Statistica Sinica **35** (2025), 1-21
doi:https://doi.org/10.5705/ss.202022.0259

ON THE ESTIMATION OF HIGH-DIMENSIONAL INTEGRATED COVARIANCE MATRIX BASED ON HIGH-FREQUENCY DATA WITH MULTIPLE TRANSACTIONS

Moming Wang[1], Jianhua Hu[1], Ningning Xia*[1] and Yong Zhou[2]

[1]*Shanghai University of Finance and Economics*
and [2]*East China Normal University*

Abstract: In addition to microstructure noise, the presence of multiple transactions at each recording time is another common feature for high-frequency data. In this paper, we consider the estimation of integrated covariance (ICV) matrix for multiple high-frequency data in a high-dimensional situation where the number of stocks and the "effective" sample size go to infinity proportionally. First, we study the limiting spectral behavior of a pre-averaged version of averaged time-variation adjusted realized covariance (PA-ATVA) matrix based on multiple noisy observations. We show that the PA-ATVA matrix has several desirable properties: it eliminates the effects of microstructure noise and multiple transactions; it allows for rather general dependence structure in the noise process, both cross-sectional and temporal; its LSD depends solely on that of ICV matrix through the Marčenko–Pastur equation. Furthermore, we show that all the aforementioned properties still hold in the presence of asynchronicity. Second, we further propose a nonlinear shrinkage estimator of the ICV matrix based on the PA-ATVA matrix. We show that the proposed estimator is not only asymptotically positive-definite, but also enjoys a desirable estimation efficiency. At last, simulation and empirical studies demonstrate impressive performance of our proposed estimator.

Key words and phrases: High-dimension, high-frequency, microstructure noise, multiple transactions, random matrix theory.

1. Introduction

1.1. Motivation

The covariance structure of stock market is of great interest to investors and researchers, as it has a critical role in financial problems such as pricing and investment. Suppose that we have p stocks whose log prices at time t are denoted by $\mathbf{X}_t = (X_t^{(1)}, \ldots, X_t^{(p)})^{\mathrm{T}}$, where T denotes the transpose. The following diffusion processes are commonly used to model financial asset price processes:

$$d\mathbf{X}_t = \boldsymbol{\mu}_t dt + \boldsymbol{\Theta}_t d\mathbf{W}_t, \qquad t \in [0,1], \tag{1.1}$$

*Corresponding author. E-mail: xia.ningning@mail.shufe.edu.cn