1/2)\Delta T_n} - \tilde{e}^{(r_2,s_2)}_{(i-1)\Delta T_n}\right)[2]\\
&+\frac{1}{\Delta T_n^2}\left(\tilde{e}^{(r_1,s_1)}_{i\Delta T_n} - \tilde{e}^{(r_1,s_1)}_{(i-1/2)\Delta T_n}\right)\left(\tilde{e}^{(r_2,s_2)}_{(i-1/2)\Delta T_n} - \tilde{e}^{(r_2,s_2)}_{(i-1)\Delta T_n}\right)[2],
\end{aligned}$$

and

$$\begin{aligned}
\hat{\varphi}^{(r_1,s_1,r_2,s_2)}_{\Delta T_n,T_{n,i-1}} - \bar{\varphi}^{(r_1,s_1,r_2,s_2)}_{\Delta T_n,T_{n,i-1}} &= \frac{1}{4}\psi^{(r_1,s_1)}_{\Delta T_n,T_{n,i-1}}\psi^{(r_2,s_2)}_{\Delta T_n,T_{n,i-1}} - \frac{1}{4}\tilde{\pi}^{(r_1,s_1)}_{\Delta T_n/2,(i-1/2)\Delta T_n}\tilde{\pi}^{(r_2,s_2)}_{\Delta T_n/2,(i-1)\Delta T_n}[2]\\
&\quad+\frac{1}{4}\left(\tilde{\pi}^{(r_1,s_1)}_{\Delta T_n/2,(i-1/2)\Delta T_n} - \tilde{\pi}^{(r_1,s_1)}_{\Delta T_n/2,(i-1)\Delta T_n}\right)\psi^{(r_2,s_2)}_{\Delta T_n,T_{n,i-1}}[2],
\end{aligned}$$

where $\psi^{(r,s)}_{\Delta T_n,T_{n,i-1}}$ is defined in (D.8). Because we can further simplify $\psi^{(r,s)}_{\Delta T_n,T_{n,i-1}}$ as follows:

$$\psi^{(r,s)}_{\Delta T_n,T_{n,i-1}} = \int_{(i-1/2)\Delta T_n}^{i\Delta T_n}\left(\frac{T_{n,i}-u}{\Delta T_n/2}\right)dc^{(r,s)}_u + \int_{(i-1)\Delta T_n}^{(i-1/2)\Delta T_n}\left(\frac{u-T_{n,i-1}}{\Delta T_n/2}\right)dc^{(r,s)}_u,$$

then we know that $\sup_i\left\|\psi^{(r,s)}_i\right\|_2 = O_p\left(\Delta T_n^{1/2}\right)$. Combining the formula (B.6), we know that $\sup_i\left|\xi^{\text{Adj-Spot-B-II}}_{\Delta T_n,T_{i-1}}\right| = O_p\left(a_n^4\Delta T_n^{-2}\right)$ and $\sup_i\mathrm{Var}\left(\xi^{\text{Adj-Spot-B-II}}_{\Delta T_n,T_{i-1}}\right) = O_p\left(a_n^4\Delta T_n^{-2}\right)$, and finally we obtain:

$$R^{\text{Adjusted-Spot-B-II}} = O_p\left(a_n^4\Delta T_n^{-2}\right) + O_p\left(a_n^2\Delta T_n^{-1/2}\right). \tag{D.9}$$

By formulas (C.18) and (D.4), we have:

$$R^{\text{Adjusted-Spot-B-III}} = O_p\left(\Delta T_n\right). \tag{D.10}$$

Plugging (D.5), (D.9), (D.10), (D.3), (C.15) and (B.1) into (D.2), if we assume that $a_n^{-1}\Delta T_n \to 0$ and $a_n^{-3/2}\Delta T_n \to \infty$, we have:

$$\tilde{\theta}_n - \theta = R^{\text{Spot-V-I}} + o_p\left(a_n\right).$$

Recall the definition (C.3), we know that

$$\left\langle R^{\text{Spot-V-I}}, R^{\text{Spot-V-I}}\right\rangle_t = \int_0^t \left(c_u^{X,X}\right)^{-2}\left(d\left[M^{X,Y}, M^{X,Y}\right]_u^{(B)} - 2\beta_u d\left[M^{X,Y}, M^{X,X}\right]_u^{(B)} + \beta_u^2 d\left[M^{X,X}, M^{X,X}\right]_u^{(B)}\right),$$

and, by Remark 3 in Section 2.2, we have $a_n^{-2}\left\langle R^{\text{Spot-V-I}}, R^{\text{Spot-V-I}}\right\rangle_t \xrightarrow{p} \Sigma_t$, and thus the theorem is proved.

E. Proof of Theorem 2

The estimation error could be decomposed as:

$$\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta} = \mathbf{R}^{\text{Spot-V}} + \mathbf{R}^{\text{Adjusted-Spot-B}} - \mathbf{R}^{\text{Discrete}}, \tag{E.1}$$

where $\mathbf{R}^{\text{Discrete}}$ is defined in (3.7), and

$$\begin{aligned} \mathbf{R}^{\text{Spot-V}} &= \mathbf{R}^{\text{Spot-V-I}} + \mathbf{R}^{\text{Spot-V-II}}, && \text{(E.2)} \\ \mathbf{R}^{\text{Adjusted-Spot-B}} &= \mathbf{R}^{\text{Adjusted-Spot-B-I}} + \mathbf{R}^{\text{Adjusted-Spot-B-II}} + \mathbf{R}^{\text{Adjusted-Spot-B-III}}, && \text{(E.3)} \end{aligned}$$

and

$$\mathbf{R}^{\text{Spot-V-I}} = \sum_{i=1}^{B} \boldsymbol{\xi}_{\Delta T_n, T_{i-1}}^{\text{Spot-V-I}} \Delta T_n \text{ and } \mathbf{R}^{\text{Spot-V-II}} = \sum_{i=1}^{B} \boldsymbol{\xi}_{\Delta T_n, T_{i-1}}^{\text{Spot-V-II}} \Delta T_n,$$

$$\begin{aligned} \mathbf{R}^{\text{Adjusted-Spot-B-I}} &= \sum_{i=1}^{B} \boldsymbol{\xi}_{\Delta T_n, T_{i-1}}^{\text{Adj-Spot-B-I}} \Delta T_n \text{ and } \mathbf{R}^{\text{Adjusted-Spot-B-II}} = \sum_{i=1}^{B} \boldsymbol{\xi}_{\Delta T_n, T_{i-1}}^{\text{Adj-Spot-B-II}} \Delta T_n, \\ \mathbf{R}^{\text{Adjusted-Spot-B-III}} &= \sum_{i=1}^{B} \left(\bar{\boldsymbol{\beta}}_{\Delta T_n, T_{i-1}} - \boldsymbol{\beta}_{T_{i-1}}\right) \Delta T_n, && \text{(E.4)} \end{aligned}$$

with

$$
\begin{aligned}
\boldsymbol{\xi}^{\text{Spot-V-I}}_{\Delta T_n,T_{i-1}} &= \left(c^{\mathbf{X},\mathbf{X}}_{T_{i-1}}\right)^{-1}\left(\check{\pi}^{\mathbf{X},Y}_{\Delta T_n,T_{i-1}} - \check{\pi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\boldsymbol{\beta}_{T_{i-1}}\right), \\
\boldsymbol{\xi}^{\text{Spot-V-II}}_{\Delta T_n,T_{i-1}} &= \hat{\boldsymbol{\beta}}_{\Delta T_n,T_{i-1}} - \bar{\boldsymbol{\beta}}_{\Delta T_n,T_{i-1}} - \boldsymbol{\xi}^{\text{Spot-V-I}}_{\Delta T_n,T_{i-1}} \\
&\quad + \left(\hat{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\tilde{\pi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\left(\bar{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\left(\tilde{\pi}^{\mathbf{X},Y}_{\Delta T_n,T_{i-1}} - \tilde{\pi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\bar{\boldsymbol{\beta}}_{\Delta T_n,T_{i-1}}\right), \\
\boldsymbol{\xi}^{\text{Adj-Spot-B-I}}_{\Delta T_n,T_{i-1}} &= \left(\hat{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\tilde{\pi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\left(\hat{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\left(\tilde{\pi}^{\mathbf{X},Y}_{\Delta T_n,T_{i-1}} - \tilde{\pi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\hat{\boldsymbol{\beta}}_{\Delta T_n,T_{i-1}}\right) \\
&\quad - \left(\hat{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\tilde{\pi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\left(\bar{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\left(\tilde{\pi}^{\mathbf{X},Y}_{\Delta T_n,T_{i-1}} - \tilde{\pi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\bar{\boldsymbol{\beta}}_{\Delta T_n,T_{i-1}}\right), \\
\boldsymbol{\xi}^{\text{Adj-Spot-B-II}}_{\Delta T_n,T_{i-1}} &= \left(\hat{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\left(\hat{\phi}^{\mathbf{X},\mathbf{X},\mathbf{X},Y}_{\Delta T_n,T_{i-1}} - \hat{\phi}^{\mathbf{X},\mathbf{X},\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\hat{\boldsymbol{\beta}}_{\Delta T_n,T_{i-1}}\right) \\
&\quad - \left(\hat{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\tilde{\pi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\left(\hat{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\left(\tilde{\pi}^{\mathbf{X},Y}_{\Delta T_n,T_{i-1}} - \tilde{\pi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\hat{\boldsymbol{\beta}}_{\Delta T_n,T_{i-1}}\right).
\end{aligned}
$$

First of all, it is easy to see that

$$\mathbf{R}^{\text{Spot-V-I}} = O_p\left(a_n\right).$$

Note that,

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{\Delta T_n,T_{i-1}} - \bar{\boldsymbol{\beta}}_{\Delta T_n,T_{i-1}} &= \left(\bar{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\left(\tilde{\pi}^{\mathbf{X},Y}_{\Delta T_n,T_{i-1}} - \tilde{\pi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\bar{\boldsymbol{\beta}}_{\Delta T_n,T_{i-1}}\right) \\
&\quad - \left(\hat{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\tilde{\pi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\left(\bar{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\left(\tilde{\pi}^{\mathbf{X},Y}_{\Delta T_n,T_{i-1}} - \tilde{\pi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\bar{\boldsymbol{\beta}}_{\Delta T_n,T_{i-1}}\right),
\end{aligned}
$$

then $\boldsymbol{\xi}^{\text{Spot-V-II}}_{\Delta T_n,T_{i-1}}$ could be further simplified as:

$$\boldsymbol{\xi}^{\text{Spot-V-II}}_{\Delta T_n,T_{i-1}} = \bar{\boldsymbol{\xi}}^{\text{Spot-V-II}}_{\Delta T_n,T_{i-1}} - \tilde{\boldsymbol{\xi}}^{\text{Spot-V-II}}_{\Delta T_n,T_{i-1}}$$

where

$$
\begin{aligned}
\tilde{\boldsymbol{\xi}}^{\text{Spot-V-II}}_{\Delta T_n,T_{i-1}} &= \left(\bar{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\bar{\pi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\left(c^{\mathbf{X},\mathbf{X}}_{T_{i-1}}\right)^{-1}\left(\check{\pi}^{\mathbf{X},Y}_{\Delta T_n,T_{i-1}} - \check{\pi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\boldsymbol{\beta}_{T_{i-1}}\right) \\
&\quad + \left(\bar{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\check{\pi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\left(\bar{\boldsymbol{\beta}}_{\Delta T_n,T_{i-1}} - \boldsymbol{\beta}_{T_{i-1}}\right), \\
\bar{\boldsymbol{\xi}}^{\text{Spot-V-II}}_{\Delta T_n,T_{i-1}} &= \frac{1}{\Delta T_n}\left(\bar{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\left[\left(\tilde{e}^{\mathbf{X},Y}_{T_i} - \tilde{e}^{\mathbf{X},Y}_{T_{i-1}}\right) - \left(\tilde{e}^{\mathbf{X},\mathbf{X}}_{T_i} - \tilde{e}^{\mathbf{X},\mathbf{X}}_{T_{i-1}}\right)\bar{\boldsymbol{\beta}}_{\Delta T_n,T_{i-1}}\right].
\end{aligned}
$$

Denote by $\tilde{\mathbf{R}}^{\text{Spot-V-II}} = \sum_{i=1}^{B}\tilde{\boldsymbol{\xi}}^{\text{Spot-V-II}}_{\Delta T_n,T_{i-1}}\Delta T_n$ and $\bar{\mathbf{R}}^{\text{Spot-V-II}} = \sum_{i=1}^{B}\bar{\boldsymbol{\xi}}^{\text{Spot-V-II}}_{\Delta T_n,T_{i-1}}\Delta T_n$, then we know that

$$\mathbf{R}^{\text{Spot-V-II}} = \bar{\mathbf{R}}^{\text{Spot-V-II}} - \tilde{\mathbf{R}}^{\text{Spot-V-II}}. \tag{E.5}$$

Following the similar discussion in the single regressor case, it is easy to see that $\tilde{\mathbf{R}}^{\text{Spot-V-II}} = o_p\left(a_n\right)$. Moreover, because

$$\begin{aligned}
\bar{\mathbf{R}}^{\text{Spot-V-II}} &= \sum_{i=1}^{B-1}\left(\bar{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\left(\bar{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_i}-\bar{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)\left(\bar{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_i}\right)^{-1}\left(\tilde{e}^{\mathbf{X},Y}_{T_i}-\tilde{e}^{\mathbf{X},\mathbf{X}}_{T_i}\bar{\boldsymbol{\beta}}_{\Delta T_n,T_i}\right)\\
&\quad+\sum_{i=1}^{B}\left(\bar{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\tilde{e}^{\mathbf{X},\mathbf{X}}_{T_i}\left(\bar{\boldsymbol{\beta}}_{\Delta T_n,T_i}-\bar{\boldsymbol{\beta}}_{\Delta T_n,T_{i-1}}\right)+O_p\left(a_n^2\right),
\end{aligned}$$

and by similar technique as in the single regressor case, we know that $\bar{\mathbf{R}}^{\text{Spot-V-II}} = O_p\left(a_n^2\Delta T_n^{-1/2}\right)$. Therefore, we obtain:

$$\mathbf{R}^{\text{Spot-V-II}} = o_p\left(a_n\right).$$

That is:

$$\mathbf{R}^{\text{Spot-V}} = \mathbf{R}^{\text{Spot-V-I}} + o_p\left(a_n\right) = O_p\left(a_n\right).$$

Next, by direct calculation, we have:

$$\begin{aligned}
\boldsymbol{\xi}^{\text{Adj-Spot-B-I}}_{\Delta T_n,T_{i-1}} &= -\left[\left(\hat{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\tilde{\pi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right]^2\left(\hat{\boldsymbol{\beta}}_{\Delta T_n,T_{i-1}}-\bar{\boldsymbol{\beta}}_{\Delta T_n,T_{i-1}}\right)\\
&\quad-\left[\left(\hat{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\tilde{\pi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right]^2\left(\bar{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\left(\tilde{\pi}^{\mathbf{X},Y}_{\Delta T_n,T_{i-1}}-\tilde{\pi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\bar{\boldsymbol{\beta}}_{\Delta T_n,T_{i-1}}\right),
\end{aligned}$$

and comparing with the case of a single regressor, we have:

$$\mathbf{R}^{\text{Adjusted-Spot-B-I}} = O_p\left(a_n^4\Delta T_n^{-2}\right)+O_p\left(a_n^3\Delta T_n^{-1}\right).$$

On the other hand, note that $\boldsymbol{\xi}^{\text{Adj-Spot-B-II}}_{\Delta T_n,T_{i-1}}$ could be rewrite as follows:

$$\boldsymbol{\xi}^{\text{Adj-Spot-B-II}}_{\Delta T_n,T_{i-1}} = \left(\hat{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\left[\tilde{\boldsymbol{\eta}}^{\mathbf{X},\mathbf{X},\mathbf{X},Y}_{\Delta T_n,T_{i-1}}-\tilde{\boldsymbol{\eta}}^{\mathbf{X},\mathbf{X},\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\hat{\boldsymbol{\beta}}_{\Delta T_n,T_{i-1}}\right],$$

where

$$\begin{aligned}
\tilde{\boldsymbol{\eta}}^{\mathbf{X},\mathbf{X},\mathbf{X},Y}_{\Delta T_n,T_{i-1}} &= \hat{\phi}^{\mathbf{X},\mathbf{X},\mathbf{X},Y}_{\Delta T_n,T_{i-1}}-\tilde{\pi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\left(\hat{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\tilde{\pi}^{\mathbf{X},Y}_{\Delta T_n,T_{i-1}},\\
\tilde{\boldsymbol{\eta}}^{\mathbf{X},\mathbf{X},\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}} &= \hat{\phi}^{\mathbf{X},\mathbf{X},\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}-\tilde{\pi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\left(\hat{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\tilde{\pi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}.
\end{aligned}$$

Let's define another intermediate variable:

$$\bar{\phi}^{\mathbf{X},\mathbf{X},\mathbf{X},Y}_{\Delta T_n,T_{i-1}} = \tilde{\varphi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\left(\hat{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\tilde{\varphi}^{\mathbf{X},Y}_{\Delta T_n,T_{i-1}},$$

where $\tilde{\varphi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}} = \left\{\tilde{\varphi}^{(r,s)}_{\Delta T_n,T_{n,i-1}}\right\}_{1\leq r,s\leq q}$ and $\tilde{\varphi}^{\mathbf{X},Y}_{\Delta T_n,T_{i-1}} = \left\{\tilde{\varphi}^{(r,q+l)}_{\Delta T_n,T_{n,i-1}}\right\}_{1\leq r\leq q}$ with

$$\tilde{\varphi}^{(r,s)}_{\Delta T_n,T_{n,i-1}} = \frac{1}{2}\left(\tilde{\pi}^{(r,s)}_{\Delta T_n/2,(i-1/2)\Delta T_n} - \tilde{\pi}^{(r,s)}_{\Delta T_n/2,(i-1)\Delta T_n}\right).$$

Similarly we could define $\bar{\phi}^{\mathbf{X},\mathbf{X},\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}} = \tilde{\varphi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\left(\hat{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\tilde{\varphi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}$.

Note that

$$\tilde{\boldsymbol{\eta}}^{\mathbf{X},\mathbf{X},\mathbf{X},Y}_{\Delta T_n,T_{i-1}} = \left(\hat{\phi}^{\mathbf{X},\mathbf{X},\mathbf{X},Y}_{\Delta T_n,T_{i-1}} - \bar{\phi}^{\mathbf{X},\mathbf{X},\mathbf{X},Y}_{\Delta T_n,T_{i-1}}\right) + \left(\bar{\phi}^{\mathbf{X},\mathbf{X},\mathbf{X},Y}_{\Delta T_n,T_{i-1}} - \tilde{\pi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\left(\hat{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\tilde{\pi}^{\mathbf{X},Y}_{\Delta T_n,T_{i-1}}\right),$$

where recall the definition ψ in (D.8), and because $\check{\varphi}^{\mathbf{X},Y}_{\Delta T_n,T_{n,i-1}} - \tilde{\varphi}^{\mathbf{X},Y}_{\Delta T_n,T_{n,i-1}} = \frac{1}{2}\psi^{\mathbf{X},Y}_{\Delta T_n,T_{n,i-1}}$ and $\check{\varphi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{n,i-1}} - \tilde{\varphi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{n,i-1}} = \frac{1}{2}\psi^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{n,i-1}}$, we have:

$$\begin{aligned}\hat{\phi}^{\mathbf{X},\mathbf{X},\mathbf{X},Y}_{\Delta T_n,T_{i-1}} - \bar{\phi}^{\mathbf{X},\mathbf{X},\mathbf{X},Y}_{\Delta T_n,T_{i-1}} \quad = \quad & \frac{1}{2}\tilde{\varphi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{n,i-1}}\left(\hat{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\psi^{\mathbf{X},Y}_{\Delta T_n,T_{n,i-1}} + \frac{1}{2}\psi^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{n,i-1}}\left(\hat{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\tilde{\varphi}^{\mathbf{X},Y}_{\Delta T_n,T_{i-1}} \\ & +\frac{1}{4}\psi^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{n,i-1}}\left(\hat{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\psi^{\mathbf{X},Y}_{\Delta T_n,T_{n,i-1}},\end{aligned}$$

and moreover because $\tilde{\pi}^{(r,s)}_{\Delta T_n,T_{i-1}} = \frac{1}{2}\tilde{\pi}^{(r,s)}_{\Delta T_n/2,(i-1/2)\Delta T_n} + \frac{1}{2}\tilde{\pi}^{(r,s)}_{\Delta T_n/2,(i-1)\Delta T_n}$ then we have:

$$\begin{aligned}& \tilde{\pi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\left(\hat{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\tilde{\pi}^{\mathbf{X},Y}_{\Delta T_n,T_{i-1}} - \bar{\phi}^{\mathbf{X},\mathbf{X},\mathbf{X},Y}_{\Delta T_n,T_{i-1}} \\ = \quad & \frac{1}{2}\tilde{\pi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n/2,(i-1/2)\Delta T_n}\left(\hat{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\tilde{\pi}^{\mathbf{X},Y}_{\Delta T_n/2,(i-1)\Delta T_n} + \frac{1}{2}\tilde{\pi}^{\mathbf{X},\mathbf{X}}_{\Delta T_n/2,(i-1)\Delta T_n}\left(\hat{c}^{\mathbf{X},\mathbf{X}}_{\Delta T_n,T_{i-1}}\right)^{-1}\tilde{\pi}^{\mathbf{X},Y}_{\Delta T_n/2,(i-1/2)\Delta T_n}.\end{aligned}$$

Finally, comparing with the single regressor case, we obtain:

$$\mathbf{R}^{\text{Adjusted-Spot-B-II}} = O_p\left(a_n^4\Delta T_n^{-2}\right) + O_p\left(a_n^2\Delta T_n^{-1/2}\right), \tag{E.6}$$

and

$$\mathbf{R}^{\text{Adjusted-Spot-B-III}} = O_p\left(\Delta T_n\right). \tag{E.7}$$

Recall the results in Lemma 2, we have $\mathbf{R}^{\text{Discrete}} = O_p\left(\Delta T_n\right)$, then we have the final representation of the estimation error as:

$$\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta} = \mathbf{R}^{\text{Spot-V-I}} + o_p\left(a_n\right).$$

In order to find the representation of $\langle \mathbf{R}^{\text{Spot-V-I}}, \mathbf{R}^{\text{Spot-V-I}} \rangle_t$, We first define:

$$\begin{aligned}
\left[M^{\mathbf{X},Y}, M^{\mathbf{X},Y}\right]_t^{(r,k,B)} &\triangleq \sum_{T_i \le t} \left(M_{T_i}^{(r,q+l)} - M_{T_{i-1}}^{(r,q+l)}\right)\left(M_{T_i}^{(k,q+l)} - M_{T_{i-1}}^{(k,q+l)}\right), \\
\left[M^{\mathbf{X},Y}, M^{\mathbf{X},\mathbf{X}}\right]_t^{(r,k,B)} &\triangleq \sum_{T_i \le t} \left(M_{T_i}^{(r,q+l)} - M_{T_{i-1}}^{(r,q+l)}\right)\left(M_{T_i}^{\mathbf{X},\mathbf{X}} - M_{T_{i-1}}^{\mathbf{X},\mathbf{X}}\right)_{\bullet,k}, \\
\left[M^{\mathbf{X},\mathbf{X}}, M^{\mathbf{X},\mathbf{X}}\right]_t^{(r,k,B)} &\triangleq \sum_{T_i \le t} \left(M_{T_i}^{\mathbf{X},\mathbf{X}} - M_{T_{i-1}}^{\mathbf{X},\mathbf{X}}\right)_{r,\bullet}^{\intercal}\left(M_{T_i}^{\mathbf{X},\mathbf{X}} - M_{T_{i-1}}^{\mathbf{X},\mathbf{X}}\right)_{k,\bullet}.
\end{aligned} \tag{E.8}$$

It is easy to see that $\left[M^{\mathbf{X},Y}, M^{\mathbf{X},Y}\right]_t^{(r,k,B)}$ is a scalar process and $\left[M^{\mathbf{X},Y}, M^{\mathbf{X},\mathbf{X}}\right]_t^{(r,k,B)}$ is a $q \times 1$ column vector process and $\left[M^{\mathbf{X},\mathbf{X}}, M^{\mathbf{X},\mathbf{X}}\right]_t^{(r,k,B)}$ is a $q \times q$ matrix process. If we define $\mathbf{A}_t \triangleq \left(c_t^{\mathbf{X},\mathbf{X}}\right)^{-1}$, then:

$$\left\langle \mathbf{R}^{\text{Spot-V-I}}, \mathbf{R}^{\text{Spot-V-I}} \right\rangle_t = \int_0^t \mathbf{A}_u d\left[\mathbf{\Phi}, \mathbf{\Phi}\right]_u^{(B)} \mathbf{A}_u,$$

where $d\left[\mathbf{\Phi}, \mathbf{\Phi}\right]_u^{(B)} = \left\{d\left[\mathbf{\Phi}, \mathbf{\Phi}\right]_u^{(r,k,B)}\right\}_{1\le r,k\le q}$ is a $q \times q$ matrix and its $(r,k)-$th element is expressed as:

$$\begin{aligned}
d\left[\mathbf{\Phi}, \mathbf{\Phi}\right]_u^{(r,k,B)} = \; & d\left[M^{\mathbf{X},Y}, M^{\mathbf{X},Y}\right]_u^{(r,k,B)} - \boldsymbol{\beta}_u^{\intercal} d\left[M^{\mathbf{X},Y}, M^{\mathbf{X},\mathbf{X}}\right]_u^{(r,k,B)} \\
& - \boldsymbol{\beta}_u^{\intercal} d\left[M^{\mathbf{X},Y}, M^{\mathbf{X},\mathbf{X}}\right]_u^{(k,r,B)} + \boldsymbol{\beta}_u^{\intercal} d\left[M^{\mathbf{X},\mathbf{X}}, M^{\mathbf{X},\mathbf{X}}\right]_u^{(r,k,B)} \boldsymbol{\beta}_u,
\end{aligned}$$

then the $(r,k)-$th element of $\langle \mathbf{R}^{\text{Spot-V-I}}, \mathbf{R}^{\text{Spot-V-I}} \rangle_t$ can be represented as follows:

$$\left\langle \mathbf{R}^{\text{Spot-V-I}}, \mathbf{R}^{\text{Spot-V-I}} \right\rangle_t^{(r,k)} = \int_0^t \left(\mathbf{A}_u\right)_{\bullet,r}^{\intercal} d\left[\mathbf{\Phi}, \mathbf{\Phi}\right]_u^{(B)} \left(\mathbf{A}_u\right)_{\bullet,k}.$$

The proof of this representation is collected in Appendix F.

From Remark 3, we know that $a_n^{-2}\left[\mathbf{\Phi}, \mathbf{\Phi}\right]_t^{(B)} \xrightarrow{p} \mathbf{\Lambda}_t$, and thus it is easy to see that

$$a_n^{-2}\left\langle \mathbf{R}^{\text{Spot-V-I}}, \mathbf{R}^{\text{Spot-V-I}} \right\rangle_t \xrightarrow{p} \mathbf{\Sigma}_t.$$

The theorem is proved. □

F. Proof of Covariance Matrix Representation

Define $\mathbf{\Phi}_{T_i} = \left\{\Phi_{T_i}^{(k)}\right\}_{1\le k\le q}$ and

$$\mathbf{\Phi}_{T_i} = \left(M_{T_i}^{\mathbf{X},Y} - M_{T_{i-1}}^{\mathbf{X},Y}\right) - \left(M_{T_i}^{\mathbf{X},\mathbf{X}} - M_{T_{i-1}}^{\mathbf{X},\mathbf{X}}\right)\boldsymbol{\beta}_{T_{i-1}},$$

then we know that for $1 \leq k \leq q$,

$$\begin{aligned}
\Phi_{T_i}^{(k)} &= \left(M_{T_i}^{(k,q+l)} - M_{T_{i-1}}^{(k,q+l)}\right) - \sum_{v=1}^{q} \beta_{T_{i-1}}^{(v)} \left(M_{T_i}^{(k,v)} - M_{T_{i-1}}^{(k,v)}\right) \\
&= \left(M_{T_i}^{(k,q+l)} - M_{T_{i-1}}^{(k,q+l)}\right) - \boldsymbol{\beta}_{T_{i-1}}^{\intercal} \left(M_{T_i}^{\mathbf{X},\mathbf{X}} - M_{T_{i-1}}^{\mathbf{X},\mathbf{X}}\right)_{\bullet,k}.
\end{aligned}$$

Therefore, we have:

$$\left\langle \mathbf{R}^{\text{Spot-V-I}}, \mathbf{R}^{\text{Spot-V-I}} \right\rangle_t = \sum_{i=1}^{B} \mathbf{A}_{T_{i-1}} \left(\mathbf{\Phi}_{T_i} \mathbf{\Phi}_{T_i}^{\intercal}\right) \mathbf{A}_{T_{i-1}}.$$

Note that the $(r,k)-$th element of $\mathbf{\Phi}_{T_i}\mathbf{\Phi}_{T_i}^{\intercal}$ can be represented as:

$$\begin{aligned}
\left(\mathbf{\Phi}_{T_i}\mathbf{\Phi}_{T_i}^{\intercal}\right)^{(r,k)} &= \Phi_{T_i}^{(r)}\Phi_{T_i}^{(k)} \\
&= \left(M_{T_i}^{(r,q+l)} - M_{T_{i-1}}^{(r,q+l)}\right)\left(M_{T_i}^{(k,q+l)} - M_{T_{i-1}}^{(k,q+l)}\right) + \left(M_{T_i}^{\mathbf{X},\mathbf{X}} - M_{T_{i-1}}^{\mathbf{X},\mathbf{X}}\right)_{r,\bullet} \boldsymbol{\beta}_{T_{i-1}} \left(M_{T_i}^{\mathbf{X},\mathbf{X}} - M_{T_{i-1}}^{\mathbf{X},\mathbf{X}}\right)_{k,\bullet} \boldsymbol{\beta}_{T_{i-1}}, \\
&\quad -\boldsymbol{\beta}_{T_{i-1}}^{\intercal} \left[\left(M_{T_i}^{(r,q+l)} - M_{T_{i-1}}^{(r,q+l)}\right)\left(M_{T_i}^{\mathbf{X},\mathbf{X}} - M_{T_{i-1}}^{\mathbf{X},\mathbf{X}}\right)_{\bullet,k}\right][2]
\end{aligned}$$

where [2] denotes the summation by switching r and k. Because $\left(M_{T_i}^{\mathbf{X},\mathbf{X}} - M_{T_{i-1}}^{\mathbf{X},\mathbf{X}}\right)_{r,\bullet} \boldsymbol{\beta}_{T_{i-1}}$ is a scalar, then we know that it is same to its transpose, and therefore,

$$\begin{aligned}
\left(M_{T_i}^{\mathbf{X},\mathbf{X}} - M_{T_{i-1}}^{\mathbf{X},\mathbf{X}}\right)_{r,\bullet} \boldsymbol{\beta}_{T_{i-1}} \left(M_{T_i}^{\mathbf{X},\mathbf{X}} - M_{T_{i-1}}^{\mathbf{X},\mathbf{X}}\right)_{k,\bullet} \boldsymbol{\beta}_{T_{i-1}} &= \left(\left(M_{T_i}^{\mathbf{X},\mathbf{X}} - M_{T_{i-1}}^{\mathbf{X},\mathbf{X}}\right)_{r,\bullet} \boldsymbol{\beta}_{T_{i-1}}\right)^{\intercal} \left(M_{T_i}^{\mathbf{X},\mathbf{X}} - M_{T_{i-1}}^{\mathbf{X},\mathbf{X}}\right)_{k,\bullet} \boldsymbol{\beta}_{T_{i-1}} \\
&= \boldsymbol{\beta}_{T_{i-1}}^{\intercal} \left(M_{T_i}^{\mathbf{X},\mathbf{X}} - M_{T_{i-1}}^{\mathbf{X},\mathbf{X}}\right)_{r,\bullet}^{\intercal} \left(M_{T_i}^{\mathbf{X},\mathbf{X}} - M_{T_{i-1}}^{\mathbf{X},\mathbf{X}}\right)_{k,\bullet} \boldsymbol{\beta}_{T_{i-1}}.
\end{aligned}$$

Finally, we know that

$$\begin{aligned}
\Phi_{T_i}^{(r)}\Phi_{T_i}^{(k)} &= \left(M_{T_i}^{(r,q+l)} - M_{T_{i-1}}^{(r,q+l)}\right)\left(M_{T_i}^{(k,q+l)} - M_{T_{i-1}}^{(k,q+l)}\right) - \boldsymbol{\beta}_{T_{i-1}}^{\intercal} \left[\left(M_{T_i}^{(r,q+l)} - M_{T_{i-1}}^{(r,q+l)}\right)\left(M_{T_i}^{\mathbf{X},\mathbf{X}} - M_{T_{i-1}}^{\mathbf{X},\mathbf{X}}\right)_{\bullet,k}\right][2] \\
&\quad +\boldsymbol{\beta}_{T_{i-1}}^{\intercal} \left(M_{T_i}^{\mathbf{X},\mathbf{X}} - M_{T_{i-1}}^{\mathbf{X},\mathbf{X}}\right)_{r,\bullet}^{\intercal} \left(M_{T_i}^{\mathbf{X},\mathbf{X}} - M_{T_{i-1}}^{\mathbf{X},\mathbf{X}}\right)_{k,\bullet} \boldsymbol{\beta}_{T_{i-1}}.
\end{aligned}$$

G. Proof of Theorems 3 and 4

Before the proof of main theorems, we first show some preliminary lemmas. We note that Lemma 3 shows eq. 4.12 in the Theorem, by replacing X by Y.

Lemma 3. *We define $\hat{c}_t = \left\{\hat{c}_{\Delta T_n,t}^{(r,s)}\right\}_{1\leq r,s\leq q+d}$ with $\Delta T_n \asymp a_n$ and $q = o(d)$. We assume Conditions 1-3. Then the elementwise max norm of estimation error has the rate $\|\hat{c}_t - c_t\|_{\max} = O_p\left((\Delta T_n \log d)^{\frac{1}{2}}\right)$.*

Proof. Based on the results of Lemma 1 (in the current paper) and Lemma 2 of Chen et al. (2020), we conclude that there exists some positive constants C_1 and C_2, such that for all $1 \leq r, s \leq q+d$, and any $x > 0$,

$$P\left(\left|\hat{c}_t^{(r,s)} - c_t^{(r,s)}\right| > x\right) \leq C_1 \exp\left(-\frac{C_2 x^2}{\Delta T_n}\right). \tag{G.1}$$

The detailed proof follows from the similar discussion in the proof of Lemma A.1 in Fan et al. (2016a). Because of the fact that

$$\left\{\|\hat{c}_t - c_t\|_{\max} > x\right\} = \bigcup_{r,s}\left\{\left|\hat{c}_t^{(r,s)} - c_t^{(r,s)}\right| > x\right\},$$

then based on the Bonferroni inequality, we obtain

$$\|\hat{c}_t - c_t\|_{\max} = O_p\left(\left[\Delta T_n \log\left(q+d\right)\right]^{\frac{1}{2}}\right)$$

using the similar technique as in Lemma A.2 (iv) of Fan et al. (2016a). Based on the assumption $q = o\left(d\right)$, which implies that $\log\left(q+d\right) = \log d + o\left(1\right)$, we finally obtain the convergence rate as stated in the lemma. □

Lemma 4. *Assume Conditions 4-5. Then for any* $0 \leq t \leq \mathcal{T}$, $\lambda_{\min}\left(\hat{c}_t^{\mathbf{X},\mathbf{X}}\right) > \frac{\vartheta_1}{2}$ *with probability approaching* 1.

Proof. First of all, by Weyl's theorem, we obtain:

$$\begin{aligned}\left|\lambda_{\min}\left(\hat{c}_t^{\mathbf{X},\mathbf{X}}\right) - \lambda_{\min}\left(c_t^{\mathbf{X},\mathbf{X}}\right)\right| &\leq \left\|\hat{c}_t^{\mathbf{X},\mathbf{X}} - c_t^{\mathbf{X},\mathbf{X}}\right\| \\ &\leq q\left\|\hat{c}_t^{\mathbf{X},\mathbf{X}} - c_t^{\mathbf{X},\mathbf{X}}\right\|_{\max},\end{aligned}$$

where the last inequality follows from the fact that $\|\mathbf{A}\| \leq \sqrt{pq}\,\|\mathbf{A}\|_{\max}$ for any $p \times q$ matrix $\mathbf{A}$.

Therefore, based on Lemma 3, we know that $\left\|\hat{c}_t^{\mathbf{X},\mathbf{X}} - c_t^{\mathbf{X},\mathbf{X}}\right\|_{\max} = O_p\left(\left(\Delta T_n \log d\right)^{\frac{1}{2}}\right)$ and

$$\left|\lambda_{\min}\left(\hat{c}_t^{\mathbf{X},\mathbf{X}}\right) - \lambda_{\min}\left(c_t^{\mathbf{X},\mathbf{X}}\right)\right| = O_p\left(q\left(\Delta T_n \log d\right)^{\frac{1}{2}}\right) = o_p\left(1\right).$$

By Condition 4, it is then easy to verify the result of this lemma. □

Based on the above lemmas, we now show the convergence rate of $\hat{c}_t^{\mathbf{B}\bullet\mathbf{X}}$ under elementwise max norm.

Proof of the rest of Theorem 3. We first define several notations: $c_t^{\mathbf{X},Y^{(l)}} = \left\{c_t^{(r,q+l)}\right\}_{1\leq r\leq q}$, a $q \times 1$ vector process, which is the $l-$th column of matrix $c_t^{\mathbf{X},\mathbf{Y}}$ for $1 \leq l \leq d$. That is, $c_t^{\mathbf{X},\mathbf{Y}} = \left(c_t^{\mathbf{X},Y^{(1)}}, c_t^{\mathbf{X},Y^{(2)}}, \ldots, c_t^{\mathbf{X},Y^{(d)}}\right)$. Similarly, we could define $\hat{c}_t^{\mathbf{X},\mathbf{Y}} = \left(\hat{c}_t^{\mathbf{X},Y^{(1)}}, \hat{c}_t^{\mathbf{X},Y^{(2)}}, \ldots, \hat{c}_t^{\mathbf{X},Y^{(d)}}\right)$. Therefore, it is easy to see that the $(r,s)-$th element of matrix $c_t^{\mathbf{B}\bullet\mathbf{X}}$ can be expressed as: $\left(c_t^{\mathbf{X},Y^{(r)}}\right)^{\intercal}\left(c_t^{\mathbf{X},\mathbf{X}}\right)^{-1} c_t^{\mathbf{X},Y^{(s)}}$, while that of matrix $\hat{c}_t^{\mathbf{B}\bullet\mathbf{X}}$ can be expressed as: $\left(\hat{c}_t^{\mathbf{X},Y^{(r)}}\right)^{\intercal}\left(\hat{c}_t^{\mathbf{X},\mathbf{X}}\right)^{-1}\hat{c}_t^{\mathbf{X},Y^{(s)}}$. Consequently, we obtain the expression of the $(r,s)-$th element of the error matrix as:

$$\begin{aligned}\left(\hat{c}_t^{\mathbf{B}\bullet\mathbf{X}} - c_t^{\mathbf{B}\bullet\mathbf{X}}\right)^{(r,s)} &= \left(\hat{c}_t^{\mathbf{X},Y^{(r)}} - c_t^{\mathbf{X},Y^{(r)}}\right)^{\intercal}\left(c_t^{\mathbf{X},\mathbf{X}}\right)^{-1}\hat{c}_t^{\mathbf{X},Y^{(s)}} + \left(c_t^{\mathbf{X},Y^{(r)}}\right)^{\intercal}\left(c_t^{\mathbf{X},\mathbf{X}}\right)^{-1}\left(\hat{c}_t^{\mathbf{X},Y^{(s)}} - c_t^{\mathbf{X},Y^{(s)}}\right) \\ &\quad + \left(\hat{c}_t^{\mathbf{X},Y^{(r)}}\right)^{\intercal}\left(\hat{c}_t^{\mathbf{X},\mathbf{X}}\right)^{-1}\left(c_t^{\mathbf{X},\mathbf{X}} - \hat{c}_t^{\mathbf{X},\mathbf{X}}\right)\left(c_t^{\mathbf{X},\mathbf{X}}\right)^{-1}\hat{c}_t^{\mathbf{X},Y^{(s)}}.\end{aligned}$$

Therefore, we have:

$$\max_{1\le r,s\le d}\left|\left(\hat{c}_t^{\mathbf{B}\bullet\mathbf{X}}-c_t^{\mathbf{B}\bullet\mathbf{X}}\right)^{(r,s)}\right| \le \max_{1\le r\le d}\left\|\hat{c}_t^{\mathbf{X},Y^{(r)}}-c_t^{\mathbf{X},Y^{(r)}}\right\|\left(\max_{1\le r\le d}\left\|\hat{c}_t^{\mathbf{X},Y^{(r)}}\right\|+\max_{1\le r\le d}\left\|c_t^{\mathbf{X},Y^{(r)}}\right\|\right)\left\|\left(c_t^{\mathbf{X},\mathbf{X}}\right)^{-1}\right\| \\ +\left(\max_{1\le r\le d}\left\|\hat{c}_t^{\mathbf{X},Y^{(r)}}\right\|\right)^2\left\|\left(\hat{c}_t^{\mathbf{X},\mathbf{X}}\right)^{-1}\right\|\left\|\left(c_t^{\mathbf{X},\mathbf{X}}\right)^{-1}\right\|\left\|c_t^{\mathbf{X},\mathbf{X}}-\hat{c}_t^{\mathbf{X},\mathbf{X}}\right\|. \tag{G.2}$$

Based on Condition 4, we have:

$$\left\|\left(c_t^{\mathbf{X},\mathbf{X}}\right)^{-1}\right\| < \vartheta_1^{-1}, \text{ and } \max_{1\le r\le d}\left\|c_t^{\mathbf{X},Y^{(r)}}\right\| < q^{1/2}\vartheta_2. \tag{G.3}$$

By the result of Lemma 4, it is easy to see that

$$\left\|\left(\hat{c}_t^{\mathbf{X},\mathbf{X}}\right)^{-1}\right\| < 2\vartheta_1^{-1} \tag{G.4}$$

with probability approaching 1. On the other hand, it is easy to see that

$$\max_{1\le r\le d}\left\|\hat{c}_t^{\mathbf{X},Y^{(r)}}-c_t^{\mathbf{X},Y^{(r)}}\right\| = O_p\left(q^{1/2}\left(\Delta T_n\log d\right)^{\frac{1}{2}}\right), \tag{G.5}$$

based on the result of Lemma 3 and the fact that $\left\|\hat{c}_t^{\mathbf{X},Y^{(r)}}-c_t^{\mathbf{X},Y^{(r)}}\right\| \le q^{1/2}\left\|\hat{c}_t^{\mathbf{X},Y^{(r)}}-c_t^{\mathbf{X},Y^{(r)}}\right\|_{\max}$. Consequently, we know that $\max_{1\le r\le d}\left\|\hat{c}_t^{\mathbf{X},Y^{(r)}}\right\| < 2q^{1/2}\vartheta_2$ with probability approaching 1.

Substituting (G.3)-(G.5) into (G.2), we finally obtain:

$$\max_{1\le r,s\le d}\left|\left(\hat{c}_t^{\mathbf{B}\bullet\mathbf{X}}-c_t^{\mathbf{B}\bullet\mathbf{X}}\right)^{(r,s)}\right| = O_p\left(\left(q^4\Delta T_n\log d\right)^{\frac{1}{2}}\right), \tag{G.6}$$

based on the result of Lemma 3 and the fact that $\left\|c_t^{\mathbf{X},\mathbf{X}}-\hat{c}_t^{\mathbf{X},\mathbf{X}}\right\| \le q\left\|c_t^{\mathbf{X},\mathbf{X}}-\hat{c}_t^{\mathbf{X},\mathbf{X}}\right\|_{\max}$. This shows (4.13). □

G. 1. *Proof of Theorem 4*

Define $\mathbf{V}_t = \hat{c}_t^{\mathbf{X},\mathbf{Y}} - c_t^{\mathbf{X},\mathbf{Y}}$ and

$$\begin{aligned}\hat{\mathbf{G}}_t &= \left[\hat{c}_t^{\mathbf{X},\mathbf{X}}+\hat{c}_t^{\mathbf{X},\mathbf{Y}}\left(\hat{\mathbf{s}}_t^*\right)^{-1}\left(\hat{c}_t^{\mathbf{X},\mathbf{Y}}\right)^{\intercal}\right]^{-1}, \\ \mathbf{G}_t &= \left[c_t^{\mathbf{X},\mathbf{X}}+c_t^{\mathbf{X},\mathbf{Y}}\mathbf{s}_t^{-1}\left(c_t^{\mathbf{X},\mathbf{Y}}\right)^{\intercal}\right]^{-1}.\end{aligned}$$

Lemma 5. *Assume the conditions of Theorem 4. Then*

(i) $\lambda_{\min}\left(c_t^{\mathbf{X},\mathbf{Y}}\mathbf{s}_t^{-1}\left(c_t^{\mathbf{X},\mathbf{Y}}\right)^{\intercal}\right) > \vartheta_5' d$ *for some* $\vartheta_5' > 0$.

(ii) $\|\mathbf{G}_t\| = O_p\left(d^{-1}\right)$.

(iii) $\|\mathbf{V}_t\|_F^2 = O_p\left(qd\Delta T_n\log d\right)$.

(iv) $\left\|\hat{c}_t^{\mathbf{X},\mathbf{Y}}\left(\hat{\mathbf{s}}_t^*\right)^{-1}\left(\hat{c}_t^{\mathbf{X},\mathbf{Y}}\right)^\intercal - c_t^{\mathbf{X},\mathbf{Y}}\mathbf{s}_t^{-1}\left(c_t^{\mathbf{X},\mathbf{Y}}\right)^\intercal\right\| = O_p\left(d\omega_n^{1-\nu}m_d\right).$

(v) $\left\|\hat{\mathbf{G}}_t\right\| = O_p\left(d^{-1}\right).$

(vi) $\left\|\left(\hat{c}_t^{\mathbf{X},\mathbf{Y}}\right)^\intercal\hat{\mathbf{G}}_t\hat{c}_t^{\mathbf{X},\mathbf{Y}}\left(\hat{\mathbf{s}}_t^*\right)^{-1}\right\| = O_p(1).$

Proof. (i) Note that

$$\lambda_{\min}\left(c_t^{\mathbf{X},\mathbf{Y}}\mathbf{s}_t^{-1}\left(c_t^{\mathbf{X},\mathbf{Y}}\right)^\intercal\right) \geq \lambda_{\min}\left(\mathbf{s}_t^{-1}\right)\lambda_{\min}\left(c_t^{\mathbf{X},\mathbf{Y}}\left(c_t^{\mathbf{X},\mathbf{Y}}\right)^\intercal\right).$$

Then by Conditions 6 and 7, we obtain that $\lambda_{\min}\left(c_t^{\mathbf{X},\mathbf{Y}}\mathbf{s}_t^{-1}\left(c_t^{\mathbf{X},\mathbf{Y}}\right)^\intercal\right) > \vartheta_5' d$ for some $\vartheta_5' > 0$ and all large d.

(ii) Because

$$\lambda_{\min}\left(c_t^{\mathbf{X},\mathbf{X}} + c_t^{\mathbf{X},\mathbf{Y}}\mathbf{s}_t^{-1}\left(c_t^{\mathbf{X},\mathbf{Y}}\right)^\intercal\right) \geq \lambda_{\min}\left(c_t^{\mathbf{X},\mathbf{Y}}\mathbf{s}_t^{-1}\left(c_t^{\mathbf{X},\mathbf{Y}}\right)^\intercal\right),$$

then the result (ii) follows immediately from result (i).

(iii) We have $\|\mathbf{V}_t\|_{\mathrm{F}}^2 \leq d\max_{1\leq r\leq d}\left\|\hat{c}_t^{\mathbf{X},Y^{(r)}} - c_t^{\mathbf{X},Y^{(r)}}\right\|^2 = O_p\left(qd\Delta T_n\log d\right)$ based on the result (G.5).

(iv) Note that

$$\begin{aligned} & \left\|\hat{c}_t^{\mathbf{X},\mathbf{Y}}\left(\hat{\mathbf{s}}_t^*\right)^{-1}\left(\hat{c}_t^{\mathbf{X},\mathbf{Y}}\right)^\intercal - c_t^{\mathbf{X},\mathbf{Y}}\mathbf{s}_t^{-1}\left(c_t^{\mathbf{X},\mathbf{Y}}\right)^\intercal\right\| \\ \leq\; & \left\|\mathbf{V}_t\left(\hat{\mathbf{s}}_t^*\right)^{-1}\mathbf{V}_t^\intercal\right\| + 2\left\|\mathbf{V}_t\left(\hat{\mathbf{s}}_t^*\right)^{-1}\left(c_t^{\mathbf{X},\mathbf{Y}}\right)^\intercal\right\| + \left\|c_t^{\mathbf{X},\mathbf{Y}}\left[\left(\hat{\mathbf{s}}_t^*\right)^{-1} - \mathbf{s}_t^{-1}\right]\left(c_t^{\mathbf{X},\mathbf{Y}}\right)^\intercal\right\|, \end{aligned}$$

where by result (iii), it is easy to verify that $\left\|\mathbf{V}_t\left(\hat{\mathbf{s}}_t^*\right)^{-1}\mathbf{V}_t^\intercal\right\| = O_p\left(qd\Delta T_n\log d\right) = o_p\left(d\omega_n\right)$, $\left\|\mathbf{V}_t\left(\hat{\mathbf{s}}_t^*\right)^{-1}\left(c_t^{\mathbf{X},\mathbf{Y}}\right)^\intercal\right\| = O_p\left(qd\left(\Delta T_n\log d\right)^{1/2}\right) = O_p\left(d\omega_n\right)$ and by Condition 7 (ii) and the result of Proposition 1, we obtain

$$\left\|c_t^{\mathbf{X},\mathbf{Y}}\left[\left(\hat{\mathbf{s}}_t^*\right)^{-1} - \mathbf{s}_t^{-1}\right]\left(c_t^{\mathbf{X},\mathbf{Y}}\right)^\intercal\right\| = O_p\left(d\omega_n^{1-\nu}m_d\right).$$

(v) First of all, $\left\|\hat{c}_t^{\mathbf{X},\mathbf{X}}\right\| \leq \left\|\hat{c}_t^{\mathbf{X},\mathbf{X}} - c_t^{\mathbf{X},\mathbf{X}}\right\| + \left\|c_t^{\mathbf{X},\mathbf{X}}\right\|$ where $\left\|\hat{c}_t^{\mathbf{X},\mathbf{X}} - c_t^{\mathbf{X},\mathbf{X}}\right\| \leq q\left\|\hat{c}_t - c_t\right\|_{\max} = O_p\left(\omega_n\right)$, and $\left\|c_t^{\mathbf{X},\mathbf{X}}\right\| = O_p(q) = o_p(d)$ by Conditions 4 and 5. Therefore, result (v) is immediately proved following from the results (i) and (iv) and the assumption $\omega_n^{1-\nu}m_d = o(1)$.

(vi) Because $\left\|\hat{c}_t^{\mathbf{X},\mathbf{Y}}\right\| \leq \|\mathbf{V}_t\| + \left\|c_t^{\mathbf{X},\mathbf{Y}}\right\|$ where $\|\mathbf{V}_t\| \leq \|\mathbf{V}_t\|_{\mathrm{F}} = O_p\left(d^{1/2}\omega_n\right)$ by result (iii), and $\left\|c_t^{\mathbf{X},\mathbf{Y}}\right\| = O_p\left(d^{1/2}\right)$ by Assumption 7 (ii). Thus we obtain $\left\|\hat{c}_t^{\mathbf{X},\mathbf{Y}}\right\|_{\mathrm{F}} = O_p\left(d^{1/2}\right)$. Finally, combing result (v) and Proposition 1, we obtain (vi). □

Proof of Theorem 4. By the Sherman-Morrison-Woodbury formula and the triangular inequality, we have

$\left\|\left(\hat{c}_t^{\mathbf{Y},\mathbf{Y},*}\right)^{-1}-\left(c_t^{\mathbf{Y},\mathbf{Y}}\right)^{-1}\right\| \leq \sum_{i=1}^{6} L_i$ where

$$
\begin{aligned}
L_1 &= \left\|\left(\hat{\mathbf{s}}_t^{*}\right)^{-1}-\mathbf{s}_t^{-1}\right\|, \\
L_2 &= \left\|\mathbf{s}_t^{-1}\left(c_t^{\mathbf{X},\mathbf{Y}}\right)^{\intercal}\mathbf{G}_t c_t^{\mathbf{X},\mathbf{Y}}\left[\left(\hat{\mathbf{s}}_t^{*}\right)^{-1}-\mathbf{s}_t^{-1}\right]\right\|, \\
L_3 &= \left\|\mathbf{s}_t^{-1}\left(c_t^{\mathbf{X},\mathbf{Y}}\right)^{\intercal}\mathbf{G}_t\left(\hat{c}_t^{\mathbf{X},\mathbf{Y}}-c_t^{\mathbf{X},\mathbf{Y}}\right)\left(\hat{\mathbf{s}}_t^{*}\right)^{-1}\right\|, \\
L_4 &= \left\|\mathbf{s}_t^{-1}\left(c_t^{\mathbf{X},\mathbf{Y}}\right)^{\intercal}\left(\hat{\mathbf{G}}_t-\mathbf{G}_t\right)\hat{c}_t^{\mathbf{X},\mathbf{Y}}\left(\hat{\mathbf{s}}_t^{*}\right)^{-1}\right\|, \\
L_5 &= \left\|\mathbf{s}_t^{-1}\left(\hat{c}_t^{\mathbf{X},\mathbf{Y}}-c_t^{\mathbf{X},\mathbf{Y}}\right)^{\intercal}\hat{\mathbf{G}}_t\hat{c}_t^{\mathbf{X},\mathbf{Y}}\left(\hat{\mathbf{s}}_t^{*}\right)^{-1}\right\|, \\
L_6 &= \left\|\left[\left(\hat{\mathbf{s}}_t^{*}\right)^{-1}-\mathbf{s}_t^{-1}\right]\left(\hat{c}_t^{\mathbf{X},\mathbf{Y}}\right)^{\intercal}\hat{\mathbf{G}}_t\hat{c}_t^{\mathbf{X},\mathbf{Y}}\left(\hat{\mathbf{s}}_t^{*}\right)^{-1}\right\|.
\end{aligned}
$$

First of all, by Proposition 1, $\left\|\left(\hat{\mathbf{s}}_t^{*}\right)^{-1}-\mathbf{s}_t^{-1}\right\| = O_p\left(\omega_n^{1-\nu}m_d\right)$. Second, by Assumptions 6 and 4 and result (ii) of Lemma 5, we have $\left\|\mathbf{s}_t^{-1}\left(c_t^{\mathbf{X},\mathbf{Y}}\right)^{\intercal}\mathbf{G}_t c_t^{\mathbf{X},\mathbf{Y}}\right\| = O_p(1)$ and consequently, $L_2 = O_p\left(\omega_n^{1-\nu}m_d\right)$. Third, based on the result (ii) of Lemma 5 and the fact that $\left\|c_t^{\mathbf{X},\mathbf{Y}}\right\| = O_p\left(d^{1/2}\right)$ and $\left\|\hat{c}_t^{\mathbf{X},\mathbf{Y}}-c_t^{\mathbf{X},\mathbf{Y}}\right\| \leq \sqrt{qd}\left\|\hat{c}_t-c_t\right\|_{\max} = O_p\left(\left(qd\Delta T_n\log d\right)^{\frac{1}{2}}\right)$, we can conclude that $L_3 = O_p\left(\omega_n\right)$. Similarly, based on the result (v) of Lemma 5, we obtain $L_5 = O_p\left(\omega_n\right)$. Fourth, note that

$$
\left\|\hat{\mathbf{G}}_t-\mathbf{G}_t\right\| = \left\|\hat{\mathbf{G}}_t\left[\mathbf{G}_t^{-1}-\hat{\mathbf{G}}_t^{-1}\right]\mathbf{G}_t\right\| \leq O_p\left(d^{-2}\right)\left\|\mathbf{G}_t^{-1}-\hat{\mathbf{G}}_t^{-1}\right\|,
$$

where $\left\|\mathbf{G}_t^{-1}-\hat{\mathbf{G}}_t^{-1}\right\| = O_p\left(d\omega_n^{1-\nu}m_d\right)+O_p\left(\omega_n\right)$ which follows from the result (iv) and the proof of result (v) in Lemma 5. Therefore, we obtain $\left\|\hat{\mathbf{G}}_t-\mathbf{G}_t\right\| = O_p\left(d^{-1}\omega_n^{1-\nu}m_d\right)$ which yields $L_4 = O_p\left(\omega_n^{1-\nu}m_d\right)$. Finally, by the result (vi) of Lemma 5, it is straightforward to see that $L_6 = O_p\left(\omega_n^{1-\nu}m_d\right)$. Thus, the theorem is proved. □