MONASH University

MONASH BUSINESS SCHOOL

DEPARTMENT OF ECONOMETRICS AND BUSINESS STATISTICS

ISSN 1440-771X

WORKING PAPER SERIES

A Simple Bootstrap Method for Panel Data Inferences

Jiti Gao
Bin Peng
Yayi Yan

Working Paper No.7/22
June 2022

A Simple Bootstrap Method for Panel Data Inferences

Jiti Gao, Bin Peng and Yayi Yan

Monash University, Australia

June 10, 2022

Abstract

In this paper, we propose a simple dependent wild bootstrap procedure for us to establish valid inferences for a wide class of panel data models including those with interactive fixed effects. The proposed method allows for the error components having weak correlation over both dimensions, and heteroskedasticity. The asymptotic properties are established under a set of simple and general conditions, and bridge the literature on bootstrap methods and the literature of heteroskedasticity and autocorrlation consistent (HAC) approaches for panel data models. The new findings fill some gaps left by the bulk literature of the block bootstrap based panel data studies. Finally, we show the superiority of our approach over several natural competitors using extensive numerical studies.

Keywords: Cross-Sectional Dependence, Edgeworth Expansion, Panel Data Bootstrap, Time Series Autocorrelation

JEL classification: C12, C18, C23

1 Introduction

The past couple of decades have seen the rising popularity of panel data models. Especially for the applied researchers, panel data models are capable of providing accurate estimates by utilizing the information along both of the cross-sectional and time dimensions. Early works on panel data studies focusing on large N but small T setting (e.g., Arellano and Bond, 1991; Blundell and Bond, 1998), where N and T represent the numbers of observations available over the cross-sectional and time dimensions respectively. To obtain valid inference, a crucial assumption widely adopted in this strand of the literature is independence among the individual units (e.g., Arellano and Bond, 1991, p. 283; Blundell and Bond, 1998, p. 118; Hsiao, 2003, p. 35; among others). Although the HAC covariance matrix estimation method (Newey and West, 1987) can potentially be applied to allow for certain cross-sectional dependence (CSD), the HAC method does not gain its recognition along this line of research. The reason might be due to the fact the HAC method requires a natural ordering of the observations, which however does not apply to the cross-sectional units in general.

As we are embracing the era of data rich environment, the literature of panel data modelling starts shifting its focus to large N and T cases. Existing examples can be seen in Hsiao (2003) and Petersen (2009) for the necessity of diverging T. Together with this shift, CSD and time series autocorrelation (TSA) become important features of panel data modelling.

At this point, it is worth commenting on two types (i.e., referred to as strong and weak) of CSD, so we can be specific on our goal later and avoid any possible misunderstanding. The strong CSD usually infers a correlation, which does not change much with respect to the distance between two individual units. Among all possible definitions, the most well known realization is through a factor structure, e.g., Eqn. (2)-(3) of Pesaran (2006). On the other hand, the weak CSD is normally modelled in a way similar to TSA, that is a correlation defined either implicitly or explicitly using the "distance" between individual units, e.g., Assumption C of Bai (2009) and Assumption A2 of Chen et al. (2012). Over the past fifteen years or so, the literature seems to agree that the strong CSD can be handled well by both of the common correlated effects approach of Pesaran (2006), and the principal component analysis approach of Bai (2009). Building on these two seminal works, a variety of extensions have been made.

However, the difficulties raised with the weak CSD have not been satisfactorily dealt in the relevant literature. In particular, these challenges are mainly about valid inferences, and become even more daunting in the simultaneous presence of weak CSD, TSA and heteroskedasticity. To formalize our arguments, let us consider a dataset having the following simple structure:

$$\{u_{it} \,|\, i \in [N],\ t \in [T]\}, \tag{1.1}$$

where u_{it}'s are scalers and observable, and $[L] = \{1, 2, \ldots, L\}$ for a positive integer L. Suppose further that

$$E[u_{it}] = 0 \quad \text{and} \quad E[u_{it}u_{js}] = \alpha_{ij,ts}, \tag{1.2}$$

where $\sum_{i,j=1}^{N} \sum_{t,s=1}^{T} |\alpha_{ij,ts}| = O(\mathbb{N})$ with $\mathbb{N} = NT$ for short. Note that (1.2) does not only allow for correlation over both dimensions of u_{it}, but also permits heteroskedasticity. On this matter, Figure 1 provides a graphical representation of the relationship:

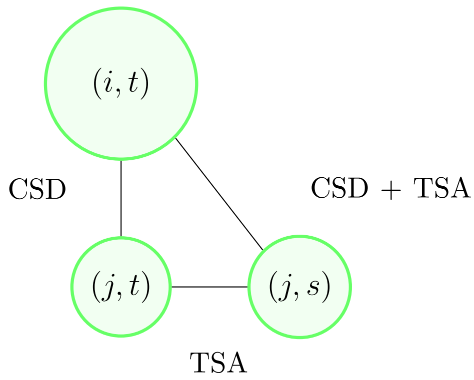

Figure 1: Heteroskedasticity and Dependence along the Two Dimensions.

In addition, one usually assumes (rather than prove) the following Central Limit Theorem (CLT) holds.

$$S_{\mathbb{N}} = \frac{1}{\sqrt{\mathbb{N}}} \sum_{i=1}^{N} \sum_{t=1}^{T} u_{it} := \frac{1}{\sqrt{\mathbb{N}}} \sum_{t=1}^{T} U_t^\top 1_N \to_D N(0, \sigma_u^2), \tag{1.3}$$

where $\sigma_u^2 = \lim_{N,T} \frac{1}{\mathbb{N}} \sum_{t,s=1}^{T} 1_N^\top E[U_t U_s^\top] 1_N > 0$, and the definitions of $U_t = (u_{1t}, \cdots, u_{Nt})^\top$ and $1_N = (1, \cdots, 1)^\top$ are standard. The conditions (1.2) and (1.3) are typical panel data assumptions (e.g., Chapter 2 of Hedeker and Gibbons, 2006; Chapter 3 of Fitzmaurice et al., 2011; Assumptions C and E of Bai, 2009). However, some fundamental questions have been left unanswered, e.g., (1). how to ensure (1.2) and (1.3) from a set of low level conditions in view of the presence of correlation over both dimensions of u_{it}; (2). does the convergence of (1.3) achieve the usual Berry-Esseen bound; and (3) is the quantity σ_u^2 estimable ?

To the best of our knowledge, the aforementioned questions have not been well addressed in the literature of panel data modelling. Among them, the estimation of σ_u^2 is especially important for the practical applications. Similar concerns have also been raised in the field of financial studies more than a decade ago. For example, Petersen (2009) writes "*Although the literature has used an assortment of methods to estimate standard errors in panel data sets, the chosen method is often incorrect and the literature provides little guidance to researchers as to which method should be used. In addition, some of the advice in the literature is simply wrong.*" Over the past couple of decades, although a variety of panel data models have been investigated, not much work has been done to improve the estimation of a quantity like σ_u^2. More often than not, one has to assume independence along at least one dimension of the dataset (e.g., Assumption 2 of Pesaran, 2006, Proposition 2 of Bai, 2009, Assumption A4 of Chen et al., 2012, Assumption LL of Moon and Weidner, 2015, Assumption A.4 of Su et al., 2015, Assumption 2.1 of Menzel (2021), among others). It is worth mentioning that Menzel (2021) uses a bootstrap technique to investigate the properties of panel data allowing cluster–dependence, which actually does not kick in through the idiosyncratic error components, and can be classified as the strong CSD mentioned above.

The studies sharing a similar concern with ours are Gonçalves (2011) and Bai et al. (2020). Specifically, Gonçalves (2011) studies a fixed effect panel data model, and proposes to use the moving blocks bootstrap (MBB) technique, which allows for the error terms to have weak correlation over both dimensions. However, how to select the optimal block size is left unanswered, and to the extent in which weak CSD can be allowed is stated vaguely using some high level conditions (i.e., Assumptions 3 and 4 of their paper). Bai et al. (2020) consider a setting similar to Gonçalves (2011), and propose using a combination of the HAC approach and the thresholding technique, which involves two tuning parameters — one is the bandwidth of the HAC approach, and the other one is the threshold for penalizing entries of the covariance matrix. Such a procedure can be computationally expensive, and optimal choices of the two tuning parameters are even more challenging.

In this article, we develop a simple dependent wild bootstrap (DWB) procedure to establish inference for a wide class of panel data models including those with interactive fixed effects. The DWB method was initially proposed by Shao (2010) to mimic the autocorrelation of time series. From here and onwards, we refer to our approach on panel data dependent wild bootstrap (PDWB) for simplicity; this method is easy to implement, and it requires only one tuning parameter. Accordingly, we derive the necessary asymptotic properties, and conduct extensive numerical studies to examine the finite sample properties and the proposed bootstrap method. It is worth pointing out that our study bridges the literature of bootstrap methods and the literature of HAC approaches, and provides a practical procedure on how to select the optimal tuning parameter. In addition, compared with the MBB, the PDWB can better handle the missing values. The block bootstrap shuffles the blocks randomly, so it may destroy the data structure to certain degree. By contrast, the PDWB method preserves the original information of the dataset in a natural manner. These new findings fill some gaps left by the bulk literature of the block bootstrap based panel data studies (e.g., Gonçalves, 2011; Palm et al., 2011).

The structure of the rest paper is as follows. Section 2 provides the basic setup, and presents the asymptotic results accordingly. In Section 3, we conduct extensive simulation studies to examine the finite sample properties of the proposed bootstrap method. Section 4 applies the proposed PDWB approach to a real dataset, inferring the aggregated mutual fund performance. Section 5 concludes. The theoretical development of the main result is presented in Appendix A of the paper, while we give the secondary results in the online supplementary Appendix B.

Before proceeding further, we introduce some mathematical symbols which will be repeatedly used in the article. $|\cdot|$ denotes the absolute value of a scalar or the spectral norm of a matrix; $\|\cdot\|$ denotes the Euclidean norm of a vector or the Frobenius norm of a matrix; $\|v\|_q \equiv (E|v|^q)^{1/q}$ for $q \geq 1$; $=_D$ denotes equality in distribution; $E^*[\cdot]$ and $\Pr^*(\cdot)$ stand for the expectation and probability operations induced by the bootstrap procedure; $\to_P$ and $\to_D$ stand for convergence in probability and convergence in distribution respectively; $\lfloor q \rfloor$ stands for the largest integer not larger than q; for two numbers a and b, $a \asymp b$ stands for $a = O(b)$ and $b = O(a)$; let $\Psi(\cdot)$ and $\psi(\cdot)$ be the cumulative distribution function (CDF) and the probability density function (PDF) of the standard normal distribution respectively; $M_A = I - A(A^\top A)^{-1}A^\top$ denotes the projection matrix for any matrix A with full column rank; 1_N and 0_N stand for a $N \times 1$ vector of ones and a $N \times 1$ vector of zeros, respectively.

2 The Setup and Asymptotic Results

In what follows, we present the basic framework, and introduce some preliminary results without specifying any model in Section 2.1, and propose the PDWB method in Section 2.2. In Section 2.3, we demonstrate that the newly proposed framework can be applied to a wide class of commonly used models. In Section 2.4, we discuss how to handle the unbalanced panel dataset using the PDWB approach in practice.

2.1 The Setup

We now focus on the dataset of (1.1). Instead of assuming (1.2) and (1.3), we provide a set of restrictions which ensures the validity of both assumptions. The following conditions also cover a wide range of data generating processes (DGPs) of the literature.

Assumption 1. *Let* $\overline{U}_t \equiv \frac{1}{\sqrt{N}} U_t^\top 1_N$*, in which* $U_t = (u_{1t}, \ldots, u_{NT})^\top$ *follows a data generating process* $U_t \equiv g(\varepsilon_t, \varepsilon_{t-1}, \ldots)$ *with* $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})^\top$ *being a sequence of independent and identically distributed (i.i.d.) random vectors,* $E[U_t] = 0_N$*, and* $g(\cdot)$ *is a measurable function. In addition, let* $\overline{U}_t^* \equiv \frac{1}{\sqrt{N}} U_t^{*\top} 1_N$*, where* $U_t^* = g(\varepsilon_t, \ldots, \varepsilon_1, \varepsilon_0', \varepsilon_{-1}', \ldots)$ *is the coupled version of* U_t*, and* $\{\varepsilon_t'\}$ *is an independent copy of* $\{\varepsilon_t\}$*. Suppose that* $\sum_{t=0}^{\infty} t^2 \lambda_{t,\delta}^U < \infty$ *for* $\delta \geq 4$*, where* $\lambda_{t,\delta}^U = \|\overline{U}_t - \overline{U}_t^*\|_\delta$*.*

Assumption 1 extends the nonlinear system of Wu (2005) to a panel data setting. We now provide some details to show that Assumption 1 covers (1.2), and covers a wide range of data generating processes.

Example 2.1. *Consider a high-dimensional MA(*∞*) process* $U_t = \sum_{j=0}^{\infty} B_j \varepsilon_{t-j}$*, where* B_j*'s are* $N \times N$ *matrices satisfying* $||B_j|| = O(\rho^j)$ *for some* $|\rho| < 1$*, and* $\{\varepsilon_{it}\}$ *is independent over* i*. Then the high-dimensional MA(*∞*) process fulfils Assumption 1.*

Assuming B_j's are diagonal matrices, Example 2.1 immediately covers Assumption 2 of Pesaran (2006):

$$u_{it} = \sum_{j=0}^{\infty} b_{ij} \varepsilon_{i,t-j}, \tag{2.1}$$

in which TSA presents and heteroskedasticity over i is allowed. A similar argument applies to Assumption 8 of Gonçalves and Perron (2014). Assuming $B_j \equiv 0$ for $j \geq 1$, it then covers Assumption A4 of Chen et al. (2012) and Assumption E of Shi and Lee (2017), where only the weak CSD matters. Providing more structure on the off-diagonal elements of B_j's, Example 2.1 satisfies Assumption C of Bai and Ng (2002) and Assumption C of Bai (2009), where both of the weak CSD and TSA exist. Therefore, Assumption 1 is general enough to cover a wide range of DGPs in the relevant literature. More examples can be seen in Onatski (2010), Gonçalves (2011), Bai et al. (2020), among others.

Note that Assumption 1 regulates the weak CSD and TSA respectively, and provides an underlying data generating process to satisfy the moment restriction:

$$E|\overline{U}_t|^4 = E\left|\frac{1}{\sqrt{N}} U_t^\top 1_N\right|^4 < \infty, \tag{2.2}$$

which has been widely adopted in the literature of panel data analysis (see, for example, Bai and Ng (2002, Assumption C)). Formally, we present the following proposition.

Proposition 2.1. *Under Assumption 1, we have*

CSD: $\|\overline{U}_t\|_{\delta^*} \leq 2 \sum_{t=0}^{\infty} \lambda_{t,\delta}^U < \infty$ *with* $1 \leq \delta^* \leq \delta$*,*

TSA: $\sum_{t=1}^{\infty} t^2 |E[\overline{U}_t \overline{U}_0]| \leq \|\overline{U}_0\|_2 \sum_{t=0}^{\infty} t^2 \lambda_{t,2}^U < \infty$.

In the first result of Proposition 2.1, by taking $\delta^* = 2$ and $\delta^* = 4$ respectively, we bound the weak CSD, and the fourth moment presented in (2.2). In the same spirit, the second result of Proposition 2.1 imposes s restriction on the TSA.

Having quantified the weak CSD and TSA properly, we are ready to show that Assumption 1 ensures the validity of (1.3) in the following proposition.

Proposition 2.2. *Under Assumption 1, as* $(N, T) \to (\infty, \infty)$*,*

1. $S_{\mathbb{N}} \to_D N(0, \sigma_u^2)$,

2. $\sup_{w\in\mathbb{R}} |\Pr(S_{\mathbb{N}} \leq w) - \Psi_{\mathbb{N}}(w)| = O(T^{-1}(\log T)^5)$,

where $\Psi_{\mathbb{N}}(w) = \Psi\left(\frac{w}{s_{\mathbb{N}}}\right) + \frac{1}{6}\kappa_{\mathbb{N}}^3 \left(1 - \frac{w^2}{s_{\mathbb{N}}^2}\right) \psi\left(\frac{w}{s_{\mathbb{N}}}\right)$*, in which* $s_{\mathbb{N}}^2 = E[S_{\mathbb{N}}^2]$ *and* $\kappa_{\mathbb{N}}^3 = E[S_{\mathbb{N}}^3]$.

The first result of Proposition 2.2 infers that the CLT holds under Assumption 1, while the second result presents an Edgeworth expansion up to the second order. By Lemma B.2.3 of the online Appendix B, we can further simplify the second result to obtain the Berry-Esseen bound:

$$\sup_{w\in\mathbb{R}} \left|\Pr(S_{\mathbb{N}} \leq w) - \Psi\left(\frac{w}{s_{\mathbb{N}}}\right)\right| = O\left(\frac{1}{\sqrt{T}}\right). \tag{2.3}$$

By looking at (2.3), when recovering the distribution of $\Pr(S_{\mathbb{N}} \leq w)$, one needs only to focus on the population quantity $E[S_{\mathbb{N}}^2]$, which is the motivation shared by both of the bootstrap and HAC methods. Therefore, the presentation of (2.3) in a sense bridges the two streams of the literature, and sheds light on how to select the tuning parameter involved in the PDWB approach for us to investigate this in the last part of Section 2.2 below.

Up to this point, we have fully demonstrated the usefulness of Assumption 1, and we are now ready to present the PDWB approach.

2.2 The PDWB Approach

In this subsection, our focus is to recover the asymptotic distribution $N(0, \sigma_u^2)$ using the PDWB approach. Specifically, for each bootstrap replication, we draw an ℓ-dependent time series ξ_t's satisfying that the following condition.

Assumption 2. *Let* $E[\xi_t] = 0$*,* $E|\xi_t|^2 = 1$*,* $E|\xi_t|^4 < \infty$*,* $E[\xi_t\xi_s] = a(\frac{t-s}{\ell})$*, where* $(\frac{1}{\ell}, \frac{\ell}{\sqrt{T}}) \to (0, 0)$*, and* $a(\cdot)$ *is a symmetric kernel function defined on* $[-1, 1]$ *satisfying that* $a(\cdot)$ *is Lipschitz continuous on* $[-1, 1]$*,* $a(0) = 1$*, and* $K_a(x) = \int_{-\infty}^{\infty} a(u)e^{-iux}\, du \geq 0$ *for* $x \in \mathbb{R}$.

The condition of $K_a(x)$ ensures the semi-positive definiteness of the covariance matrix of $\{\xi_t\}_{t=1}^T$, while the restrictions on $a(\cdot)$ are satisfied by a few commonly used kernels, such as the Bartlett and Parzen kernels. In practice, one may generate $\xi \equiv (\xi_1, \ldots, \xi_T)^\top \sim N(0, \Sigma_\xi)$, where $\Sigma_\xi = \{a\left(\frac{t-s}{\ell}\right)\}_{T\times T}$. The normal distribution is not really necessary, but it fulfils Assumption 2, and is easy to implement.

Accordingly, the bootstrap version of $S_{\mathbb{N}}$ is constructed as follows:

$$S_{\mathbb{N}}^* = \frac{1}{\sqrt{\mathbb{N}}}\sum_{t=1}^T U_t^\top 1_N \xi_t, \tag{2.4}$$

which, in connection with Assumptions 1 and 2, yields the following theorem of the paper.

Theorem 2.1. *Under Assumptions 1 and 2, as $(N,T) \to (\infty, \infty)$,*

$$\sup_{w\in\mathbb{R}} |\Pr^*(S_{\mathbb{N}}^* \le w) - \Pr(S_{\mathbb{N}} \le w)| = o_P(1).$$

Theorem 2.1 indicates that the bootstrap procedure can fully recover the asymptotic distribution $N(0, \sigma_u^2)$. As a consequence, one can easily use the bootstrap draws to establish the confidence interval while allowing for both of the weak CSD and TSA. In addition, the bootstrap draws also offer a sample version of the form:

$$E^*[S_{\mathbb{N}}^{*2}] = \frac{1}{T}\sum_{t,s=1}^T \overline{U}_t \overline{U}_s a\left(\frac{t-s}{\ell}\right), \tag{2.5}$$

to consistently estimate $E[S_{\mathbb{N}}^2]$. Therefore, to select the optimal ℓ, we minimise the mean squared error (MSE) between $E^*[S_{\mathbb{N}}^{*2}]$ and $E[S_{\mathbb{N}}^2]$, which in fact a widely adopted criterion in the literature of HAC methods (e.g., Andrews, 1991).

Remark 2.1. *The result of Theorem 2.1 and the aforementioned discussion still hold if the weak CSD and/or TSA vanish. We further verify this argument through the simulation studies of Section 3.*

Optimal ℓ — We are now ready to consider the choice of ℓ. Before proceeding further, we impose one more condition on the kernel function $a(\cdot)$.

Assumption 3. *For $q \in [2]$, suppose that $\lim_{|x|\to 0} \frac{1-a(x)}{|x|^q} = c_q$ for some real number $0 < c_q < \infty$.*

Assumption 3 is standard in the literature. For example, for the Bartlett kernel, $q = 1$ and $c_1 = 1$; for the Parzen, Tukey-Hanning, QS kernels, and the trapezoidal functions, $q = 2$ and the values of c_2 vary but all satisfy the condition $c_2 < \infty$. We refer interested readers to

Kiefer and Vogelsang (2002) and Paparoditis and Politis (2001) for the properties of Bartlett kernel and trapezoidal functions respectively, and to Andrews (1991) for discussions on the other kernel functions.

Theorem 2.2. *Under Assumptions 1-3, as* $(N, T) \to (\infty, \infty)$,

Bias: $E\left(E^*[S_{\mathbb{N}}^{*2}]\right) - E[S_{\mathbb{N}}^2] = -\frac{c_q}{\ell^q}\Delta_1 + o(\ell^{-q})$,

Variance: $\mathrm{Var}(E^*[S_{\mathbb{N}}^{*2}]) = \frac{2\ell}{T}\Delta_2 + o(\ell/T)$,

where $\Delta_1 = \sum_{k=-\infty}^{\infty} |k|^q E[\overline{U}_0 \overline{U}_k]$ *and* $\Delta_2 = (E[S_{\mathbb{N}}^2])^2 \int_{-1}^{1} a^2(x)\mathrm{d}x$.

Theorem 2.2 summaries the bias and variance associated with the estimator of $E[S_{\mathbb{N}}^2]$ yielded by the PDWB procedure. Simple algebra shows that the corresponding MSE is minimized when the bandwidth is

$$\ell_{\text{opt}} = \left(q c_q^2 \Delta_1^2 / \Delta_2\right)^{1/(2q+1)} T^{1/(2q+1)}. \tag{2.6}$$

Obviously,

$$\ell_{\text{opt}} \asymp \begin{cases} T^{1/3}, & \text{if } q = 1 \\ T^{1/5}, & \text{if } q = 2 \end{cases}. \tag{2.7}$$

Up to this point, we would like to point out that the bulk literature of the block bootstrap based panel data studies (e.g., Gonçalves, 2011; Palm et al., 2011) has not been able to connect the bootstrap procedure with the HAC approach. Also, there is no theoretical guide on how to select the optimal bandwidth. In Section 3.1, we provide more details about the numerical implementation of (2.6) and (2.7).

Till now, no specific model has been investigated. In the following subsection, we apply the PDWB approach to some specific panel data models including those with interactive fixed effects, which have attracted considerable attention since the seminal papers of Pesaran (2006) and Bai (2009).

2.3 Applications of the PDWB

From now on, we treat u_{it}'s as unobservable idiosyncratic errors, and the model to be studied is

$$Y_t = F_t(\theta_0) + U_t, \tag{2.8}$$

where Y_t is a $N \times 1$ observable vector, θ_0 is a $d \times 1$ vector of interest with d being finite, and $F_t(\cdot)$ includes the observable regressors and may also include unknown parameters such as interactive fixed effects. For simplicity, we suppose $\{U_t\}$ is independent of $\{F_t(\theta_0)\}$ throughout this subsection, and will no longer mention this again.

Remark 2.2. *In this study, we focus on the parametric model. One may further generalize the specification of $F_t(\cdot)$ to incorporate nonparametric (e.g., $y_{it} = f(x_{it}) + u_{it}$) and semi-parametric models (e.g., $y_{it} = f(x_{it}'\theta_0) + u_{it}$). Although we believe the PDWB method can be applied to these flexible models, formalizing the restrictions and deriving the properties should be left for future study.*

We now provide two specific examples below.

Example 2.2. *Consider a simple panel data model:*

$$Y_t = X_t\theta_0 + U_t, \tag{2.9}$$

where X_t is a $N \times d$ observable matrix. In this case, we have

$$F_t(\theta_0) \equiv X_t\theta_0. \tag{2.10}$$

Example 2.3. *Consider a panel data model with interactive fixed effects:*

$$Y_t = X_t\theta_0 + \Gamma_0 f_t + U_t, \tag{2.11}$$

where Γ_0 is a $N \times p$ unknown matrix, f_t is a $p \times 1$ unknown vector, p is a known finite positive integer for simplicity, and X_t is still a $N \times d$ observable matrix. In this case,

$$F_t(\theta_0) = X_t\theta_0 + \Gamma_0 f_t. \tag{2.12}$$

Other examples include, for instance, Eq. (1.1) of Chen et al. (2012) and some classical ones in Hsiao (2003).

To carry on, we suppose that there is a consistent estimator of θ_0, say $\widehat{\theta}$, which admits the following asymptotic distribution:

$$\sqrt{\mathbb{N}}(\widehat{\theta} - \theta_0) \to_D N(\mu, \Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}) \tag{2.13}$$

with Σ_1 and Σ_2 being $d \times d$ deterministic invertible matrices, and μ being the unknown mean vector.

Remark 2.3. *For some simple models, such as Example 2.2, we may have $\mu = 0_d$, while μ is non-zero for cases such as Theorem 3 of Bai (2009) and its extensions. Here, it is worth emphasizing that our goal is not about the bias correction. In the literature, great efforts have been devoted to correct asymptotic biases for different models and estimation methods, such as Section 7 of Bai (2009), Dhaene and Jochmans (2015), Section 4.3 of Chen et al. (2021), among others.*

Instead, we aim to recover the covariance matrix of $N(\mu, \Sigma_1^{-1}\Sigma_2\Sigma_1^{-1})$ when certain weak correlation exists along both dimensions of u_{it}. To obtain valid inferences for panel data models in practice, it is impossible without valid information on $\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}$, the estimation of which as explained in the introduction is not an easy job, and has not been explored much. It is therefore part of the objective of this paper.

To recover $N(\mu, \Sigma_1^{-1}\Sigma_2\Sigma_1^{-1})$, we first note that the estimation of Σ_1 is usually straightforward, we thus suppose that there is an estimator $\widehat{\Sigma}_1$ such that

$$\|\widehat{\Sigma}_1 - \Sigma_1\| \to_P 0. \tag{2.14}$$

Also, in view of the literature mentioned in Remark 2.3, it is reasonable to assume that there is an estimator of $\widehat{\mu}$ (say, constructed using the sample analog) such that

$$\|\widehat{\mu} - \mu\| \to_P 0 \quad \text{if} \quad \mu \neq 0. \tag{2.15}$$

We assume further that the distribution of (2.13) (without μ) is due to

$$L_{\mathbb{N}} = \widehat{\Sigma}_1^{-1} \frac{1}{\sqrt{\mathbb{N}}} \sum_{t=1}^{T} Z_t^\top U_t \to_D N(0, \Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}), \tag{2.16}$$

where Z_t is a $N \times d$ matrix, and is generated from some intermediate step of the asymptotic analysis. For the above two examples, Z_t has the following presentations:

$$\begin{aligned} &\text{Example 2.2} — \quad Z_t \equiv X_t, \\ &\text{Example 2.3} — \quad Z_t \equiv M_{\Gamma_0} X_t - \frac{1}{T}\sum_{s=1}^{T} (f_t'(F'F/T)^{-1} f_s) M_{\Gamma_0} X_s, \end{aligned} \tag{2.17}$$

in which Z_t of Example 2.3 is identical to that on Bai (2009, p. 1245) but interchanging the i and t dimension, and $F = (f_1, \ldots, f_T)^\top$.

According to (2.16), we define its bootstrap version as follows:

$$L_{\mathbb{N}}^* = \widehat{\Sigma}_1^{-1} \frac{1}{\sqrt{\mathbb{N}}} \sum_{t=1}^{T} Z_t^\top \widehat{U}_t \xi_t, \tag{2.18}$$

where $\widehat{U}_t$ is the estimated version of U_t, and usually depends on the model. For the time being, we take its existence as granted, and shall come back to its detailed definition in the discussion of Examples 2.1–2.3 below. Our goal is to show that

$$\sup_{w\in\mathbb{R}}\left|\Pr^*(L_{\mathbb{N}}^* + \widehat{\mu} \le w) - \Pr(\sqrt{\mathbb{N}}(\widehat{\theta} - \theta_0) \le w)\right| = o_P(1). \tag{2.19}$$

To facilitate the development, we introduce the following conditions.

Assumption 4. *Let* (2.14)-(2.16) *hold, and let* $\frac{1}{\sqrt{\mathbb{N}}}\sum_{t=1}^{T} Z_t^\top \widehat{U}_t \xi_t = \frac{1}{\sqrt{\mathbb{N}}}\sum_{t=1}^{T} Z_t^\top U_t \xi_t + o_P(1)$.

We have justified (2.14)-(2.16) above. Moreover, when it comes to a specific model, instead of assuming (2.16), it can easily be proved using Assumption 1 and the independence between $\{Z_t\}$ and $\{U_t\}$. Thus, for a detailed model, the necessary high–level conditions in Assumption 4 can be easily verified. The extra condition of Assumption 4 is also a high level one. In Corollaries 2.1 and 2.2 below, we show it can be easily fulfilled for Examples 2.2 and 2.3, respectively.

Under Assumption 4, the next theorem follows.

Theorem 2.3. *Under Assumptions 1, 2 and 4, as* $(N, T) \to (\infty, \infty)$*, equation* (2.19) *holds.*

Theorem 2.3 infers that the bootstrap procedure can fully recover the distribution of (2.13). We now come back to Examples 2.2 and 2.3 to explain the usefulness and applicability of Theorem 2.3.

Example 2.2 (Cont.) — The ordinary least squares (OLS) estimator is defined by

$$\widehat{\theta} = \left(\sum_{t=1}^{T} X_t^\top X_t\right)^{-1} \sum_{t=1}^{T} X_t^\top Y_t \tag{2.20}$$

which immediately yields that $\widehat{U}_t = Y_t - X_t\widehat{\theta}$. Also, it is obvious that $\widehat{\Sigma}_1 = \frac{1}{\mathbb{N}}\sum_{t=1}^{T} X_t^\top X_t$. Using these notations, the bootstrap procedure can be specified as follows.

1. For each bootstrap replication, let $Y_t^* = X_t^\top \widehat{\theta} + \widehat{U}_t \xi_t$ for each t.
2. Calculate $\widetilde{\theta}$ as in (2.20) using the bootstrap samples $\{Y_t^*, X_t\}_{t=1}^{T}$.
3. Repeat the first two steps R times.

By design,

$$\sqrt{\mathbb{N}}(\widetilde{\theta} - \widehat{\theta}) = \widehat{\Sigma}_1^{-1}\frac{1}{\sqrt{\mathbb{N}}}\sum_{t=1}^{T} X_t^\top U_t \xi_t + \widehat{\Sigma}_1^{-1}\left(\frac{1}{\mathbb{N}}\sum_{t=1}^{T} X_t^\top X_t \xi_t\right)\sqrt{\mathbb{N}}(\theta_0 - \widehat{\theta})$$

$$= \widehat{\Sigma}_1^{-1} \frac{1}{\sqrt{\mathbb{N}}} \sum_{t=1}^{T} X_t^\top U_t \xi_t + o_P(1).$$

Thus, the next corollary immediately holds.

Corollary 2.1. *Consider Example 2.2 and suppose that* (2.14) *holds and* $\max_{i,t} E|X_{it}|^4 < \infty$. *Let further Assumptions 1 and 2 hold. As* $(N, T) \to (\infty, \infty)$,

$$\sup_{w \in \mathbb{R}} \left| \Pr^*(\sqrt{\mathbb{N}}(\widetilde{\theta} - \widehat{\theta}) \le w) - \Pr(\sqrt{\mathbb{N}}(\widehat{\theta} - \theta_0) \le w) \right| = o_P(1).$$

Example 2.3 (Cont.) — For this example, we consider the following objective function

$$Q_{\mathbb{N}}(\theta, \Gamma) = \sum_{t=1}^{T} (Y_t - X_t\theta)^\top M_\Gamma (Y_t - X_t\theta), \tag{2.21}$$

where Γ is a generic $N \times p$ matrix and satisfies that $\frac{1}{N}\Gamma^\top\Gamma = I_p$ for the purpose of identification. In (2.21), we have interchanged the i and t dimensions compared to the estimation conducted in Bai (2009). Accordingly, we estimate θ_0 and Γ_0 by minimizing (2.21):

$$(\widehat{\theta}, \widehat{\Gamma}) = \underset{\theta, \Gamma}{\operatorname{argmin}}\, Q_{\mathbb{N}}(\theta, \Gamma). \tag{2.22}$$

Also, we obtain $\widehat{f}_t = \frac{1}{N}\widehat{\Gamma}^\top(Y_t - X_t\widehat{\theta})$ and $\widehat{U}_t = Y_t - X_t\widehat{\theta} - \widehat{\Gamma}\widehat{f}_t$.

To establish the valid inference for θ_0, the bootstrap procedure is as follows.

1. For each bootstrap replication, let $Y_t^* = X_t^\top\widehat{\theta} + \widehat{\Gamma}\widehat{f}_t + \widehat{U}_t\xi_t$.

2. Treat $\widehat{\Gamma}\widehat{f}_t$ as observable, and calculate $\widetilde{\theta}$ as follows.

$$\widetilde{\theta} = \left(\sum_{t=1}^{T} X_t^\top M_{\widehat{\Gamma}} X_t \right)^{-1} \sum_{t=1}^{T} X_t^\top M_{\widehat{\Gamma}} Y_t^*.$$

3. Repeat the first two steps R times.

By design, after some algebra (presented in the proof of Corollary 2.2), we obtain

$$\widetilde{\theta} - \widehat{\theta} = \widehat{\Sigma}_1^{-1} \cdot \frac{1}{\mathbb{N}} \sum_{t=1}^{T} X_t^\top M_{\widehat{\Gamma}} (U_t + \Gamma_0 f_t)\xi_t + o_P(1). \tag{2.23}$$

In view of the proof of Theorem 3 of Bai (2009), one can decompose $M_{\widehat{\Gamma}}(U_t + \Gamma_0 f_t)\xi_t$, so the two bias terms will arise. However, both terms include $\{\xi_t\}$ that is independent of all the other variables and has mean 0, as a consequence these two terms will vanish asymptotically under Assumption 2.

That said, the following corollary holds for Example 2.3.

Corollary 2.2. *Consider Example 2.3. Suppose that the following conditions hold:*

1. $\max_{i,t} E\|x_{it}\|^4 < \infty$, *and* $\inf_{\Gamma \in \mathcal{G}} D(\Gamma) > 0$, *where* $\mathcal{G} = \{\Gamma \,|\, \frac{1}{N}\Gamma^\top \Gamma = I_d\}$*;*

2. $\max_t E\|f_t\|^4 < \infty$ *and* $\frac{1}{T}F^\top F \to_P \Sigma_F$ *for some* $p \times p$ *matrix* Σ_F*;* $\max_i E\|\gamma_i\|^4 < \infty$ *and* $\frac{1}{N}\Gamma_0^\top \Gamma_0 \to_P \Sigma_\Gamma$ *for some* $p \times p$ *matrix* Σ_Γ*, where* $\gamma_i^\top$ *is the* i^{th} *row of* Γ_0*;*

3. *Let Assumptions 1 and 2 hold. There exist* $\widehat{\mu}_B$ *and* $\widehat{\mu}_C$ *such that* $\|\widehat{\mu}_B - B\| \to_P 0$ *and* $\|\widehat{\mu}_C - C\| \to_P 0$*;*

where

$$
\begin{aligned}
D(\Gamma) &= \frac{1}{NT}\sum_{t=1}^T X_t^\top M_\Gamma X_t - \frac{1}{NT^2}\sum_{t=1}^T\sum_{s=1}^T X_t^\top M_\Gamma X_s a_{ts},\\
B &= -D(\Gamma_0)^{-1}\sum_{t=1}^T\sum_{s=1}^T (X_t - V_t)^\top \Gamma_0 \left(\Gamma_0^\top \Gamma_0\right)^{-1}\left(F^\top F\right)^{-1} f_s \frac{E[U_t^\top U_s]}{N},\\
C &= -D(\Gamma_0)^{-1}\sum_{t=1}^T X_t^\top M_{\Gamma_0}\Omega_u \Gamma_0 \left(\Gamma_0^\top \Gamma_0\right)^{-1}\left(F^\top F\right)^{-1} f_t,\\
a_{ts} &= f_t^\top (F^\top F/T)^{-1} f_s, \quad V_t = \frac{1}{T}\sum_{s=1}^T a_{ts}X_s, \quad \Omega_u = E[U_1 U_1^\top].
\end{aligned}
$$

As $(N,T) \to (\infty,\infty)$ *and* $T/N \to \rho > 0$,

$$
\sup_w \left|\Pr{}^*(\sqrt{\mathbb{N}}(\widetilde{\theta} - \widehat{\theta}) + \rho^{1/2}\widehat{\mu}_B + \rho^{-1/2}\widehat{\mu}_C \le w) - \Pr(\sqrt{\mathbb{N}}(\widehat{\theta} - \theta_0) \le w)\right| = o_P(1).
$$

The first and second conditions of Corollary 2.2 are Assumptions A and B of Bai (2009). As presented in Propositions 2.1 and 2.2, Assumption 1 in fact indicates Assumptions C and E of Bai (2009). Also, Remark 2.3 has explained it is reasonable to assume the existence of $\widehat{\mu}_B$ and $\widehat{\mu}_C$. Thus, the asymptotic distribution of $\sqrt{\mathbb{N}}(\widehat{\theta} - \theta_0)$ becomes straightforward in view of the development of Theorem 3 of Bai (2009). In the appendix, we focus on $\sqrt{\mathbb{N}}(\widetilde{\theta} - \widehat{\theta})$. In Section 3.4, we further examine this corollary with extensive simulation studies.

2.4 On Unbalanced Dataset

To close our investigation on the PDWB method, we consider the following unbalanced panel dataset

$$
\{u_{it} \,|\, i \in [N_t] \text{ for } \forall t,\ t \in [T]\}, \tag{2.24}
$$

where N_t varies with respect to t, and $\mathbb{N} = \sum_{t=1}^{T} N_t$. The structure of (2.24) is widely adopted in the literature of panel data modelling (e.g., Chapter 9 of Baltagi, 2005; Chapter 4 of Hedeker and Gibbons, 2006), and also suits the mutual fund dataset of Section 4.

To accommodate the missing values, we can rewrite $\overline{U}_t$ of Assumption 1 as

$$\overline{U}_t = \frac{1}{\sqrt{N_t}} U_t^\top \mathscr{L}_t, \tag{2.25}$$

where $\mathscr{L}_t$ is a $N \times 1$ vector with elements being 1 and 0 only to represent non-missing and missing respectively. By (2.24) and (2.25), $\|\mathscr{L}_t\| = \sqrt{N_t}$. Under some trivial modification, one can show that the established results still hold. For example, we may adopt the following condition.

Assumption 5. *Suppose that* $\frac{\overline{N}T}{\mathbb{N}} \to c \in (0, \infty)$*, where* $\overline{N} = \max_t N_t$ *and* c *is a constant.*

Assumption 5 allows $\underline{N} = \min_t N_t$ to be a fixed value, however, the number of N_t's being finite has to be negligible.

Corollary 2.3. *Under Assumption 1, 2 and 5, as* $(N, T) \to (\infty, \infty)$*,*

$$\sup_{w \in \mathbb{R}} |\Pr^*(S_{\mathbb{N}}^* \leq w) - \Pr(S_{\mathbb{N}} \leq w)| = o_P(1),$$

where $S_{\mathbb{N}}^*$ *and* $S_{\mathbb{N}}$ *are defined in an obvious manner using* (2.25)*.*

Remark 2.4. *Compared with the block bootstrap based studies, one more advantage of PDWB is that it can better handle the missing values. Note that the block bootstrap shuffles the blocks randomly, as a consequence the positions of missing values will be different for each bootstrap replication. In a sense, shuffling the blocks may destroy the data structure to certain degree. By contrast, the PDWB method preserves the original information of the dataset much better.*

Up to this point, we have finished the theoretical investigation. In the next section, we examine the aforementioned results using extensive simulation studies, and compare the PDWB method with some existing ones.

3 Simulations

In this section, we conduct simulations to support the theoretical findings of Section 2. First, we explain how to select ℓ_{opt} in practice in Section 3.1. Then we evaluate Theorem 2.1 in Section 3.2, and examine Examples 2.2 and 2.3 in Sections 3.3 and 3.4 respectively.

3.1 Numerical Implementation

We now comment on how to calculate ℓ_{opt} in practice. It is worth mentioning that the quantity $qc_q^2\Delta_1^2/\Delta_2$ in (2.6) in fact can be consistently estimated, so there is a data-driven $\widehat{\ell}_{\text{opt}}$ in practice. To see this, first note that c_q is decided by the kernel function, and is known. Thus, we just need to focus on Δ_1 and Δ_2.

By Theorem 2.2.1, $\frac{1}{T}\sum_{t,s=1}^{T}\overline{U}_t\overline{U}_s a\left(\frac{t-s}{T^{\nu_q}}\right) \to_P E[S_{\mathbb{N}}^2]$, where $\nu_q = 1/3$ if $q = 1$, and $\nu_q = 1/5$ if $q = 2$ by (2.7). Thus,

$$\widehat{\Delta}_2 \equiv \left(\frac{1}{T}\sum_{t,s=1}^{T}\overline{U}_t\overline{U}_s a\left(\frac{t-s}{T^{\nu_q}}\right)\right)^2 \int_{-1}^{1} a^2(x)\mathrm{d}x \to_P \Delta_2.$$

For Δ_1, let

$$\widehat{\Delta}_1 \equiv 2\sum_{k=1}^{Q_T}\frac{k^q}{T}\sum_{t=1}^{T-k}\overline{U}_t\overline{U}_{t+k}$$

where $Q_T \asymp T^{2/(4q+5)}$ is a truncation parameter. Since $\text{Var}\left(\frac{1}{T}\sum_{t=1}^{T-k}\overline{U}_t\overline{U}_{t+k} - \sigma(k)\right) = O(1/T)$ by Lemma B.2.4, we require $Q_T \asymp T^{2/(4q+5)}$ to ensure $\widehat{\Delta}_1 \to_P \Delta_1$.

Finally, we recommend the following data-driven bandwidth:

$$\widehat{\ell}_{\text{opt}} = \widehat{\ell} \vee \ell_{\min},$$

where $\widehat{\ell} = (qc_q^2\widehat{\Delta}_1^2/\widehat{\Delta}_2)^{1/(2q+1)}T^{1/(2q+1)}$, and $\ell_{\min}$ is a fixed value (say[1], $\ell_{\min} = 10$). The reason for having $\ell_{\min}$ is to boost the finite sample performance when T is relatively small. Note that even for $T = 200$, $T^{1/5}$ only returns 2.89, and it is also not guaranteed that the term $(qc_q^2\widehat{\Delta}_1^2/\widehat{\Delta}_2)^{1/(2q+1)}$ will return a value greater than 1. Therefore, to avoid an unreasonably small $\widehat{\ell}$, we use $\ell_{\min}$ to bound $\widehat{\ell}_{\text{opt}}$ from below in the numerical implementation, which does not alter any aforementioned theoretical argument. For the models considered in Section 2.3, we may simply replace $\{U_t\}$ with $\{\widehat{U}_t\}$.

3.2 Examination of Theorem 2.1

We are now ready to conduct the simulation study. The data generating process is as follows.

$$U_t^* = \rho_u U_{t-1}^* + \epsilon_t,$$

where $\epsilon_t \sim N(0_N, \Sigma_N^{\epsilon})$, and $\Sigma_N^{\epsilon} = \{\delta_{\epsilon}^{|i-j|}\}_{N\times N}$. We let $\rho_u, \rho_\epsilon \in \{0.25, 0.5\}$. To introduce heteroscedasticity, we further let $\mathbb{U}_i = \sqrt{1+i/N}\widetilde{U}_i$, where $\mathbb{U}_i = (u_{i1}, \ldots, u_{iT})^\top$, and $\widetilde{U}_i =$

[1] As indicated in the simulation study of Chen et al. (2019), 10 observations are already able to pick up a moderate serial correlation with reasonably well performance.

$(U_{i1}^*, \ldots, U_{iT}^*)^\top$ with U_{it}^* being the i^{th} element of U_t^*. Thus, u_{it} has weak correlation over both dimensions, and also has heteroskedasticity over i.

To implement the bootstrap procedure, ξ_t's are generated in the same way as mentioned under Assumption 2. We specifically consider two kernels in the following simulations.

1. The Bartlett kernel: $\psi(w) = (1 - |w|)I(|w| \le 1)$.

2. A Trapezoidal function: $a(x) = \frac{\int_{-1}^{1} w(u)w(u+|x|)\mathrm{d}u}{\int_{-1}^{1} w^2(u)\mathrm{d}u}$, where

$$w(u) = \frac{u}{0.43} I\left(u \in [0, 0.43)\right) + I\left(u \in [0.43, 0.57]\right) + \frac{1-u}{0.43} I\left(u \in (0.57, 1]\right).$$

The Bartlett kernel is well adopted in the literature for its simplicity (e.g., Andrews, 1991; Gonçalves, 2011; Bai et al., 2020; among others), while the specific form of the trapezoidal function can be seen in Shao (2010). It is worth mentioning that both kernel functions represent the cases with $q = 1$ and $q = 2$ respectively regarding (2.7). The bandwidths ℓ_B and ℓ_T of the Bartlett kernel and the trapezoidal function are selected as in Section 3.1, and, for each kernel we further consider $0.8\ell_j$ and $1.2\ell_j$ for $j \in \{B, T\}$ to examine the sensitivity.

For every generated dataset, we record the value $S_{\mathbb{N}}$ of (1.3) and the 95% confidence interval (CI) yielded by the 399 bootstrap draws. After R replications, we report

$$\text{Size} = \frac{1}{R} \sum_{m=1}^{R} I(S_{\mathbb{N},m} \notin \text{CI}_m), \tag{3.1}$$

where $S_{\mathbb{N},m}$ and CI_m respectively stand for the value of $S_{\mathbb{N}}$ and the 95% CI from the m^{th} replication.

For the purpose of comparison, we first consider three traditional methods to calculate the 95% CI. Specifically, we estimate the variances as follows:

$$s_1^2 = \frac{1}{\mathbb{N}} \sum_{i=1}^{N} \sum_{t=1}^{T} u_{it}^2, \quad s_2^2 = \frac{1}{\mathbb{N}} \sum_{i=1}^{N} \sum_{t,s=1}^{T} u_{it} u_{is}, \quad s_3^2 = \frac{1}{\mathbb{N}} \sum_{i,j=1}^{N} \sum_{t=1}^{T} u_{it} u_{jt}, \tag{3.2}$$

where s_1^2 is a consistent estimator of the variance when u_{it} is independent over (i, t), and s_2^2 and s_3^2 are consistent estimators of the variance provided that the observed u_{it} is independent over either the cross-sectional or time dimension.

The second method considered for comparison is the MBB method of Gonçalves (2011). The block length ℓ_M is generated in the same way as in Gonçalves (2011), so we omit the details here. We further consider $\lfloor 0.8\ell_M \rfloor$ and $\lfloor 1.2\ell_M \rfloor$ to examine the sensitivity. For each dataset, we also do 399 bootstrap draws to obtain the 95% CI.

The third method included for comparison is the approach of Bai et al. (2020) (referred to BCL below). The implementation is identical to Section 2.1 of Bai et al. (2020). For the two tuning parameters L (for HAC) and M (for penalization), we use $L = 3, 7, 11$ and $M = 0.1, 0.15, 0.2, 0.25$ as in Section 3 of their paper. We do not further provide the details of their approach as it is quite lengthy.

We let $R = 1000$, and summarize the results in Table 1 to Table 4. For the traditional methods, the size is always greater than 5%, which is not surprising. As all of s_j^2 for $j = 1, 2, 3$ just include a proportion of the asymptotic variance, we do expect the three traditional methods will over reject. It seems that the MBB method always generates a CI larger than it should be, which might be caused by the sensitivity of the MBB procedure itself. The BCL method tends to over reject over all. As the values of ρ_u and ρ_ϵ increase, the over reject rate increases, which might be due to the fact that many weak correlations get penalized by the thresholding method. The PDWB approach always works well throughout the four tables, and the sizes are very close to the nominal one (i.e., 5%). It is noteworthy that the PDWB procedure is not very sensitive to the choice of the kernel function, and furthermore is also robust to different choices of the bandwidth. The finite sample performance is quite consistent across Tables 1 to 4.

Table 1: Results for Section 3.2 ($\rho_u = 0.25$ and $\delta_\epsilon = 0.25$)

		PDWB						MBB			Traditional		
		Bartlett			Trapezoidal								
N	T	$0.8\ell_B$	ℓ_B	$1.2\ell_B$	$0.8\ell_T$	ℓ_T	$1.2\ell_T$	$0.8\ell_M$	ℓ_M	$1.2\ell_M$	s_1^2	s_2^2	s_3^2
50	50	0.023	0.018	0.010	0.037	0.027	0.020	0.000	0.001	0.000	0.246	0.137	0.120
	100	0.038	0.033	0.034	0.037	0.034	0.035	0.000	0.000	0.000	0.224	0.114	0.109
	200	0.051	0.045	0.047	0.058	0.049	0.046	0.000	0.000	0.000	0.258	0.123	0.140
	400	0.067	0.063	0.062	0.066	0.058	0.059	0.000	0.000	0.000	0.243	0.126	0.129
100	50	0.028	0.016	0.009	0.033	0.025	0.018	0.000	0.001	0.000	0.220	0.105	0.115
	100	0.031	0.030	0.026	0.030	0.032	0.027	0.000	0.000	0.000	0.230	0.124	0.117
	200	0.045	0.040	0.037	0.047	0.047	0.038	0.000	0.000	0.000	0.240	0.124	0.124
	400	0.045	0.052	0.051	0.044	0.043	0.043	0.000	0.000	0.000	0.212	0.104	0.104
200	50	0.022	0.012	0.007	0.034	0.026	0.019	0.000	0.000	0.000	0.232	0.120	0.113
	100	0.054	0.044	0.038	0.057	0.052	0.046	0.000	0.000	0.000	0.252	0.145	0.154
	200	0.050	0.047	0.048	0.053	0.048	0.047	0.000	0.000	0.000	0.238	0.122	0.120
	400	0.058	0.053	0.053	0.058	0.056	0.053	0.000	0.000	0.000	0.239	0.127	0.130
		BCL											
		$M = 0.1$			$M = 0.15$			$M = 0.2$			$M = 0.25$		
		$L = 3$	$L = 7$	$L = 11$	$L = 3$	$L = 7$	$L = 11$	$L = 3$	$L = 7$	$L = 11$	$L = 3$	$L = 7$	$L = 11$
50	50	0.089	0.066	0.059	0.091	0.077	0.084	0.090	0.087	0.130	0.094	0.110	0.143
	100	0.074	0.063	0.062	0.073	0.064	0.071	0.075	0.076	0.099	0.080	0.095	0.114
	200	0.108	0.088	0.086	0.113	0.093	0.096	0.114	0.098	0.121	0.114	0.119	0.136
	400	0.112	0.097	0.096	0.111	0.097	0.098	0.112	0.100	0.108	0.112	0.107	0.121
100	50	0.079	0.056	0.052	0.078	0.070	0.082	0.083	0.086	0.109	0.083	0.105	0.121
	100	0.080	0.068	0.060	0.085	0.074	0.082	0.084	0.084	0.108	0.084	0.105	0.126
	200	0.092	0.073	0.077	0.093	0.079	0.086	0.093	0.090	0.106	0.093	0.100	0.129
	400	0.076	0.070	0.067	0.076	0.070	0.071	0.079	0.074	0.086	0.080	0.081	0.093
200	50	0.087	0.065	0.061	0.086	0.069	0.082	0.089	0.087	0.119	0.091	0.112	0.120
	100	0.115	0.098	0.094	0.116	0.105	0.113	0.116	0.119	0.139	0.120	0.131	0.151
	200	0.094	0.086	0.080	0.096	0.089	0.094	0.099	0.095	0.109	0.096	0.107	0.123
	400	0.098	0.089	0.086	0.099	0.090	0.099	0.099	0.098	0.105	0.099	0.104	0.125

Table 2: Results for Section 3.2 ($\rho_u = 0.25$ and $\delta_\epsilon = 0.5$)

		PDWB						MBB			Traditional		
		Bartlett			Trapezoidal								
N	T	$0.8\ell_B$	ℓ_B	$1.2\ell_B$	$0.8\ell_T$	ℓ_T	$1.2\ell_T$	$0.8\ell_M$	ℓ_M	$1.2\ell_M$	s_1^2	s_2^2	s_3^2
50	50	0.022	0.017	0.010	0.035	0.027	0.018	0.000	0.000	0.000	0.391	0.265	0.117
	100	0.036	0.032	0.032	0.037	0.033	0.032	0.000	0.000	0.000	0.379	0.253	0.108
	200	0.051	0.043	0.046	0.051	0.048	0.045	0.000	0.000	0.000	0.380	0.278	0.135
	400	0.067	0.061	0.061	0.066	0.063	0.059	0.000	0.000	0.000	0.385	0.268	0.131
100	50	0.027	0.017	0.010	0.035	0.026	0.018	0.000	0.001	0.000	0.373	0.240	0.113
	100	0.032	0.032	0.031	0.033	0.033	0.029	0.000	0.000	0.000	0.370	0.240	0.113
	200	0.044	0.039	0.036	0.042	0.042	0.038	0.000	0.000	0.000	0.371	0.261	0.124
	400	0.049	0.050	0.051	0.044	0.045	0.049	0.000	0.000	0.000	0.342	0.219	0.105
200	50	0.022	0.012	0.006	0.032	0.022	0.017	0.000	0.000	0.000	0.378	0.255	0.113
	100	0.056	0.043	0.037	0.058	0.050	0.046	0.000	0.000	0.000	0.392	0.273	0.148
	200	0.050	0.046	0.048	0.054	0.049	0.047	0.000	0.000	0.000	0.373	0.258	0.125
	400	0.061	0.051	0.054	0.059	0.053	0.052	0.000	0.000	0.000	0.362	0.256	0.125
		BCL											
		$M = 0.1$			$M = 0.15$			$M = 0.2$			$M = 0.25$		
		$L = 3$	$L = 7$	$L = 11$	$L = 3$	$L = 7$	$L = 11$	$L = 3$	$L = 7$	$L = 11$	$L = 3$	$L = 7$	$L = 11$
50	50	0.126	0.102	0.098	0.125	0.115	0.143	0.127	0.141	0.224	0.130	0.178	0.274
	100	0.115	0.094	0.097	0.115	0.105	0.118	0.117	0.125	0.165	0.123	0.141	0.213
	200	0.142	0.131	0.127	0.144	0.130	0.141	0.146	0.142	0.154	0.148	0.148	0.191
	400	0.136	0.124	0.120	0.136	0.125	0.132	0.136	0.131	0.149	0.137	0.141	0.161
100	50	0.119	0.091	0.089	0.118	0.108	0.134	0.124	0.130	0.208	0.124	0.171	0.251
	100	0.119	0.101	0.099	0.119	0.108	0.125	0.120	0.128	0.168	0.123	0.151	0.220
	200	0.129	0.113	0.111	0.130	0.116	0.128	0.127	0.132	0.157	0.132	0.146	0.184
	400	0.112	0.102	0.100	0.113	0.106	0.108	0.115	0.113	0.124	0.118	0.117	0.142
200	50	0.111	0.102	0.097	0.117	0.115	0.154	0.119	0.147	0.239	0.125	0.196	0.256
	100	0.152	0.131	0.135	0.152	0.148	0.163	0.155	0.163	0.213	0.155	0.195	0.262
	200	0.128	0.121	0.117	0.129	0.121	0.130	0.127	0.133	0.158	0.133	0.144	0.193
	400	0.132	0.119	0.116	0.131	0.122	0.130	0.136	0.132	0.148	0.138	0.142	0.164

Table 3: Results for Section 3.2 ($\rho_u = 0.5$ and $\delta_\epsilon = 0.25$)

		PDWB						MBB			Traditional		
		Bartlett			Trapezoidal								
N	T	$0.8\ell_B$	ℓ_B	$1.2\ell_B$	$0.8\ell_T$	ℓ_T	$1.2\ell_T$	$0.8\ell_M$	ℓ_M	$1.2\ell_M$	s_1^2	s_2^2	s_3^2
50	50	0.037	0.022	0.010	0.045	0.040	0.024	0.000	0.001	0.001	0.388	0.138	0.274
	100	0.043	0.038	0.035	0.052	0.045	0.039	0.000	0.000	0.000	0.388	0.110	0.247
	200	0.073	0.061	0.057	0.071	0.071	0.055	0.000	0.000	0.000	0.382	0.126	0.274
	400	0.078	0.069	0.066	0.088	0.078	0.070	0.000	0.000	0.000	0.379	0.128	0.261
100	50	0.033	0.021	0.010	0.046	0.031	0.024	0.000	0.000	0.000	0.373	0.111	0.248
	100	0.040	0.038	0.038	0.051	0.042	0.034	0.000	0.000	0.000	0.378	0.128	0.252
	200	0.060	0.051	0.047	0.069	0.059	0.044	0.000	0.000	0.000	0.369	0.122	0.258
	400	0.062	0.057	0.054	0.060	0.056	0.057	0.000	0.000	0.000	0.347	0.107	0.220
200	50	0.034	0.021	0.013	0.049	0.033	0.021	0.000	0.001	0.000	0.374	0.120	0.252
	100	0.069	0.051	0.043	0.076	0.059	0.055	0.000	0.000	0.000	0.386	0.148	0.280
	200	0.066	0.055	0.055	0.067	0.060	0.053	0.000	0.000	0.000	0.375	0.121	0.249
		BCL											
		$M = 0.1$			$M = 0.15$			$M = 0.2$			$M = 0.25$		
		$L = 3$	$L = 7$	$L = 11$	$L = 3$	$L = 7$	$L = 11$	$L = 3$	$L = 7$	$L = 11$	$L = 3$	$L = 7$	$L = 11$
50	50	0.129	0.084	0.073	0.130	0.094	0.096	0.136	0.105	0.140	0.143	0.130	0.154
	100	0.124	0.076	0.070	0.124	0.084	0.085	0.123	0.095	0.109	0.126	0.111	0.128
	200	0.154	0.116	0.102	0.155	0.119	0.115	0.155	0.132	0.133	0.152	0.142	0.143
	400	0.151	0.110	0.105	0.152	0.118	0.109	0.155	0.121	0.122	0.157	0.124	0.132
100	50	0.122	0.072	0.060	0.123	0.084	0.090	0.124	0.098	0.126	0.127	0.128	0.133
	100	0.128	0.087	0.074	0.128	0.091	0.091	0.128	0.099	0.115	0.129	0.119	0.131
	200	0.142	0.100	0.087	0.139	0.104	0.100	0.144	0.115	0.120	0.147	0.127	0.136
	400	0.125	0.088	0.075	0.124	0.088	0.082	0.127	0.091	0.089	0.125	0.100	0.104
200	50	0.123	0.079	0.065	0.122	0.090	0.090	0.126	0.104	0.138	0.129	0.134	0.143
	100	0.159	0.120	0.107	0.160	0.125	0.125	0.163	0.138	0.150	0.166	0.159	0.171
	200	0.138	0.103	0.094	0.138	0.109	0.109	0.137	0.117	0.124	0.143	0.126	0.134
	400	0.143	0.110	0.101	0.143	0.113	0.110	0.144	0.116	0.120	0.145	0.124	0.137

Table 4: Results for Section 3.2 ($\rho_u = 0.5$ and $\delta_\epsilon = 0.5$)

		PDWB						MBB			Traditional		
		Bartlett			Trapezoidal								
N	T	$0.8\ell_B$	ℓ_B	$1.2\ell_B$	$0.8\ell_T$	ℓ_T	$1.2\ell_T$	$0.8\ell_M$	ℓ_M	$1.2\ell_M$	s_1^2	s_2^2	s_3^2
50	50	0.035	0.024	0.013	0.047	0.038	0.024	0.000	0.000	0.001	0.529	0.269	0.278
	100	0.043	0.041	0.038	0.053	0.043	0.040	0.000	0.000	0.000	0.525	0.252	0.253
	200	0.069	0.062	0.054	0.075	0.065	0.054	0.000	0.000	0.000	0.530	0.278	0.278
	400	0.080	0.074	0.068	0.088	0.075	0.066	0.000	0.000	0.000	0.502	0.267	0.271
100	50	0.039	0.023	0.007	0.051	0.032	0.019	0.000	0.002	0.000	0.490	0.256	0.257
	100	0.063	0.050	0.045	0.073	0.060	0.049	0.000	0.000	0.000	0.535	0.278	0.273
	200	0.071	0.061	0.056	0.074	0.064	0.064	0.000	0.000	0.000	0.505	0.275	0.267
	400	0.052	0.048	0.048	0.058	0.049	0.045	0.000	0.000	0.000	0.480	0.240	0.247
200	50	0.029	0.018	0.008	0.046	0.035	0.021	0.000	0.001	0.001	0.503	0.251	0.243
	100	0.051	0.051	0.041	0.066	0.054	0.048	0.000	0.000	0.000	0.494	0.239	0.251
	200	0.071	0.060	0.060	0.074	0.067	0.056	0.000	0.000	0.000	0.515	0.259	0.256
	400	0.058	0.052	0.044	0.059	0.055	0.050	0.000	0.000	0.000	0.514	0.262	0.257
		BCL											
		$M = 0.1$			$M = 0.15$			$M = 0.2$			$M = 0.25$		
		$L = 3$	$L = 7$	$L = 11$	$L = 3$	$L = 7$	$L = 11$	$L = 3$	$L = 7$	$L = 11$	$L = 3$	$L = 7$	$L = 11$
50	50	0.175	0.128	0.113	0.179	0.142	0.149	0.177	0.169	0.222	0.183	0.198	0.291
	100	0.166	0.122	0.107	0.168	0.129	0.133	0.174	0.148	0.173	0.178	0.165	0.220
	200	0.192	0.151	0.143	0.193	0.159	0.152	0.196	0.160	0.164	0.198	0.169	0.206
	400	0.197	0.147	0.136	0.198	0.151	0.145	0.199	0.157	0.162	0.200	0.167	0.172
100	50	0.175	0.132	0.115	0.177	0.143	0.159	0.178	0.161	0.244	0.182	0.200	0.288
	100	0.198	0.145	0.136	0.195	0.156	0.161	0.198	0.172	0.202	0.198	0.189	0.259
	200	0.195	0.147	0.132	0.195	0.155	0.156	0.197	0.163	0.178	0.195	0.178	0.217
	400	0.174	0.145	0.135	0.175	0.147	0.143	0.175	0.150	0.153	0.176	0.156	0.164
200	50	0.160	0.117	0.106	0.163	0.136	0.165	0.164	0.170	0.250	0.171	0.210	0.270
	100	0.187	0.134	0.128	0.187	0.149	0.162	0.189	0.172	0.205	0.195	0.190	0.241
	200	0.178	0.139	0.130	0.181	0.150	0.151	0.182	0.162	0.179	0.183	0.173	0.212
	400	0.190	0.152	0.131	0.189	0.154	0.147	0.192	0.164	0.166	0.195	0.169	0.178

3.3 Examination of Example 2.2

We now examine the results associated with Example 2.2. The data generating process is as follows:

$$Y_t = X_t\theta_0 + U_t,$$

where U_t is generated in exactly the same way as in Section 3.2, and $\theta_0 = 1$ for simplicity. The regressor X_t is generated by $X_t = 1 + N(0_N, \Sigma_N^x)$, where $\Sigma_N^x = \{0.2^{|i-j|}\}_{N\times N}$. For each generated dataset, we record $\sqrt{\mathbb{N}}(\widehat{\theta}-\theta_0)$ and the 95% CI yielded by the 399 bootstrap draws of $\sqrt{\mathbb{N}}(\widetilde{\theta}-\widehat{\theta})$. After R replications, we report

$$\text{Size} = \frac{1}{R}\sum_{m=1}^{R} I(\sqrt{\mathbb{N}}(\widehat{\theta}_m - \theta_0) \notin \text{CI}_m),$$

where $\widehat{\theta}_m$ and CI_m respectively stand for the value of $\widehat{\theta}$ and the 95% CI recorded in the m^{th} replication. The methods used for comparison are adjusted accordingly in an obvious manner, so we omit the details. Again, we let $R = 1000$.

The results are reported in Table 5 to Tabel 8. First, it is interesting to see that the MBB method no longer under rejects and in fact it over rejects across the four tables. The reason might be due to the sensitivity nature of the MBB method together with the estimation errors introduced through estimation. Except the unusual changes of the MBB method, the overall pattern presented by Table 5 to Table 8 is very similar to those shown in Section 3.2. Notably, in some (not all) cases due to the estimation error, the PDWB method converges to the nominal size slower than those seen in Section 3.2. The difference becomes more obvious when the values of ρ_u and ρ_ϵ get larger. For instances, the difference of the performance of PDWB between Table 1 and Table 5 is rather minor, but the difference between Table 4 and Table 8 is relatively significant, which is especially true when the sample size is small.

Table 5: Results for Example 2.2 of Section 3.3 ($\rho_u = 0.25$ and $\delta_\epsilon = 0.25$)

		PDWB						MBB			Traditional		
		Bartlett			Trapezoidal								
N	T	$0.8\ell_B$	ℓ_B	$1.2\ell_B$	$0.8\ell_T$	ℓ_T	$1.2\ell_T$	$0.8\ell_M$	ℓ_M	$1.2\ell_M$	s_1^2	s_2^2	s_3^2
50	50	0.114	0.122	0.130	0.097	0.109	0.117	0.159	0.158	0.163	0.155	0.111	0.112
	100	0.073	0.079	0.089	0.068	0.066	0.072	0.125	0.114	0.105	0.158	0.111	0.093
	200	0.074	0.073	0.070	0.071	0.068	0.072	0.105	0.079	0.089	0.162	0.108	0.103
	400	0.058	0.060	0.058	0.057	0.057	0.056	0.083	0.072	0.069	0.143	0.107	0.085
100	50	0.093	0.110	0.120	0.080	0.090	0.097	0.156	0.149	0.145	0.147	0.091	0.093
	100	0.079	0.081	0.086	0.066	0.070	0.081	0.136	0.123	0.110	0.173	0.118	0.114
	200	0.089	0.090	0.093	0.083	0.080	0.082	0.128	0.107	0.112	0.173	0.128	0.115
	400	0.067	0.057	0.064	0.063	0.059	0.061	0.095	0.075	0.080	0.180	0.120	0.108
200	50	0.114	0.128	0.137	0.104	0.110	0.115	0.171	0.181	0.162	0.152	0.108	0.108
	100	0.080	0.089	0.090	0.071	0.080	0.085	0.138	0.111	0.114	0.156	0.099	0.090
	200	0.056	0.057	0.064	0.053	0.055	0.055	0.097	0.080	0.084	0.146	0.098	0.089
	400	0.055	0.060	0.062	0.058	0.055	0.056	0.084	0.077	0.081	0.153	0.102	0.092
		BCL											
		$M = 0.1$			$M = 0.15$			$M = 0.2$			$M = 0.25$		
		$L = 3$	$L = 7$	$L = 11$	$L = 3$	$L = 7$	$L = 11$	$L = 3$	$L = 7$	$L = 11$	$L = 3$	$L = 7$	$L = 11$
50	50	0.102	0.106	0.113	0.099	0.104	0.104	0.102	0.100	0.106	0.106	0.106	0.112
	100	0.085	0.082	0.087	0.086	0.083	0.094	0.088	0.089	0.098	0.087	0.097	0.100
	200	0.091	0.085	0.088	0.092	0.088	0.092	0.092	0.092	0.091	0.093	0.091	0.099
	400	0.076	0.075	0.075	0.078	0.078	0.081	0.079	0.079	0.083	0.082	0.086	0.092
100	50	0.081	0.089	0.090	0.083	0.089	0.090	0.088	0.082	0.083	0.086	0.089	0.091
	100	0.099	0.094	0.093	0.097	0.101	0.100	0.098	0.102	0.115	0.100	0.112	0.117
	200	0.105	0.106	0.105	0.107	0.104	0.109	0.108	0.111	0.119	0.109	0.117	0.121
	400	0.095	0.092	0.086	0.098	0.093	0.100	0.097	0.098	0.107	0.099	0.104	0.114
200	50	0.104	0.099	0.104	0.102	0.102	0.105	0.101	0.102	0.107	0.101	0.106	0.108
	100	0.083	0.089	0.095	0.085	0.091	0.096	0.086	0.091	0.097	0.087	0.095	0.095
	200	0.077	0.077	0.079	0.078	0.083	0.087	0.081	0.086	0.092	0.084	0.088	0.095
	400	0.081	0.078	0.079	0.080	0.076	0.082	0.080	0.083	0.088	0.084	0.088	0.099

Table 6: Results for Example 2.2 of Section 3.3 ($\rho_u = 0.25$ and $\delta_\epsilon = 0.5$)

		PDWB						MBB			Traditional		
		Bartlett			Trapezoidal								
N	T	$0.8\ell_B$	ℓ_B	$1.2\ell_B$	$0.8\ell_T$	ℓ_T	$1.2\ell_T$	$0.8\ell_M$	ℓ_M	$1.2\ell_M$	s_1^2	s_2^2	s_3^2
50	50	0.125	0.132	0.145	0.114	0.118	0.123	0.172	0.160	0.164	0.266	0.207	0.119
	100	0.079	0.082	0.090	0.074	0.078	0.078	0.131	0.126	0.115	0.263	0.205	0.113
	200	0.066	0.069	0.068	0.070	0.068	0.069	0.104	0.090	0.089	0.261	0.201	0.117
	400	0.063	0.065	0.065	0.063	0.064	0.062	0.087	0.074	0.081	0.254	0.202	0.099
100	50	0.104	0.116	0.125	0.082	0.094	0.103	0.162	0.145	0.144	0.245	0.189	0.103
	100	0.084	0.094	0.100	0.071	0.076	0.078	0.141	0.124	0.115	0.284	0.228	0.126
	200	0.085	0.085	0.086	0.083	0.078	0.084	0.137	0.108	0.108	0.280	0.219	0.122
	400	0.065	0.064	0.069	0.066	0.064	0.063	0.101	0.076	0.082	0.276	0.225	0.124
200	50	0.109	0.127	0.142	0.102	0.108	0.118	0.176	0.164	0.159	0.264	0.198	0.108
	100	0.080	0.080	0.093	0.072	0.081	0.079	0.134	0.122	0.114	0.282	0.213	0.102
	200	0.059	0.063	0.063	0.060	0.059	0.058	0.102	0.087	0.085	0.236	0.179	0.106
	400	0.062	0.061	0.059	0.060	0.058	0.058	0.088	0.069	0.077	0.261	0.199	0.105
		BCL											
		$M = 0.1$			$M = 0.15$			$M = 0.2$			$M = 0.25$		
		$L = 3$	$L = 7$	$L = 11$	$L = 3$	$L = 7$	$L = 11$	$L = 3$	$L = 7$	$L = 11$	$L = 3$	$L = 7$	$L = 11$
50	50	0.131	0.144	0.157	0.131	0.149	0.165	0.135	0.153	0.187	0.134	0.164	0.204
	100	0.124	0.129	0.138	0.124	0.129	0.149	0.127	0.146	0.170	0.130	0.163	0.191
	200	0.127	0.124	0.127	0.127	0.124	0.131	0.126	0.129	0.153	0.129	0.143	0.170
	400	0.107	0.106	0.102	0.107	0.105	0.110	0.107	0.109	0.116	0.107	0.113	0.130
100	50	0.122	0.131	0.145	0.121	0.137	0.159	0.122	0.149	0.175	0.129	0.167	0.188
	100	0.140	0.140	0.152	0.139	0.144	0.158	0.142	0.158	0.194	0.152	0.178	0.219
	200	0.134	0.131	0.131	0.135	0.133	0.142	0.136	0.136	0.163	0.140	0.149	0.191
	400	0.134	0.122	0.128	0.133	0.132	0.142	0.136	0.139	0.161	0.139	0.158	0.177
200	50	0.129	0.146	0.160	0.132	0.143	0.167	0.135	0.160	0.198	0.140	0.185	0.208
	100	0.119	0.120	0.128	0.119	0.130	0.139	0.120	0.135	0.177	0.127	0.155	0.206
	200	0.117	0.110	0.111	0.118	0.118	0.125	0.119	0.123	0.141	0.122	0.129	0.165
	400	0.114	0.107	0.110	0.114	0.114	0.124	0.118	0.122	0.139	0.119	0.130	0.156

Table 7: Results for Example 2.2 of Section 3.3 ($\rho_u = 0.5$ and $\delta_\epsilon = 0.25$)

		PDWB						MBB			Traditional		
		Bartlett			Trapezoidal								
N	T	$0.8\ell_B$	ℓ_B	$1.2\ell_B$	$0.8\ell_T$	ℓ_T	$1.2\ell_T$	$0.8\ell_M$	ℓ_M	$1.2\ell_M$	s_1^2	s_2^2	s_3^2
50	50	0.144	0.146	0.155	0.128	0.129	0.135	0.220	0.189	0.193	0.255	0.122	0.207
	100	0.090	0.092	0.095	0.089	0.084	0.085	0.162	0.135	0.126	0.270	0.119	0.197
	200	0.078	0.080	0.079	0.081	0.073	0.073	0.129	0.091	0.103	0.255	0.123	0.195
	400	0.069	0.067	0.070	0.075	0.065	0.066	0.101	0.079	0.085	0.255	0.119	0.178
100	50	0.118	0.131	0.137	0.107	0.110	0.119	0.203	0.177	0.168	0.230	0.108	0.178
	100	0.103	0.099	0.105	0.094	0.085	0.094	0.169	0.128	0.136	0.267	0.126	0.213
	200	0.097	0.098	0.097	0.094	0.095	0.090	0.154	0.111	0.122	0.265	0.132	0.197
	400	0.085	0.073	0.078	0.086	0.080	0.073	0.123	0.083	0.092	0.284	0.129	0.215
200	50	0.133	0.135	0.148	0.124	0.128	0.131	0.217	0.184	0.182	0.260	0.109	0.187
	100	0.096	0.096	0.098	0.095	0.091	0.091	0.169	0.125	0.135	0.267	0.109	0.189
	200	0.069	0.068	0.070	0.069	0.064	0.062	0.125	0.089	0.097	0.234	0.100	0.169
	400	0.067	0.061	0.061	0.075	0.067	0.058	0.106	0.076	0.084	0.249	0.109	0.177
		BCL											
		$M = 0.1$			$M = 0.15$			$M = 0.2$			$M = 0.25$		
		$L = 3$	$L = 7$	$L = 11$	$L = 3$	$L = 7$	$L = 11$	$L = 3$	$L = 7$	$L = 11$	$L = 3$	$L = 7$	$L = 11$
50	50	0.141	0.125	0.123	0.138	0.122	0.124	0.139	0.126	0.123	0.142	0.126	0.127
	100	0.128	0.106	0.104	0.129	0.105	0.106	0.127	0.112	0.119	0.127	0.120	0.124
	200	0.127	0.103	0.091	0.127	0.101	0.099	0.128	0.105	0.113	0.128	0.117	0.125
	400	0.115	0.096	0.090	0.114	0.100	0.093	0.115	0.104	0.100	0.115	0.107	0.106
100	50	0.120	0.117	0.115	0.122	0.112	0.111	0.120	0.112	0.107	0.122	0.112	0.107
	100	0.134	0.111	0.112	0.136	0.116	0.113	0.136	0.120	0.129	0.136	0.129	0.139
	200	0.145	0.117	0.111	0.148	0.123	0.120	0.149	0.126	0.134	0.146	0.131	0.135
	400	0.136	0.111	0.105	0.136	0.117	0.115	0.137	0.124	0.124	0.138	0.126	0.130
200	50	0.130	0.118	0.123	0.133	0.118	0.122	0.136	0.119	0.123	0.134	0.131	0.122
	100	0.128	0.104	0.106	0.127	0.105	0.114	0.126	0.112	0.120	0.131	0.122	0.128
	200	0.108	0.085	0.085	0.109	0.089	0.093	0.112	0.097	0.100	0.110	0.102	0.106
	400	0.121	0.099	0.093	0.122	0.098	0.098	0.122	0.104	0.111	0.123	0.108	0.116

Table 8: Results for Example 2.2 of Section 3.3 ($\rho_u = 0.5$ and $\delta_\epsilon = 0.5$)

		PDWB						MBB			Traditional		
		Bartlett			Trapezoidal								
N	T	$0.8\ell_B$	ℓ_B	$1.2\ell_B$	$0.8\ell_T$	ℓ_T	$1.2\ell_T$	$0.8\ell_M$	ℓ_M	$1.2\ell_M$	s_1^2	s_2^2	s_3^2
50	50	0.151	0.152	0.160	0.134	0.137	0.137	0.244	0.179	0.202	0.376	0.228	0.231
	100	0.097	0.096	0.101	0.094	0.093	0.090	0.173	0.129	0.124	0.372	0.226	0.211
	200	0.081	0.081	0.080	0.083	0.072	0.081	0.135	0.101	0.105	0.381	0.235	0.211
	400	0.078	0.074	0.072	0.075	0.075	0.071	0.100	0.081	0.088	0.363	0.230	0.202
100	50	0.142	0.161	0.170	0.133	0.133	0.147	0.230	0.195	0.190	0.389	0.240	0.238
	100	0.097	0.099	0.098	0.094	0.091	0.092	0.182	0.127	0.143	0.374	0.230	0.222
	200	0.090	0.084	0.085	0.095	0.082	0.078	0.154	0.115	0.118	0.420	0.255	0.240
	400	0.074	0.067	0.067	0.070	0.064	0.065	0.105	0.074	0.077	0.361	0.219	0.203
200	50	0.117	0.130	0.141	0.107	0.114	0.116	0.200	0.166	0.165	0.342	0.215	0.201
	100	0.104	0.108	0.107	0.097	0.098	0.093	0.170	0.140	0.132	0.378	0.210	0.205
	200	0.077	0.083	0.086	0.084	0.071	0.075	0.145	0.104	0.113	0.395	0.229	0.216
	400	0.067	0.064	0.067	0.068	0.071	0.058	0.110	0.084	0.087	0.375	0.226	0.207
		BCL											
		$M = 0.1$			$M = 0.15$			$M = 0.2$			$M = 0.25$		
		$L = 3$	$L = 7$	$L = 11$	$L = 3$	$L = 7$	$L = 11$	$L = 3$	$L = 7$	$L = 11$	$L = 3$	$L = 7$	$L = 11$
50	50	0.199	0.180	0.189	0.198	0.186	0.198	0.200	0.197	0.222	0.201	0.207	0.247
	100	0.166	0.145	0.148	0.168	0.151	0.163	0.168	0.161	0.187	0.170	0.173	0.216
	200	0.168	0.144	0.137	0.167	0.147	0.151	0.168	0.156	0.167	0.168	0.162	0.189
	400	0.157	0.117	0.116	0.157	0.124	0.124	0.157	0.131	0.136	0.157	0.136	0.157
100	50	0.197	0.188	0.199	0.197	0.190	0.203	0.198	0.203	0.228	0.203	0.212	0.252
	100	0.180	0.148	0.152	0.181	0.158	0.169	0.183	0.169	0.200	0.182	0.195	0.227
	200	0.194	0.157	0.153	0.194	0.162	0.168	0.195	0.173	0.192	0.195	0.186	0.232
	400	0.161	0.135	0.126	0.161	0.138	0.142	0.162	0.145	0.149	0.163	0.151	0.162
200	50	0.166	0.155	0.166	0.169	0.159	0.177	0.171	0.176	0.222	0.176	0.203	0.227
	100	0.169	0.146	0.147	0.169	0.157	0.167	0.167	0.166	0.192	0.171	0.181	0.220
	200	0.168	0.143	0.146	0.169	0.152	0.159	0.171	0.163	0.177	0.173	0.174	0.206
	400	0.153	0.127	0.121	0.155	0.131	0.131	0.159	0.137	0.147	0.162	0.144	0.169

3.4 Examination of Example 2.3

Having shown the superiority of the PDWB over the other methods, in this subsection we consider Example 2.3, and focus on the PDWB method only. The DGP is as follows:

$$Y_t = X_t\theta_0 + \Gamma_0 f_t + U_t,$$

where $\theta_0 = 1$, and U_t follows the identical DGP as in Section 3.2. For the factor structure, we let $\Gamma_0 = (\gamma_{01}, \ldots, \gamma_{0N})^\top$ with $\gamma_{0i,\ell} \sim U(0.2, 2.2)$, and $f_t \sim N(0_p, I_p)$, where $\gamma_{0i,\ell}$ stands for the ℓ^{th} element of γ_{0i}. To introduce a correlation between the regressors and the factor structure, we let $X_t = \Gamma_0 1_p + v_t$, where $v_t \sim N(0_N, I_N)$. We let $p = 2$.

Note that according to the above DGP, $\{f_{0t}\}$ is in fact independent of $\{x_{it}\}$ and has mean zero, so by Corollary 2.2, it is not hard to see that no asymptotic biases will arise. Recall that as explained in Remark 2.3 our focus lies in inferences with correlation presenting in both dimensions of u_{it}, therefore such a simplification will not lose any generality. The estimation procedure and the bootstrap draws are obtained in exactly the same way as documented above Corollary 2.2. Still, we calculate the size as for Example 2.2.

After 1000 replications, we report the results in Table 9. The pattern is very much similar to those shown by Table 5 to Table 8. For example, when the values of $(\rho_u, \delta_\epsilon)$ increase, we require more sample to ensure a reasonable size. Compared to the simple panel data model of Example 2.2, it seems that having a factor structure does not yield results worse than those in Table 5 to Table 8.

Table 9: Results of Example 2.3

		$(\rho_u, \delta_\epsilon) = (0.25, 0.25)$						$(\rho_u, \delta_\epsilon) = (0.25, 0.5)$					
		Bartlett			Trapezoidal			Bartlett			Trapezoidal		
N	T	$0.8\ell_B$	ℓ_B	$1.2\ell_B$	$0.8\ell_T$	ℓ_T	$1.2\ell_T$	$0.8\ell_B$	ℓ_B	$1.2\ell_B$	$0.8\ell_T$	ℓ_T	$1.2\ell_T$
50	50	0.149	0.165	0.185	0.145	0.154	0.164	0.149	0.171	0.181	0.146	0.157	0.167
	100	0.098	0.101	0.103	0.089	0.098	0.101	0.089	0.103	0.111	0.089	0.097	0.105
	200	0.091	0.093	0.093	0.080	0.089	0.089	0.082	0.096	0.090	0.081	0.086	0.092
	400	0.056	0.063	0.061	0.057	0.056	0.055	0.060	0.061	0.069	0.056	0.062	0.058
100	50	0.162	0.177	0.187	0.151	0.164	0.170	0.160	0.170	0.192	0.147	0.158	0.172
	100	0.103	0.113	0.119	0.096	0.108	0.110	0.094	0.095	0.108	0.093	0.095	0.098
	200	0.063	0.066	0.066	0.061	0.064	0.067	0.073	0.070	0.070	0.067	0.074	0.074
	400	0.073	0.072	0.077	0.067	0.069	0.070	0.073	0.075	0.071	0.074	0.072	0.073
200	50	0.163	0.174	0.177	0.142	0.156	0.172	0.141	0.168	0.182	0.130	0.141	0.155
	100	0.082	0.088	0.090	0.078	0.087	0.087	0.089	0.094	0.096	0.083	0.087	0.093
	200	0.063	0.067	0.067	0.049	0.058	0.063	0.058	0.059	0.066	0.056	0.054	0.060
	400	0.055	0.057	0.050	0.053	0.047	0.053	0.044	0.046	0.048	0.044	0.043	0.045
		$(\rho_u, \delta_\epsilon) = (0.5, 0.25)$						$(\rho_u, \delta_\epsilon) = (0.5, 0.5)$					
		Bartlett			Trapezoidal			Bartlett			Trapezoidal		
		$0.8\ell_B$	ℓ_B	$1.2\ell_B$	$0.8\ell_T$	ℓ_T	$1.2\ell_T$	$0.8\ell_B$	ℓ_B	$1.2\ell_B$	$0.8\ell_T$	ℓ_T	$1.2\ell_T$
50	50	0.211	0.220	0.239	0.204	0.215	0.214	0.163	0.188	0.202	0.148	0.163	0.185
	100	0.115	0.127	0.136	0.115	0.121	0.122	0.102	0.112	0.125	0.098	0.103	0.113
	200	0.099	0.107	0.109	0.098	0.104	0.102	0.089	0.093	0.101	0.080	0.088	0.092
	400	0.072	0.072	0.070	0.067	0.071	0.070	0.061	0.062	0.061	0.056	0.059	0.062
100	50	0.236	0.257	0.260	0.238	0.241	0.246	0.151	0.165	0.181	0.147	0.157	0.162
	100	0.142	0.142	0.149	0.138	0.144	0.147	0.114	0.120	0.132	0.103	0.116	0.122
	200	0.077	0.077	0.077	0.072	0.071	0.078	0.072	0.070	0.078	0.065	0.065	0.070
	400	0.058	0.060	0.060	0.058	0.060	0.059	0.060	0.056	0.062	0.063	0.061	0.061
200	50	0.257	0.263	0.279	0.248	0.249	0.256	0.174	0.189	0.208	0.161	0.168	0.190
	100	0.129	0.131	0.131	0.124	0.129	0.130	0.084	0.095	0.106	0.072	0.087	0.096
	200	0.074	0.075	0.075	0.069	0.070	0.076	0.069	0.074	0.076	0.063	0.066	0.068
	400	0.052	0.058	0.057	0.052	0.054	0.056	0.053	0.050	0.053	0.050	0.050	0.056

4 An Empirical Study

In this section, we apply the proposed PDWB approach to a real dataset, evaluating the aggregated mutual fund performance.

A vast literature of financial economics has been devoted to evaluating the skills of the mutual fund managers, however the existing results present many discrepancies (Berk and Van Binsbergen, 2015), which may be due to the fact that the analyses suffer from various modelling problems. For example, the traditional approach ignores the panel nature of the dataset (Blake et al., 2014), so the inter-fund information of the cross-sectional dimension has been largely ignored. In the same spirit, Fama and French (2010, p. 1939) suggest that the TSA of the regression residuals may also alter the size of the usual fund alpha test. In this empirical study, we apply the PDWB method of Section 2, and aim to settle the discrepancies by accounting for the weak dependences along both dimensions of the panel dataset.

We obtain active U.S. equity mutual funds data from the Center for Research in Security Prices (CRSP) Survivor-Bias-Free Mutual Fund database for the period over Feb 1987 – Sep 2017, and exclude the passive index funds (e.g., Harvey and Liu, 2018). As the data are monthly collected, so T is equal to 368. We only include the funds which have initial total net assets above 10 million, and have more than 80% of their holdings in equity markets. To mitigate degree of the unbalanced panel data structure, we consider three datasets by removing the funds with more than 20%, 25%, and 30% missing values[2] during the entire period respectively, which leave us with 97, 114, and 132 mutual funds for different thresholds.

We consider the following unbalanced panel data model:

$$y_{it} = \alpha + x_t^\top \beta + u_{it},$$

where y_{it} is the net return (excluding fees and expenses) for fund i, x_t includes the Fama-French-Carhart four-factor (including the market excess return factor, the Small-Minus-Big size factor, the High-Minus-Low value factor, the momentum factor), β includes the slope coefficients, and α measures the abnormal performance of the mutual fund industry. We are interested in inferring α, which is usually considered as an average indicator of the managerial skill of fund managers since the seminal work of Jensen (1968).

[2]The thresholds 20%, 25%, and 30% are set arbitrarily. After different attempts, we note that the conclusion is not sensitive to the thresholds adopted here. In addition, we regard the three choices of the threshold as one type of robustness check.

After running the OLS regression, we obtain the estimates of α and β as follows:

$$\begin{aligned}
\text{When } N = 97,\ (\widehat{\alpha}, \widehat{\beta}^\top) &= (0.0006,\ 07573,\ 0.0224,\ 0.0319,\ -0.0022),\\
\text{When } N = 114,\ (\widehat{\alpha}, \widehat{\beta}^\top) &= (0.0005,\ 0.7542,\ 0.0295,\ 0.0170, -0.0025),\\
\text{When } N = 132,\ (\widehat{\alpha}, \widehat{\beta}^\top) &= (0.0006,\ 0.7299,\ 0.0373,\ 0.0137,\ 0.0032).
\end{aligned}$$

The estimated residuals can then be calculated as follows:

$$\widehat{u}_{it} = y_{it} - \widehat{\alpha} - x_t^\top \widehat{\beta}.$$

To show the necessity of accounting for the weak CSD and TSA, we first conduct the following two tests:

1. Examine the TSA by conducting the Ljung-Box Q-test for each time series (i.e., $\{\widehat{u}_{i1}, \ldots, \widehat{u}_{iT}\}$), and report the percentage of individuals having non-negligible autocorrelation;

2. Examine the weak CSD by conducting the CD test[3] of Pesaran (2021) on $\widehat{u}_{it}$'s, and report the test statistics.

As shown in Table 10, a non-negligible portion of individuals show evidences of the TSA, while the CD test statistic always yields a significantly large value, which indicates the presence of CSD among the residuals. It is noteworthy that the test statistic value of the CD test increases, as the threshold (of removing individuals) becomes less restrictive, so it is a strong sign of the CSD of the dataset.

Table 10: The Test Results of the Ljung-Box Q-Test and the CD Test

	Ljung-Box Q-test	CD test statistic
$N = 97$	34.02%	72.3653
$N = 114$	39.47%	76.0659
$N = 132$	40.91%	82.0605

Below, we start reporting the 95% confidence intervals (CIs) by using different methods. First, in Table 11, we present the CIs using the three traditional methods as in Section 3.3. It is clear the CIs generated by s_1^2 and s_2^2 indicate that the annualized aggregate mutual fund alpha is positively significant, that is saying the overall mutual fund industry can actually beat the market. However, the CIs generated by s_3^2 tell a different story. The results are not

[3]The asymptotic distribution of the CD test follows the standard normal distribution, so at the 5% significance level, the critical values are ± 1.96. We refer interested readers to Pesaran (2021) for more details.

very surprising given that Tabel 10 shows a reasonable amount of individuals fail to reject the null of the Ljung-Box Q-Test that assumes no time series autocorrelation.

Table 11: Estimates of the Annualized Aggregate Mutual Fund Alpha and the Slope Coefficients

		Estimates	95% CI (s_1^2)	95% CI (s_2^2)	95% CI (s_3^2)
$N = 97$	Annualized Alpha	0.72%	(0.24%, 1.08%)	(0.12%, 1.20%)	(-0.12%, 1.44%)
	Market Factor	0.7573	(0.7443, 0.7703)	(0.6811, 0.8335)	(0.7327, 0.7819)
	Size Factor	0.0224	(0.0071, 0.0376)	(-0.0154, 0.0601)	(-0.0061, 0.0508)
	Value Factor	0.0319	(0.0164, 0.0475)	(-0.0118, 0.0756)	(-0.0039, 0.0678)
	Momentum Factor	-0.0022	(-0.0130, 0.0086)	(-0.0163, 0.0119)	(-0.0224, 0.0180)
$N = 114$	Annualized Alpha	0.60%	(0.24%, 1.08%)	(0.12%, 1.20%)	(-0.24%, 1.44%)
	Market Factor	0.7542	(0.7420, 0.7664)	(0.6833, 0.8251)	(0.7295, 0.7789)
	Size Factor	0.0295	(0.0151, 0.0438)	(-0.0049, 0.0639)	(0.0023, 0.0567)
	Value Factor	0.0170	(0.0024, 0.0317)	(-0.0228, 0.0569)	(-0.0188, 0.0529)
	Momentum Factor	-0.0025	(-0.0127, 0.0077)	(-0.0158, 0.0108)	(-0.0215, 0.0165)
$N = 132$	Annualized Alpha	0.72%	(0.36%, 1.08%)	(0.24%, 1.2%)	(-0.12%, 1.56%)
	Market Factor	0.7299	(0.7181, 0.7416)	(0.6622, 0.7976)	(0.7058, 0.7540)
	Size Factor	0.0373	(0.0230, 0.0516)	(0.0033, 0.0714)	(0.0119, 0.0627)
	Value Factor	0.0137	(-0.0002, 0.0276)	(-0.0236, 0.0510)	(-0.0213, 0.0488)
	Momentum Factor	0.0032	(-0.0065, 0.0128)	(-0.0094, 0.0158)	(-0.0157, 0.0220)

In what follows, we consider the BCL, MBB, and PDWB methods and focus on the CIs associated with the annualized alpha. The implementation of these methods is identical to that of Section 3.3. The results are summarized in Table 12. Note that in Table 12 the BCL and MBB methods show mixed conclusions, while the PDWB method consistently supports the result of $\alpha = 0$ regardless the different combinations of the sample size, the bandwidth parameter, and the kernel function. Also, the consistent finding from the PDWB method agrees with Fama and French (2010), in which they conclude that the mutual fund industry cannot beat the market.

Finally, in connection with the numerical results presented in Section 3, we argue that the PDWB method shows strong evidence of its superiority over some natural competitors in finite sample studies, we thus believe the PDWB method will yield more reliable results in practice.

5 Conclusion

Although a variety of panel data models have been investigated over the past couple of decades, not much work has been done to improve these inferences associated with the

Table 12: The 95% CIs for Annualized Aggregate Mutual Fund Alpha

		BCL		
		$L = 3$	$L = 7$	$L = 11$
$N = 97$	$M = 0.1$	(-0.00%, 1.32%)	(0.00%, 1.32%)	(0.00%, 1.32%)
	$M = 0.15$	(-0.00%, 1.32%)	(0.00%, 1.32%)	(0.00%, 1.32%)
	$M = 0.2$	(-0.00%, 1.32%)	(0.00%, 1.32%)	(0.00%, 1.32%)
	$M = 0.25$	(-0.00%, 1.32%)	(0.00%, 1.32%)	(0.00%, 1.32%)
$N = 114$	$M = 0.1$	(-0.12%, 1.44%)	(-0.00%, 1.32%)	(-0.00%, 1.32%)
	$M = 0.15$	(-0.12%, 1.44%)	(-0.00%, 1.32%)	(-0.00%, 1.32%)
	$M = 0.2$	(-0.12%, 1.44%)	(-0.00%, 1.32%)	(-0.00%, 1.32%)
	$M = 0.25$	(-0.12%, 1.44%)	(-0.00%, 1.32%)	(-0.00%, 1.32%)
$N = 132$	$M = 0.1$	(0.00%, 1.44%)	(0.00%, 1.32%)	(0.00%, 1.32%)
	$M = 0.15$	(0.00%, 1.44%)	(0.00%, 1.32%)	(0.00%, 1.32%)
	$M = 0.2$	(0.00%, 1.44%)	(0.00%, 1.32%)	(0.00%, 1.32%)
	$M = 0.25$	(0.00%, 1.44%)	(0.00%, 1.32%)	(0.00%, 1.32%)
		MBB		
		$\lfloor 0.8\ell_M \rfloor$	$\lfloor \ell_M \rfloor$	$\lfloor 1.2\ell_M \rfloor$
	$N = 97$	(-0.00%, 1.44%)	(0.00%, 1.44%)	(0.00%, 1.44%)
	$N = 114$	(-0.12%, 1.56%)	(0.00%, 1.44%)	(-0.12%, 1.44%)
	$N = 132$	(0.00%, 1.56%)	(0.12%, 1.56%)	(0.00%, 1.56%)
		PDWB		
		0.8ℓ	ℓ	1.2ℓ
$N = 97$	Bartlett	(-0.12%, 1.44%)	(-0.12%, 1.44%)	(-0.12%, 1.44%)
	Trapezoidal	(-0.12%, 1.44%)	(-0.12%, 1.32%)	(-0.12%, 1.32%)
$N = 114$	Bartlett	(-0.12%, 1.56%)	(-0.12%, 1.68%)	(-0.12%, 1.68%)
	Trapezoidal	(-0.24%, 1.56%)	(-0.24%, 1.56%)	(-0.12%, 1.56%)
$N = 132$	Bartlett	(-0.00%, 1.56%)	(-0.12%, 1.68%)	(-0.00%, 1.56%)
	Trapezoidal	(-0.12%, 1.56%)	(-0.12%, 1.56%)	(-0.00%, 1.56%)

estimation of the parameters–of–interest. In this paper, we have developed a simple dependent wild bootstrap procedure to establish inferences for a wide class of panel data models, including those with interactive fixed effects. The proposed method allows for the error components to have weak cross-sectional dependence, time series autocorrelation, and heteroskedasticity. The asymptotic properties are established under a set of simple and general conditions. From a methodological point of view, our study bridges the literature of bootstrap method and the literature of HAC approach, provides a theoretical guide on how to select the optimal tuning parameter, and can better handle missing values. These new findings fill some gaps left by the bulk literature of the block bootstrap based panel data studies (e.g., Gonçalves, 2011; Palm et al., 2011). In addition, the newly proposed approach is easy to implement, and requires only one tuning parameter. Finally, we show the superiority of our approach over some natural competitors using extensive numerical studies.

In this paper, we have considered stationary time series for all individuals. We are aware of the growing literature on using bootstrap assisted methods to establish inferences for co-integrated time series models (e.g., Cavaliere et al., 2015; Reichold and Jentsch, 2022). Along this line of research, Shao (2015) provides a recent review on the bootstrap techniques frequently adopted. It would be interesting to investigate co-integrated panel data models (associated with certain cross-sectional dependence) using bootstrap methods. Such settings should be appealing in view of the increasing availability of large financial datasets over the past two decades. We leave possible extensions for future research.

References

Anderson, T. W. (1971), *The Statistical Analysis of Time Series*, first edn, John Wiley & Sons, Inc.

Andrews, D. W. K. (1991), 'Heteroskedasticity and autocorrelation consistent covariance matrix estimation', *Econometrica* **59**(3), 817–858.

Arellano, M. and Bond, S. (1991), 'Some tests of specification for panel data: Monte carlo evidence and an application to employment equations', *Review of Economic Studies* **58**(2), 277–297.

Bai, J. (2009), 'Panel data models with interactive fixed effects', *Econometrica* **77**(4), 1229–1279.

Bai, J., Choi, S. H. and Liao, Y. (2020), 'Standard errors for panel data models with unknown clusters', *Journal of Econometrics* p. forthcoming.

Bai, J. and Ng, S. (2002), 'Determining the number of factors in approximate factor models', *Econometrica* **70**(1), 191–221.

Baltagi, B. H. (2005), *Econometric Analysis of Panel Data*, third edn, John Wiley & Sons Ltd.

Berk, J. B. and Van Binsbergen, J. H. (2015), 'Measuring skill in the mutual fund industry', *Journal of Financial Economics* **118**(1), 1–20.

Blake, D., Caulfield, T., Ioannidis, C. and Tonks, I. (2014), 'Improved inference in the evaluation of mutual fund performance using panel bootstrap methods', *Journal of Econometrics* **183**(2), 202–210.

Blundell, R. and Bond, S. (1998), 'Initial conditions and moment restrictions in dynamic panel data models', *Journal of Econometrics* **87**(1), 115–143.

Cavaliere, G., Nielsen, H. B. and Rahbek, A. (2015), 'Bootstrap testing of hypotheses on co-integration relations in vector autoregressive models', *Econometrica* **83**(2), 813–831.

Chen, J., Gao, J. and Li, D. (2012), 'Semiparametric trending panel data models with cross-sectional dependence', *Journal of Econometrics* **171**(1), 71–85.

Chen, J., Li, D. and Linton, O. (2019), 'A new semiparametric estimation approach for large dynamic covariance matrices with multiple conditioning variables', *Journal of Econometrics* **212**(1), 155–176.

Chen, M., Fernández-Val, I. and Weidner, M. (2021), 'Nonlinear factor models for network and panel data', *Journal of Econometrics* **220**(2), 296–324.

Dhaene, G. and Jochmans, K. (2015), 'Split-panel Jackknife Estimation of Fixed-effect Models', *Review of Economic Studies* **82**(3), 991–1030.

Fama, E. F. and French, K. R. (2010), 'Luck versus skill in the cross-section of mutual fund returns', *Journal of Finance* **65**(5).

Feller, W. (1970), *An Introduction to Probability Theory and its Applications, Vol 2*, John Wiley & Sons.

Fitzmaurice, G. M., Laird, N. M. and Ware, J. H. (2011), *Applied Longitudinal Data Analysis*, second edn, John Wiley & Sons, Inc.

Gonçalves, S. (2011), 'The moving blocks bootstrap for panel linear regression models with individual fixed effects', *Econometric Theory* **27**(5), 1048–1082.

Gonçalves, S. and Perron, B. (2014), 'Bootstrapping factor-augmented regression models', *Journal of Econometrics* **182**(1), 156–173.

Harvey, C. R. and Liu, Y. (2018), 'Detecting repeatable performance', *Review of Financial Studies* **31**(7), 2499–2552.

Hedeker, D. and Gibbons, R. D. (2006), *Longitudinal Data Analysis*, first edn, John Wiley & Sons, Inc.

Hsiao, C. (2003), *Analysis of Panel Aata*, second edn, New York: Cambridge University Press.

Jensen, M. C. (1968), 'The performance of mutual funds in the period 1945-1964', *Journal of Finance* **23**(2), 389–416.

Jirak, M., Wu, W. B. and Zhao, O. (2021), 'Sharp connections between berry-esseen characteristics and edgeworth expansions for stationary processes', *Transactions of the American Mathematical Society* **374**(6), 4129–4183.

Kiefer, N. M. and Vogelsang, T. J. (2002), 'Heteroskedasticity-autocorrelation robust standard errors using the bartlett kernel without truncation', *Econometrica* **70**(5), 2093–2095.

Liu, W., Xiao, H. and Wu, W. B. (2013), 'Probability and moment inequalities under dependence', *Statistica Sinica* **23**(3), 1257–1272.

Menzel, K. (2021), 'Bootstrap with cluster-dependence in two or more dimensions', *Econometrica* **89**(5), 2143–2188.

Moon, H. R. and Weidner, M. (2015), 'Linear regression for panel with unknown number of factors as interactive fixed effects', *Econometrica* **83**(4), 1543–1579.

Nagaev, S. V. (1979), 'Large deviations of sums of independent random variables', *Annals of Probability* **7**(5), 745 – 789.

Newey, W. K. and West, K. D. (1987), 'A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix', *Econometrica* **55**(3), 703–708.

Onatski, A. (2010), 'Determining the number of factors from empirical distribution of eigenvalues', *Review of Economics and Statistics* **92**(4), 1004–1016.

Palm, F., Smeekes, S. and Urbain, J.-P. (2011), 'Cross-sectional dependence robust block bootstrap panel unit root tests', *Journal of Econometrics* **163**(1), 85–104.

Paparoditis, E. and Politis, D. N. (2001), 'Tapered block bootstrap', *Biometrika* **88**(4), 1105–1119.

Pesaran, M. H. (2006), 'Estimation and inference in large heterogeneous panels with a multifactor error structure', *Econometrica* **74**(4), 967–1012.

Pesaran, M. H. (2021), 'General diagnostic tests for cross section dependence in panels', *Empirical Economics* **60**, 13–50.

Petersen, M. A. (2009), 'Estimating standard errors in finance panel data sets: Comparing approaches', *Review of Financial Studies* **22**(1), 435–480.

Reichold, K. and Jentsch, C. (2022), A bootstrap-assisted self-normalization approach to inference in cointegrating regressions. `arXiv:2204.01373`.

Shao, X. (2010), 'The dependent wild bootstrap', *Journal of the American Statistical Association* **105**(489), 218–235.

Shao, X. (2015), 'Self-normalization for time series: A review of recent developments', *Journal of the American Statistical Association* **110**(512), 1797–1817.

Shi, W. and Lee, L. (2017), 'Spatial dynamic panel data models with interactive fixed effects', *Journal of Econometrics* **197**(2), 323–347.

Su, L., Jin, S. and Zhang, Y. (2015), 'Specification test for panel data models with interactive fixed effects', *Journal of Econometrics* **186**(1), 222–244.

Wu, W. B. (2005), 'Nonlinear system theory: Another look at dependence', *Proceedings of the National Academy of Sciences* **102**(40), 14150–14154.

Appendix A

The structure of the appendix is as follows. Appendix A.1 outlines the roadmap of the theoretical development of this paper. Appendix A.2 presents some notations which will be used throughout the theoretical development, and also provides some useful bounds. In Appendix A.3, we present the proofs of some of the main results.

A.1 Outline of the Theoretical Development

The roadmap of the theoretical development is as follows. In Appendix A.2, we introduce a few notations and provide some useful bounds, which facilitate the development of Proposition A.1, Proposition 2.1 and the first result of Proposition 2.2.

To derive the second result of Proposition 2.2, we prepare Lemmas B.1-B.8, some basic results (e.g., the moments conditions of Lemma B.2) of which will be also used to select the optimal bandwidth (i.e., Theorem 2.2). Due to space limit, Lemmas B.1-B.8 with their proofs are presented in the online supplementary Appendix B.

With the aforementioned results in hand, we develop Theorem 2.1. Notably, although Theorem 2.1 is stated before Theorem 2.2 in the main text, the proof of Theorem 2.2 can significantly shorten the development of Theorem 2.1.

After establishing the above results, we prove Corollary 2.2. Due to similarity, the proofs of Theorem 2.3, Corollary 2.1, and Corollary 2.3 are provided in the online supplementary Appendix B.

A.2 Mathematical Symbols and Useful Bounds

We now introduce some notations and useful bounds to facilitate the development.

For $0 \le a \le b$, we define the Berry-Esseen tail associated with $S_{\mathbb{N}}$ as follows:

$$\mathscr{T}_a^b(w) = \int_{a\le |x|\le b} e^{-ixw} E(e^{ixS_{\mathbb{N}}}) \left(1 - \frac{|x|}{b}\right) \frac{1}{x} \mathrm{d}x,$$

which arises in Berry's smoothing inequality. For $a > 0$, we define the following Berry-Esseen characteristic

$$\mathscr{C}_a = \inf_{b\ge a} \left(\sup_{w\in\mathbb{R}} |\mathscr{T}_a^b(w)| + 1/b \right).$$

Let $U_t' \equiv g(\varepsilon_t, \ldots, \varepsilon_1, \varepsilon_0', \varepsilon_{-1}, \ldots)$ and $\overline{U}_t' \equiv \frac{1}{\sqrt{N}} U_t'^\top 1_N$, where ε_0' is an independent copy of ε_0. Define $\theta_{t,\delta}^U = \|\overline{U}_t - \overline{U}_t'\|_\delta$, which is fully bounded by $\lambda_{t,\delta}^U + \lambda_{t+1,\delta}^U$. To see this, write

$$\begin{aligned}\theta_{t,\delta}^U &= \|\overline{U}_t - \overline{U}_t'\|_\delta \le \|\overline{U}_t - \overline{U}_t^*\|_\delta + \|\overline{U}_t^* - \overline{U}_t'\|_\delta \\ &= \lambda_{t,\delta}^U + \|\overline{U}_{t+1}^* - \overline{U}_{t+1}\|_\delta = \lambda_{t,\delta}^U + \lambda_{t+1,\delta}^U, \end{aligned} \tag{A.1}$$

in which the second equality follows from $\overline{U}_t^* - \overline{U}_t' =_D \overline{U}_{t+1}^* - \overline{U}_{t+1}$. (A.1) infers that the conditions imposed on $\lambda_{t,\delta}^U$ in Assumption 1 also apply to $\theta_{t,\delta}^U$. For the same purpose, for $0 \le m \le t$, let $U_t^{(m,\prime)} \equiv g(\varepsilon_t, \ldots, \varepsilon_{t-m+1}, \varepsilon_{t-m}', \varepsilon_{t-m-1}, \ldots)$, and $U_t^{(m,*)} \equiv g(\varepsilon_t, \ldots, \varepsilon_{t-m+1}, \varepsilon_{t-m}', \varepsilon_{t-m-1}', \ldots)$. Accordingly, we have $\overline{U}_t^{(m,\prime)}$ and $\overline{U}_t^{(m,*)}$. In addition, define $U_t^{(m,**)}$ and $\overline{U}_t^{(m,**)}$ using $\{\varepsilon_t''\}$, which is another independent copy of $\{\varepsilon_t\}$.

Define the σ-field $\mathscr{F}_t = \sigma(\varepsilon_t, \varepsilon_{t-1}, \ldots)$ and the projection operator:

$$\mathcal{P}_t(\cdot) = E[\cdot \mid \mathscr{F}_t] - E[\cdot \mid \mathscr{F}_{t-1}].$$

For $1 \leq \delta^* \leq \delta$ and some integer $0 \leq m \leq t$,

$$\begin{aligned}
\|\mathcal{P}_{t-m}(\overline{U}_t)\|_{\delta^*} &= \|E[\overline{U}_t \mid \mathscr{F}_{t-m}] - E[\overline{U}_t \mid \mathscr{F}_{t-m-1}]\|_{\delta^*} = \|E[\overline{U}_t \mid \mathscr{F}_{t-m}] - E[\overline{U}_t^{(m,\prime)} \mid \mathscr{F}_{t-m-1}]\|_{\delta^*} \\
&= \|E[\overline{U}_t - \overline{U}_t^{(m,\prime)} \mid \mathscr{F}_{t-m}]\|_{\delta^*} \leq \|\overline{U}_t - \overline{U}_t^{(m,\prime)}\|_{\delta^*} \leq \|\overline{U}_t - \overline{U}_t^{(m,\prime)}\|_{\delta} \\
&= \|\overline{U}_m - \overline{U}'_m\|_{\delta} = \theta^U_{m,\delta}, \qquad \text{(A.2)}
\end{aligned}$$

where the first inequality follows from Jensen's inequality, the second inequality follows from the moments monotonicity, and the fourth equality follows from $\overline{U}_t - \overline{U}_t^{(m,\prime)} =_D \overline{U}_m - \overline{U}'_m$.

Let $\gamma \equiv \gamma_T \to \infty$ and $\gamma/T \to 0$. For $t \geq s$, define $\mathscr{F}_{t,s} = \sigma(\varepsilon_t, \ldots, \varepsilon_s)$, $\mathscr{F}^*_{t,s} = \sigma(\varepsilon'_t, \ldots, \varepsilon'_s)$, and $\overline{U}_{t\gamma} = E[\overline{U}_t \mid \mathscr{F}_{t,t-\gamma}]$. Further, we let

$$\overline{U}^*_{t\gamma} = \begin{cases} E[\overline{U}^*_t \mid \mathscr{F}_{t,1}, \mathscr{F}^*_{0,t-\gamma}], & \text{for } 1 \leq t \leq \gamma \\ \overline{U}_{t\gamma}, & \text{for } t > \gamma \end{cases}.$$

With the above definitions in hand, we are ready to present a proposition which will be used repeatedly in the theoretical development.

Proposition A.1. *Under Assumption 1, we have*

1. $\|\sum_{t=1}^T \overline{U}_t\|_\delta = O(\sqrt{T})$,
2. $\sum_{t=1}^\infty t^2 \|\overline{U}_{t\gamma} - \overline{U}^*_{t\gamma}\|_\delta < \infty$.

A.3 Proofs of Some Main Results

Proof of Proposition A.1:

(1). Write

$$\begin{aligned}
&\left\|\sum_{t=1}^T \overline{U}_t\right\|_\delta = \left\|\sum_{t=1}^T \sum_{m=0}^\infty \mathcal{P}_{t-m}(\overline{U}_t)\right\|_\delta \leq \sum_{m=0}^\infty \left\|\sum_{t=1}^T \mathcal{P}_{t-m}(\overline{U}_t)\right\|_\delta \\
&\leq O(1) \sum_{m=0}^\infty \left(E\left[\sum_{t=1}^T |\mathcal{P}_{t-m}(\overline{U}_t)|^2\right]^{\delta/2}\right)^{\frac{1}{\delta}} = O(1) \sum_{m=0}^\infty \left(E\left[\sum_{t=1}^T |\mathcal{P}_{t-m}(\overline{U}_t)|^2\right]^{\delta/2}\right)^{\frac{2}{\delta}\cdot\frac{1}{2}} \\
&\leq O(1) \sum_{m=0}^\infty \left\{\sum_{t=1}^T \left(E\left[|\mathcal{P}_{t-m}(\overline{U}_t)|^2\right]^{\delta/2}\right)^{\frac{2}{\delta}}\right\}^{\frac{1}{2}} = O(1) \sum_{m=0}^\infty \left\{\sum_{t=1}^T \left(E|\mathcal{P}_{t-m}(\overline{U}_t)|^\delta\right)^{\frac{2}{\delta}}\right\}^{\frac{1}{2}} \\
&\leq O(1) \sum_{m=0}^\infty \left(T|\theta^U_{m,\delta}|^2\right)^{\frac{1}{2}} = O(1) T^{\frac{1}{2}} \sum_{m=0}^\infty \theta^U_{m,\delta} = O(T^{1/2}),
\end{aligned}$$

where the first inequality follows from the triangle inequality, the second inequality follows from the Burkholder's inequality, the third inequality follows from the Minkowski inequality, the fourth inequality follows from (A.2), and the fifth equality follows from (A.1) and Assumption 1.

(2). Similar to $\overline{U}_{t\gamma}^*$, we define $\overline{U}_{t\gamma}^{**}$ using $\{\varepsilon_t''\}$, which is another independent copy of $\{\varepsilon_t\}$. Then for $1 \le t \le \gamma$

$$\begin{aligned}
\|\overline{U}_{t\gamma} - \overline{U}_{t\gamma}^*\|_\delta &= \|E[\overline{U}_{t\gamma} - \overline{U}_{t\gamma}^* + \overline{U}_{t\gamma}^* \mid \mathscr{F}_{t,t-\gamma}] - E[\overline{U}_{t\gamma}^* - \overline{U}_{t\gamma}^{**} + \overline{U}_{t\gamma}^{**} \mid \mathscr{F}_{t,1}, \mathscr{F}_{0,t-\gamma}^*]\|_\delta \\
&= \|E[\overline{U}_{t\gamma} - \overline{U}_{t\gamma}^* \mid \mathscr{F}_{t,t-\gamma}] - E[\overline{U}_{t\gamma}^* - \overline{U}_{t\gamma}^{**} \mid \mathscr{F}_{t,1}, \mathscr{F}_{0,t-\gamma}^*]\|_\delta \\
&\le 2\|\overline{U}_t - \overline{U}_t^*\|_\delta = 2\lambda_{t,\delta}^U,
\end{aligned}$$

where the second equality follows from the fact that $E[\overline{U}_{t\gamma}^* \mid \mathscr{F}_{t,t-\gamma}] = E[\overline{U}_{t\gamma}^{**} \mid \mathscr{F}_{t,1}, \mathscr{F}_{0,t-\gamma}^*]$, and the inequality follows from the Jensen's inequality. For $t > \gamma$,

$$\|\overline{U}_{t\gamma} - \overline{U}_{t\gamma}^*\|_\delta = 0.$$

In connection with Assumption 1, the result follows immediately. ■

Verification for the statement of Example 2.1:

Without loss of generality, let $\delta = 4$ in what follows. Then write

$$\begin{aligned}
\|\overline{U}_t - \overline{U}_t^*\|_4 &= \left(E\left|\sum_{j=t}^{\infty} \frac{1}{\sqrt{N}} 1_N^\top B_j(\varepsilon_{t-j} - \varepsilon_{t-j}')\right|^4\right)^{1/4} \\
&\le \left(\sum_{j=t}^{\infty} E\left|\frac{1}{\sqrt{N}} 1_N^\top B_j(\varepsilon_{t-j} - \varepsilon_{t-j}')\right|^4\right)^{1/4} + 6\left(\sum_{j=t}^{\infty} \frac{1}{N} 1_N^\top B_j B_j^\top 1_N\right)^{1/2} \\
&:= A_1 + A_2,
\end{aligned}$$

where the definitions of A_1 and A_2 are obvious.

Consider A_1. For notational simplicity, let $B_j = \{B_{j,kl}\}_{k,l\in[N]}$ and $\frac{1}{\sqrt{N}} 1_N^\top B_j = (B_{j,\cdot 1}, \ldots, B_{j,\cdot N})$. As $\{\varepsilon_{it}\}$ are independent over i, we can write

$$\begin{aligned}
& E\left|\frac{1}{\sqrt{N}} 1_N^\top B_j(\varepsilon_{k-j} - \varepsilon_{k-j}')\right|^4 \\
&= E\left|\sum_{l=1}^{N} B_{j,\cdot l}^2 (\varepsilon_{l,k-j} - \varepsilon_{l,k-j}')^2\right|^2 + 4E\left|\sum_{l=1}^{N-1}\sum_{k=l+1}^{N} B_{j,\cdot l} B_{j,\cdot k} (\varepsilon_{l,t-j} - \varepsilon_{l,t-j}')(\varepsilon_{k,t-j} - \varepsilon_{k,t-j}')\right|^2 \\
&\le O(1)\left(\sum_{l=1}^{N} B_{j,\cdot l}^2\right)^2 + O(1)\sum_{l=1}^{N-1}\sum_{k=l+1}^{N} B_{j,\cdot l}^2 B_{j,\cdot k}^2 \\
&\le O(1)\left(\frac{1}{N} 1_N^\top B_j B_j^\top 1_N\right)^2,
\end{aligned}$$

where the first inequality follows from some direct calculation, and the second inequality follows from $\sum_{l=1}^{N-1}\sum_{k=l+1}^{N} B_{j,.l}^2 B_{j,.k}^2 \le (\sum_{l=1}^{N} B_{j,.l}^2)^2$. Finally, we can write

$$A_1 \le O(1)\left\{\sum_{j=t}^{\infty}\left(\frac{1}{N}1_N^\top B_j B_j^\top 1_N\right)^2\right\}^{1/4} \le O(1)\left\{\sum_{j=t}^{\infty}|B_j|^4\right\}^{1/4} = O(\rho^t).$$

For A_2, it is obvious that $6\left\{\sum_{j=t}^{\infty}|B_j|^2\right\}^{1/2} = O(\rho^t)$.

Based on the above development,

$$\sum_{t=0}^{\infty} t^2\|\overline{U}_t - \overline{U}_t^*\|_4 \le O(1)\sum_{t=0}^{\infty} t^2\rho^t < \infty,$$

so Assumption 1 is met. The proof is now completed. ■

Proof of Proposition 2.1:

CSD — Note that $\overline{U}_t$ can be decomposed as $\overline{U}_t = \sum_{m=0}^{\infty}\mathcal{P}_{t-m}(\overline{U}_t)$, so we write

$$\|\overline{U}_t\|_{\delta^*} \le \sum_{m=0}^{\infty}\|\mathcal{P}_{t-m}(\overline{U}_t)\|_{\delta^*} \le \sum_{m=0}^{\infty}\theta_{m,\delta}^U \le \sum_{m=0}^{\infty}(\lambda_{m,\delta}^U + \lambda_{m+1,\delta}^U) < \infty,$$

where the second inequality follows from (A.2), the third inequality follows from (A.1), and the last inequality follows from Assumption 1. Then the first result follows.

TSA — Since $E[\overline{U}_t] = E[\overline{U}_t^* \mid \mathscr{F}_0] = 0$, we have

$$\sum_{t=1}^{\infty} t^2|E[\overline{U}_t\overline{U}_0]| = \sum_{t=1}^{\infty} t^2|E[E(\overline{U}_t - \overline{U}_t^* \mid \mathscr{F}_0)\overline{U}_0]| \le \sum_{t=1}^{\infty} t^2\|\overline{U}_t - \overline{U}_t^*\|_2\|\overline{U}_0\|_2 = O(1),$$

where the first equality follows from the independence between $\overline{U}_t^*$ and $\overline{U}_0$, and the inequality follows from Cauchy-Schwarz inequality. The proof is now completed. ■

Proposition of 2.2:

(1). Let $S_{\mathbb{N},L} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{l=0}^{L-1}\mathcal{P}_{t-l}(\overline{U}_t)$ and $\widehat{S}_{\mathbb{N},L} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{l=0}^{L-1}\mathcal{P}_t(\overline{U}_{t+l})$, in which $L \to \infty$ and $L/T \to 0$.

Note that $\overline{U}_t = \sum_{l=0}^{\infty}\mathcal{P}_{t-l}(\overline{U}_t)$ and $\{\mathcal{P}_{t-l}(\overline{U}_t)\}_{t=1}^T$ is a sequence of martingale differences, and thus

$$\begin{aligned}\|S_{\mathbb{N},L} - S_{\mathbb{N}}\|_2 &= \left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{l=L}^{\infty}\mathcal{P}_{t-l}(\overline{U}_t)\right\|_2 \le \frac{1}{\sqrt{T}}\sum_{l=L}^{\infty}\left\|\sum_{t=1}^{T}\mathcal{P}_{t-l}(\overline{U}_t)\right\|_2 \\ &\le O(1)\frac{1}{\sqrt{T}}\sum_{l=L}^{\infty}\left\{\sum_{t=1}^{T}E\left[\left(\mathcal{P}_{t-l}(\overline{U}_t)\right)^2\right]\right\}^{1/2} \le O(1)\sum_{l=L}^{\infty}\theta_{l,2}^U \to 0,\end{aligned}$$

where the second inequality follows from Burkholder's inequality, the third inequality follows from (A.2), and the last step follows from $L \to \infty$. Similarly, by using Burkholder's inequality, we have

$$\|\widehat{S}_{\mathbb{N},L} - S_{\mathbb{N},L}\|_2 \le \frac{1}{\sqrt{T}}\sum_{l=0}^{L-1}\left\|\sum_{t=1}^{l}\mathcal{P}_{t-l}(\overline{U}_t)\right\|_2 + \frac{1}{\sqrt{T}}\sum_{l=0}^{L-1}\left\|\sum_{t=T-l+1}^{T}\mathcal{P}_t(\overline{U}_{t+l})\right\|_2 = O(1)\frac{\sqrt{L}}{\sqrt{T}} \to 0.$$

Hence, we have $\|\widehat{S}_{\mathbb{N},L} - S_{\mathbb{N}}\|_2 \to 0$.

Note that $\{\sum_{l=0}^{L-1} \mathcal{P}_t(\overline{U}_{t+l})\}_{t=1}^T$ is a sequence of martingale differences subject to $\mathscr{F}_t$, the asymptotic normality can be easily obtained by using a standard martingale central limit theory. The proof of the first result is now complete.

(2). We first define $S_{\mathbb{N}\gamma} = \frac{1}{\sqrt{T}} \sum_{t=1}^T \overline{U}_{t\gamma}$, and establish some rates of convergence associated with $S_{\mathbb{N}}$ and $S_{\mathbb{N}\gamma}$. Using the same argument as in the proof of Theorem 1 of Liu et al. (2013), it follows that

$$\|S_{\mathbb{N}} - S_{\mathbb{N}\gamma}\|_\delta \le O(1)\gamma^{-2} \sum_{t=\gamma}^{\infty} t^2 \|\overline{U}_t - \overline{U}_t^*\|_\delta = O(\gamma^{-2}).$$

Hence, by Cauchy-Schwarz inequality and Proposition A.1, we have

$$|E(S_{\mathbb{N}}^2) - E(S_{\mathbb{N}\gamma}^2)| \le \|S_{\mathbb{N}} - S_{\mathbb{N}\gamma}\|_2 \cdot \|S_{\mathbb{N}} + S_{\mathbb{N}\gamma}\|_2 = O(\gamma^{-2}).$$

Similarly, by Hölder's inequality, we have

$$|E(S_{\mathbb{N}}^3) - E(S_{\mathbb{N}\gamma}^3)| \le \|S_{\mathbb{N}} - S_{\mathbb{N}\gamma}\|_3 \cdot \left(\|S_{\mathbb{N}}\|_3^2 + \|S_{\mathbb{N}}\|_3 \|S_{\mathbb{N}\gamma}\|_3 + \|S_{\mathbb{N}\gamma}\|_3^2\right) = O(\gamma^{-2}).$$

We are now ready to start the investigation:

$$\begin{aligned}
\sup_{w\in\mathbb{R}} |\Pr(S_{\mathbb{N}} \le w) - \Psi_{\mathbb{N}}(w)| &\le \sup_{w\in\mathbb{R}} |\Pr(S_{\mathbb{N}} \le w) - \Pr(\overline{Z} + \widetilde{Z} \le w)| \\
&\quad + \sup_{w\in\mathbb{R}} |\Pr(\overline{Z} + \widetilde{Z} \le w) - \Psi_{\mathbb{N}}(w)| \\
&:= I_{T,1} + I_{T,2},
\end{aligned}$$

where $\overline{Z}$ and $\widetilde{Z}$ are defined in Section B.2 of the supplementary Appendix B.

Consider $I_{T,1}$. By $e^{ix} = \cos(x) + i\sin(x)$ and Lipschitz continuity, we have

$$|E(e^{ixS_{\mathbb{N}}}) - E(e^{ixS_{\mathbb{N}\gamma}})| = O(|x|\gamma^{-2}).$$

Hence, by the Berry's smoothing inequality (Lemma 2, XVI.3 in Feller, 1970), we have

$$\begin{aligned}
I_{T,1} &\le \int_{-c\sqrt{T}}^{c\sqrt{T}} |E(e^{ixS_{\mathbb{N}}}) - E(e^{ix(\overline{Z}+\widetilde{Z})})| \frac{1}{|x|} \mathrm{d}x + \mathscr{C}_{c\sqrt{T}} \\
&= \mathscr{U}_T + O(\sqrt{T}/\gamma^2) + \mathscr{C}_{c\sqrt{T}},
\end{aligned}$$

where $\mathscr{U}_T$ is defined in Section B.2, and $\mathscr{C}_{c\sqrt{T}}$ is defined with respect to $\overline{U}_t$ and is given in the beginning of Section A.2. Selecting γ large enough and using Lemma B.8.2, we have

$$\sup_{w\in\mathbb{R}} |\Pr(S_{\mathbb{N}} \le w) - \Pr(\overline{Z} + \widetilde{Z} \le w)| = O(T^{-1}(\log T)^5) + \mathscr{C}_{c\sqrt{T}}.$$

We now consider $\mathscr{C}_{c\sqrt{T}}$. Recall that we have defined $(\mathscr{T}_a^b)^\diamond$, H_t's, and $S_{\mathbb{N}}^\diamond$ in Section B.2 of the online Appendix B below. First, note that selecting $a > 0$ such that $c > ab$ and using $E(e^{ixS_{\mathbb{N}}^\diamond}) = 0$ for $|x| > \sqrt{T}|ab|$, we have $\sup_{w\in\mathbb{R}}(\mathscr{T}_{c\sqrt{T}}^\infty)^\diamond(w) = 0$. Also, note that by Taylor expansion,

$$e^{ixS_{\mathbb{N}}^\diamond} - e^{ixS_{\mathbb{N}}} = \frac{\partial e^{ixz}}{\partial z}\Big|_{z=S_{\mathbb{N}}} \cdot \frac{1}{\sqrt{T}}(H_T - H_0) + \frac{\partial^2 e^{ixz}}{\partial z^2}\Big|_{z=S_{\mathbb{N}}} \cdot \frac{1}{T}(H_T - H_0)^2 + o\left(\frac{1}{T}(H_T - H_0)^2\right).$$

Hence, it is easy to know that $|E(e^{ixS_{\mathbb{N}}}) - E(e^{ixS_{\mathbb{N}}^\diamond})| = O(T^{-1})$, which thus yields $\mathscr{C}_{c\sqrt{T}} = O(T^{-1})$. We now can conclude that $I_{T,1} = O(T^{-1}(\log T)^5)$.

Finally, we consider $I_{T,2}$. Define $\Psi_{\mathbb{N}\gamma}(w)$ in analogy to $\Psi_{\mathbb{N}}(w)$ with respect to $S_{\mathbb{N}\gamma}$. Write

$$I_{T,2} \le \sup_{w\in\mathbb{R}} |\Pr(\overline{Z} + \widetilde{Z} \le w) - \Psi_{\mathbb{N}\gamma}(w)| + \sup_{w\in\mathbb{R}} |\Psi_{\mathbb{N}\gamma}(w) - \Psi_{\mathbb{N}}(w)| = O(\gamma^{-1} + \gamma^{-2}),$$

where the last step follows from Lemma B.6.2, and the facts that $|E(S_{\mathbb{N}}^2) - E(S_{\mathbb{N}\gamma}^2)| = O(\gamma^{-2})$ and $|E(S_{\mathbb{N}}^3) - E(S_{\mathbb{N}\gamma}^3)| = O(\gamma^{-2})$ as shown in the beginning of the proof of this result. Selecting $\gamma = T/(\log T)^5$, we have $I_{T,2} = O(T^{-1}(\log T)^5)$.

Collecting the above results, the proof is now completed. ■

Proof of Theorem 2.2:

(1). For ease of notation, let $s_{\mathbb{N}}^{*2} := E^*[S_{\mathbb{N}}^{*2}] = \frac{1}{T}\sum_{t,s=1}^T \overline{U}_t\overline{U}_s a\left(\frac{t-s}{\ell}\right)$, and write

$$\begin{aligned}\ell^q\left(E[s_{\mathbb{N}}^{*2}] - s_{\mathbb{N}}^2\right) &= \ell^q \sum_{k=-\ell}^{\ell} [a(k/\ell) - 1]\, E(\overline{U}_k\overline{U}_0) - 2\ell^q \sum_{k=1}^{\ell} \frac{k}{T}[a(k/\ell) - 1]\, E(\overline{U}_k\overline{U}_0)\\ &\quad -2\ell^q \sum_{k=\ell+1}^{T-1} \frac{T-k}{T} E(\overline{U}_k\overline{U}_0) := I_1 + I_2 + I_3,\end{aligned}$$

where the definitions of I_1 to I_3 are obvious. We consider these three terms one by one.

Consider I_1. By Assumption 3, for $\forall\epsilon > 0$, we choose $\nu_\epsilon > 0$ such that

$$|k/\ell| < \nu_\epsilon \quad \text{and} \quad \left|\frac{1 - a(k/\ell)}{|k/\ell|^q} - c_q\right| < \epsilon.$$

Letting $\ell_T^* = \lfloor \nu_\epsilon \ell \rfloor$, write

$$I_1 = \sum_{k=-\ell_T^*}^{\ell_T^*} \frac{a(k/\ell) - 1}{|k/\ell|^q} |k|^q E(\overline{U}_k\overline{U}_0) + 2\sum_{k=\ell_T^*+1}^{\ell} \frac{a(k/\ell) - 1}{|k/\ell|^q} |k|^q E(\overline{U}_k\overline{U}_0).$$

Then, by Proposition 2.1, it is easy to see that the first term of the right-hand side converges to $-c_q\sum_{k=-\infty}^{\infty} |k|^q E(\overline{U}_0\overline{U}_k)$. For the second term, since $|a(\cdot)| \le M$, $\left|\frac{1-a(k/\ell)}{|k/\ell|^q}\right| \le (M+1)/\nu_\epsilon^q$ due to the fact that $k/\ell \ge \nu_\epsilon$. Then this term is bounded by

$$(M+1)/\nu_\epsilon^q \sum_{k=\ell_T^*+1}^{\infty} k^2 |E(\overline{U}_k\overline{U}_0)|,$$

which, by Proposition 2.1 again, converges to 0 as $\ell \to \infty$.

Consider I_2. As $\ell^q/T \to 0$ of Assumption 2 and $\sum_{k=1}^{\infty} k^2|E(\overline{U}_k\overline{U}_0)| < \infty$ of Proposition 2.1, we have

$$|I_2| \le \frac{\ell^q}{T} 2(M+1) \sum_{k=1}^{\infty} k^2 |E(\overline{U}_k\overline{U}_0)| \to 0.$$

Consider I_3 and write

$$|I_3| \le 2\frac{\ell^q}{\ell^2} \sum_{k=\ell+1}^{\infty} k^2|E(\overline{U}_k\overline{U}_0)| \to 0,$$

where the last steps follows from the facts that $\frac{\ell^q}{\ell^2}$ is bounded and $\ell \to \infty$. The proof of the first result is now completed.

(2). Define $v(r,s,t) = E(\overline{U}_0\overline{U}_r\overline{U}_s\overline{U}_t)$, $\sigma(r) = E(\overline{U}_0\overline{U}_r)$ and

$$\kappa(r,s,t) = v(r,s,t) - \sigma(r)\sigma(s-t) - \sigma(s)\sigma(r-t) - \sigma(t)\sigma(r-s).$$

Then write

$$\begin{aligned}
\frac{T}{\ell}\mathrm{Var}\left(s_{\mathbb{N}}^{*2}\right) &= \frac{T}{\ell}\sum_{i=-\ell}^{\ell}\sum_{k=-\ell}^{\ell} a\left(\frac{i}{\ell}\right) a\left(\frac{k}{\ell}\right) \mathrm{Cov}\left(\frac{1}{T}\sum_{t=1+|i|}^{T}\overline{U}_t\overline{U}_{t-|i|}, \frac{1}{T}\sum_{t=1+|k|}^{T}\overline{U}_t\overline{U}_{t-|k|}\right) \\
&= \frac{1}{\ell}\sum_{i=-\ell}^{\ell}\sum_{k=-\ell}^{\ell} a\left(\frac{i}{\ell}\right) a\left(\frac{k}{\ell}\right)\left(\frac{1}{T}\sum_{t=1+|i|}^{T}\sum_{s=1+|k|}^{T} E(\overline{U}_t\overline{U}_{t-|i|}\overline{U}_s\overline{U}_{s-|k|})\right. \\
&\quad \left. -\frac{1}{T}\sum_{t=1+|i|}^{T}\sum_{s=1+|k|}^{T} E(\overline{U}_t\overline{U}_{t-|i|})E(\overline{U}_s\overline{U}_{s-|k|})\right) \\
&= \frac{1}{\ell}\sum_{i=-\ell}^{\ell}\sum_{k=-\ell}^{\ell} a\left(\frac{i}{\ell}\right) a\left(\frac{k}{\ell}\right)\sum_{r=1-T}^{T-1}\phi_T(r,i,k)[\sigma(r)\sigma(r+k-i) + \\
&\quad \sigma(r-i)\sigma(r+k) + \kappa(k,-r,i-r)] := I_1 + I_2 + I_3,
\end{aligned}$$

where, by tedious but trivial calculation (e.g., Chapter 9 of Anderson, 1971), it is easy to know that $\lim_{T\to\infty}\phi_T(r,i,k) = 1$ for every r,i,k, $0 \le \phi_T(r,i,k) \le 1$ and $\phi_T(r,i,k) \ge 1 - \frac{|r|+|i|+|k|}{T}$.

Consider I_1. Write

$$\begin{aligned}
I_1 &= \frac{1}{\ell}\sum_{u=1-T}^{T-1}\sum_{v=u-2\ell}^{u+2\ell}\sum_{s=\max(u,v)-\ell}^{\min(u,v)+\ell} a\left(\frac{u-s}{\ell}\right) a\left(\frac{v-s}{\ell}\right)\phi_T(u,u-s,v-s)\sigma(u)\sigma(v) \\
&= \frac{1}{\ell}\sum_{u,v=-m}^{m}\sum_{s=\max(u,v)-\ell}^{\min(u,v)+\ell} a\left(\frac{u-s}{\ell}\right) a\left(\frac{v-s}{\ell}\right)\phi_T(u,u-s,v-s)\sigma(u)\sigma(v) + o(1) \\
&= \frac{1}{\ell}\sum_{u,v=-m}^{m}\sum_{s=-\ell}^{\ell} a\left(\frac{s}{\ell}\right)^2 \phi_T(u,u-s,v-s)\sigma(u)\sigma(v) + o(1) \\
&= \frac{1}{\ell}\sum_{u,v=-m}^{m}\sum_{s=-\ell}^{\ell} a\left(\frac{s}{\ell}\right)^2 \sigma(u)\sigma(v) + o(1) \to s_{\mathbb{N}}^4 \int_{-1}^{1} a^2(x)\mathrm{d}x
\end{aligned}$$

by $\sum_{i=-\infty}^{\infty} |\sigma(i)| < \infty$ and selecting some $m \to \infty$ and $m/\ell \to 0$. Similarly, we have

$$I_2 \to s_{\mathbb{N}}^4 \int_{-1}^{1} a^2(x)\mathrm{d}x.$$

To complete the proof, it suffices to show $I_3 \to 0$. In view of the facts that $a(\cdot)$ is finite and $1/\ell \to 0$, we just need to show $\sum_{r,s,t=-\infty}^{\infty} |\kappa(r,s,t)| < \infty$. By construction, we have $\sum_{r,s,t=-\infty}^{\infty} |\kappa(r,s,t)| \le O(1) \sum_{0\le r\le s\le t<\infty} |\kappa(r,s,t)|$, so focus on $\sum_{0\le r\le s\le t<\infty} |\kappa(r,s,t)|$ below.

$$\begin{aligned}\sum_{0\le r\le s\le t<\infty} |\kappa(r,s,t)| &= \sum_{0<r<s<t<\infty} |\kappa(r,s,t)| + \sum_{0=r<s<t<\infty} |\kappa(r,s,t)| \\ &\quad + \sum_{0\le r<s=t<\infty} |\kappa(r,s,t)| + \sum_{0\le r=s\le t<\infty} |\kappa(r,s,t)| \\ &:= I_4 + I_5 + I_6 + I_7.\end{aligned}$$

The most difficult term to deal with is I_4. Since

$$E[\overline{U}_0]E\left[\left(\overline{U}_r\overline{U}_s\overline{U}_t\right)^{(t-0,*)}\right] = E\left[\overline{U}_0\left(\overline{U}_r\overline{U}_s\overline{U}_t\right)^{(t-0,*)}\right] = E\left[\overline{U}_0 E\left[\left(\overline{U}_r\overline{U}_s\overline{U}_t\right)^{(t-0,*)} \mid \mathscr{F}_0\right]\right] = 0,$$

using Jensen's inequality and Hölder's inequality yields that

$$\begin{aligned}|E\left[\overline{U}_0\overline{U}_r\overline{U}_s\overline{U}_t\right]| &= \left|E\left[\overline{U}_0 E\left[\overline{U}_r\overline{U}_s\overline{U}_t \mid \mathscr{F}_0\right]\right] - E\left[\overline{U}_0 E\left[\left(\overline{U}_r\overline{U}_s\overline{U}_t\right)^{(t-0,*)} \mid \mathscr{F}_0\right]\right]\right| \\ &\le \|\overline{U}_0\|_4^3\left(\|\overline{U}_r - \overline{U}_r^{(r-0,*)}\|_4 + \|\overline{U}_s - \overline{U}_s^{(s-0,*)}\|_4 + \|\overline{U}_t - \overline{U}_t^{(t-0,*)}\|_4\right) \\ &= O(1)(\lambda_{r,4} + \lambda_{s,4} + \lambda_{t,4}).\end{aligned}$$

Similarly, we have

$$\left|E\left[\overline{U}_0\overline{U}_r\left(\overline{U}_s\overline{U}_t - (\overline{U}_s\overline{U}_t)^{(t-r,*)}\right)\right]\right| = O(1)(\lambda_{s-r,4} + \lambda_{t-r,4}),$$

and

$$\left|E\left[\overline{U}_0\overline{U}_r\overline{U}_s\left(\overline{U}_t - \overline{U}_t^{(t-s,*)}\right)\right]\right| = O(1)\lambda_{t-s,4}.$$

Putting the above results together, we have

$$\left|E\left[\overline{U}_0\overline{U}_r\overline{U}_s\overline{U}_t\right]\right| \le O(1)\min\left(\lambda_{r,4} + \lambda_{s,4} + \lambda_{t,4}, \lambda_{s-r,4} + \lambda_{t-r,4}, \lambda_{t-s,4}\right).$$

Next, define the following three sets:

$$\begin{aligned}\mathcal{S}_{ts} &= \{r, 0 : (t-s) \ge \max(s-r, r-0)\}, \\ \mathcal{S}_{sr} &= \{0, t : (s-r) \ge \max(t-s, r-0)\}, \\ \mathcal{S}_{r0} &= \{s, t : (r-0) \ge \max(s-r, t-s)\}.\end{aligned}$$

Note that the cardinalities (denoted by #) of these sets are bounded as follows:

$$\#\mathcal{S}_{ts} \leq O(1)(t-s), \quad \#\mathcal{S}_{sr} \leq O(1)(s-r), \quad \text{and} \quad \#\mathcal{S}_{r0} \leq O(1)(r-0)^2,$$

which further yields that

$$\begin{aligned}
&\sum_{0<r<s<t<\infty} |v(r,s,t) - \sigma(r)\sigma(s-t)| \\
&= \sum_{0<r<s<t<\infty} \left| E(\overline{U}_0\overline{U}_r\overline{U}_s\overline{U}_t - \overline{U}_0\overline{U}_r(\overline{U}_s\overline{U}_t)^{(t-r,*)}) \right| \\
&\leq O(1)\left(\sum_{r=1}^{\infty} r^2\lambda_{r,4} + \sum_{s=2}^{\infty}\sum_{r=1}^{s-1}(s-r)\lambda_{s-r,4} + \sum_{t=2}^{\infty}\sum_{s=1}^{t-1}(t-s)\lambda_{t-s,4} \right) = O(1).
\end{aligned}$$

In addition, we have

$$\sum_{0<r<s<t<\infty} |\sigma(s)\sigma(r-t)| \leq \sum_{s=1}^{\infty} |\sigma(s)| \left(\sum_{t=2}^{\infty}\sum_{r=1}^{t-1} |\sigma(r-t)| \right) \leq \sum_{s=1}^{\infty} |\sigma(s)| \sum_{j=1}^{\infty} j|\sigma(j)| = O(1).$$

Similarly, we have $\sum_{0<r<s<t<\infty} |\sigma(t)\sigma(r-s)| = O(1)$. Thus, we can obtain

$$I_4 \leq \sum_{0<r<s<t<\infty} (|v(r,s,t) - \sigma(r)\sigma(s-t) + |\sigma(s)\sigma(r-t)| + |\sigma(t)\sigma(r-s)|) = O(1).$$

Similarly, we have $|I_j| < \infty$ for $j = 5, 6, 7$. Collecting the above results, the proof of the second result is now completed. ∎

Proof of Theorem 2.1:

Our goal is to show that

$$S_{\mathbb{N}}^* \to_{D^*} N(0, \sigma_u^2), \tag{A.3}$$

which in connection with Proposition 2.2 immediately yields the result. In order to do so, we rewrite $S_{\mathbb{N}}^*$ as follows:

$$S_{\mathbb{N}}^* = \sum_{j=1}^{K} \nu_j^* + \sum_{j=1}^{K} \varpi_j^*, \tag{A.4}$$

where $\nu_j^* = \sum_{t=B_j+1}^{B_j+r_1} \frac{U_t^\top 1_N \xi_t}{\sqrt{\mathbb{N}}}$, $\varpi_j^* = \sum_{t=B_j+r_1+1}^{B_j+r_1+r_2} \frac{U_t^\top 1_N \xi_t}{\sqrt{\mathbb{N}}}$, and $B_j = (j-1)(r_1+r_2)$. Without loss of generality, suppose that $K = T/(r_1+r_2)$ is an integer for simplicity. Otherwise, one needs to include the remaining terms in (A.4) which are negligible for an obvious reason. In addition, we let

$$(r_1, r_2) \to (\infty, \infty), \quad \left(\frac{r_2}{r_1}, \frac{r_1}{T}\right) \to (0, 0), \quad r_1 \geq \ell, \tag{A.5}$$

so the blocks ϖ_j^*'s are mutually independent by the construction of ξ_t's. Note that by $\frac{r_2}{r_1} \to 0$ of (A.5), it is easy to know that $\frac{Kr_2}{T} \to 0$ and $\frac{Kr_1}{T} \to 1$.

We now write

$$EE^*\left[\left(\sum_{j=1}^K \varpi_j^*\right)^2\right] = \sum_{j=1}^K EE^*[(\varpi_j^*)^2]$$

$$\leq \frac{1}{\mathbb{N}}\sum_{j=1}^K \sum_{s=-r_2+1}^{r_2-1} \sum_{t=B_j+r_1+1}^{B_j+r_1+r_2-|s|} a\left(\frac{s}{\ell}\right)|E[U_t^\top 1_N U_{t+s}^\top 1_N]|$$

$$\leq O(1)\frac{1}{T}\sum_{j=1}^K \sum_{t=B_j+r_1+1}^{B_j+r_1+r_2} \frac{1}{N}\sum_{s=-r_2+1}^{r_2-1} |E[U_t^\top 1_N U_{t+s}^\top 1_N]| \leq O(1)\frac{Kr_2}{T} = o(1),$$

where the second inequality follows from $a(\cdot)$ being bounded on $[-1,1]$, and the third inequality follows from $\frac{1}{N}\sum_{s=-r_2+1}^{r_2-1}|E[U_t^\top 1_N U_{t+s}^\top 1_N]| = O(1)$ by Proposition 2.1. Therefore, the term $\sum_{j=1}^K \varpi_j^*$ of (A.4) is negligible.

Next, we employ Lindeberg CLT to establish the asymptotic normality of $\sum_{j=1}^K \nu_j^*$. Note that by the first result of Theorem 2.2.1, we know that $E^*[(S_{\mathbb{N}}^*)^2] \to_P \sigma_u^2$. As we have shown that $\sum_{j=1}^K \varpi_j^*$ of (A.4) is negligible, we conclude that $E^*[\sum_{j=1}^K \nu_j^*]^2 \to_P \sigma_u^2$. That said, we need only to verify that for $\forall \epsilon > 0$

$$\sum_{j=1}^K E^*\left[(\nu_j^*)^2 \cdot I\left(|\nu_j^*| > \epsilon\right)\right] = o_P(1), \tag{A.6}$$

which follows from

$$\sum_{j=1}^K E|E^*\left[(\nu_j^*)^2 \cdot I\left(|\nu_j^*| > \epsilon\right)\right]| = \sum_{j=1}^K E\left[(\nu_j^*)^2 \cdot I\left(|\nu_j^*| > \epsilon\right)\right] = o(1).$$

Thus, we write

$$\sum_{j=1}^K E[(\nu_j^*)^2 \cdot I(|\nu_j^*| > \epsilon)] \leq \epsilon^{-2}\sum_{j=1}^K E|\nu_j^*|^4 = \epsilon^{-2}\frac{1}{T^2}\sum_{j=1}^K E\left(\sum_{t=B_j+1}^{B_j+r_1} \overline{U}_t\xi_t\right)^4$$

$$= O(1)\frac{Kr_1^2}{T^2} = O(1)\frac{r_1}{T} = o(1),$$

where the first inequality follows from Cauchy-Schwarz inequality and Chebyshev's inequality, the second equality follows from the fact $E(\sum_{t=B_j+1}^{B_j+r_1}\overline{U}_t\xi_t)^4 = O(r_1^2)$ by Proposition A.1.1. Thus, we can conclude the validity of (A.6).

Based on the above development, we are ready to conclude that (A.3) holds. The proof is now completed. ■

Proof of Corollary 2.2:

By design, it is easy to know that

$$\widetilde{\theta} - \widehat{\theta} = \widehat{\Sigma}_1^{-1} \cdot \frac{1}{\mathbb{N}}\sum_{t=1}^T X_t^\top M_{\widehat{\Gamma}}\widehat{U}_t\xi_t$$

$$= \widehat{\Sigma}_1^{-1} \cdot \frac{1}{\mathbb{N}} \sum_{t=1}^{T} X_t^\top M_{\widehat{\Gamma}} U_t \xi_t + \frac{1}{\mathbb{N}} \sum_{t=1}^{T} X_t^\top M_{\widehat{\Gamma}} (X_t \theta_0 + \Gamma_0 f_t - X_t \widehat{\theta} - \widehat{\Gamma} \widehat{f}_t) \xi_t. \quad \text{(A.7)}$$

Note that by bringing $\widehat{f}_t = \frac{1}{N} \widehat{\Gamma}^\top (Y_t - X_t \widehat{\theta})$ in $X_t \theta_0 + \Gamma_0 f_t - X_t \widehat{\theta} - \widehat{\Gamma} \widehat{f}_t$, we can obtain

$$X_t \theta_0 + \Gamma_0 f_t - X_t \widehat{\theta} - \widehat{\Gamma} \widehat{f}_t = M_{\widehat{\Gamma}} X_t (\theta_0 - \widehat{\theta}) + M_{\widehat{\Gamma}} \Gamma_0 f_t + P_{\widehat{\Gamma}} U_t. \quad \text{(A.8)}$$

Bringing (A.8) in (A.7), the term $P_{\widehat{\Gamma}} U_t$ disappear automatically as $M_{\widehat{\Gamma}} P_{\widehat{\Gamma}} U_t = 0$. Thus, we can write further

$$\widetilde{\theta} - \widehat{\theta} = \widehat{\Sigma}_1^{-1} \cdot \frac{1}{\mathbb{N}} \sum_{t=1}^{T} X_t^\top M_{\widehat{\Gamma}} (U_t + \Gamma_0 f_t) \xi_t + \widehat{\Sigma}_1^{-1} \cdot \frac{1}{\mathbb{N}} \sum_{t=1}^{T} X_t^\top M_{\widehat{\Gamma}} X_t (\theta_0 - \widehat{\theta}) \xi_t.$$

Moreover we know that

$$\frac{1}{\mathbb{N}} \sum_{t,s=1}^{T} (\theta_0 - \widehat{\theta})^\top X_t^\top M_{\widehat{\Gamma}} Z_t Z_s^\top M_{\widehat{\Gamma}} X_t (\theta_0 - \widehat{\theta}) E[\xi_t \xi_s] = o_P(1),$$

where the last step is obvious in view of the fact that $\|\theta_0 - \widehat{\theta}\| = O_P\left(\frac{1}{\sqrt{\mathbb{N}}}\right)$ by the development almost identical to Theorem 3 of Bai (2009), and Assumption 2.

Therefore, in order to establish the asymptotic distribution of $\sqrt{\mathbb{N}}(\widetilde{\theta} - \widehat{\theta})$, we need only to study

$$\widehat{\Sigma}_1^{-1} \cdot \frac{1}{\mathbb{N}} \sum_{t=1}^{T} X_t^\top M_{\widehat{\Gamma}} (U_t + \Gamma_0 f_t) \xi_t.$$

In view of the proof of Theorem 3 of Bai (2009), one can decompose $M_{\widehat{\Gamma}} (U_t + \Gamma_0 f_t) \xi_t$, so the two bias terms will arise. However, both terms include $\{\xi_t\}$ that is independent of all the other variables and has mean 0, as a consequence these two terms will vanish asymptotically under Assumption 2. Therefore, we have

$$\widehat{\Sigma}_1^{-1} \cdot \frac{1}{\mathbb{N}} \sum_{t=1}^{T} X_t^\top M_{\widehat{\Gamma}} (U_t + \Gamma_0 f_t) \xi_t \to_{D^*} N(0, \Sigma_1^{-1} \Sigma_2 \Sigma_1^{-1}).$$

Further using the conditions that $\|\widehat{\mu}_B - B\| \to_P 0$ and $\|\widehat{\mu}_C - C\| \to_P 0$, the proof is now completed. ■

Online Supplementary Appendix B to "A Simple Bootstrap Method for Panel Data Inferences"

Jiti Gao, Bin Peng and Yayi Yan

Monash University, Australia

In this appendix, we first present the omitted proofs of Theorem 2.3, Corollary 2.1, and Corollary 2.3 in Appendix B.1. We then introduce a few definitions in Appendix B.2 to facilitate development of the preliminary lemmas. Finally, we summarize the preliminary lemmas in Appendix B.3 and give their proofs in Appendix B.4.

B.1 Omitted Proofs of the Main Results

Proof of Theorem 2.3:

In fact, our goal can be simplified as follows.

$$L_{\mathbb{N}}^* \to_{D^*} N(0, \Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}), \tag{B.1}$$

which in connection with (2.15) and (2.16) yields (2.19).

By Assumption 4, we have

$$L_{\mathbb{N}}^* = \widehat{\Sigma}_1^{-1}\frac{1}{\sqrt{\mathbb{N}}}\sum_{t=1}^{T} Z_t^\top \widehat{U}_t\xi_t = \widehat{\Sigma}_1^{-1}\frac{1}{\sqrt{\mathbb{N}}}\sum_{t=1}^{T} Z_t^\top U_t\xi_t + o_P(1).$$

Using a procedure similar to that of Theorem 2.1, it is easy to show (B.1). The proof is now complete. ■

Proof of Corollary 2.1:

The proof of Corollary 2.1 is straightforward in view of the fact that $\widehat{\theta} - \theta_0 = \left(\frac{1}{\sqrt{\mathbb{N}}}\right)$. Thus, the details are omitted. ■

Proof of Corollary 2.3:

Still, we decompose $S_{\mathbb{N}}^*$ as follows.

$$S_{\mathbb{N}}^* = \sum_{j=1}^{K}\nu_j^* + \sum_{j=1}^{K}\varpi_j^*,$$

where $\nu_j^* = \sum_{t=B_j+1}^{B_j+r_1}\frac{U_t^\top 1_{N_t}\xi_t}{\sqrt{\mathbb{N}}}$, $\varpi_j^* = \sum_{t=B_j+r_1+1}^{B_j+r_1+r_2}\frac{U_t^\top 1_{N_t}\xi_t}{\sqrt{\mathbb{N}}}$, and $B_j = (j-1)(r_1+r_2)$. Without loss of generality we suppose that $K = T/(r_1+r_2)$ is an integer for simplicity In addition, we let

$$(r_1, r_2) \to (\infty, \infty), \quad \left(\frac{r_2}{r_1}, \frac{r_1}{T}\right) \to (0, 0), \quad r_1 \geq \ell,$$

so the blocks ϖ_j^*'s are mutually independent by the construction of ξ_t's. Note that

$$EE^*\left[\left(\sum_{j=1}^{K}\varpi_j^*\right)^2\right]=\sum_{j=1}^{K}EE^*[(\varpi_j^*)^2]\leq\frac{1}{\mathbb{N}}\sum_{j=1}^{K}\sum_{s=-r_2+1}^{r_2-1}\sum_{t=B_j+r_1+1}^{B_j+r_1+r_2-|s|}a\left(\frac{s}{\ell}\right)|E[U_t^\top 1_{N_t}U_{t+s}^\top 1_{N_{t+s}}]|$$

$$\leq O(1)\frac{\overline{N}}{\mathbb{N}}\sum_{j=1}^{K}\sum_{t=B_j+r_1+1}^{B_j+r_1+r_2}\frac{1}{\overline{N}}\sum_{s=-r_2+1}^{r_2-1}|E[U_t^\top 1_{N_t}U_{t+s}^\top 1_{N_{t+s}}]|\leq O(1)\frac{Kr_2}{T}=o(1),$$

where the second inequality follows from $a(\cdot)$ being bounded on $[-1,1]$, and the third inequality follows from $\frac{1}{\overline{N}}\sum_{s=-r_2+1}^{r_2-1}|E[U_t^\top 1_{N_t}U_{t+s}^\top 1_{N_{t+s}}]|=O(1)$ by $\frac{\overline{N}T}{\mathbb{N}}\to c\in(0,\infty)$ of Assumption 5. Therefore, the term $\sum_{j=1}^{K}\varpi_j^*$ is negligible.

We can then establish the asymptotic normality of $\sum_{j=1}^{K}\nu_j^*$ in a way almost identical to those presented in the proof of Theorem 2.1. The proof is now completed. ∎

B.2 Notation and Definitions

We now prepare a few notations before proceeding further, which will be repeatedly used in the following development. First, we would like to emphasize that from this point and onwards, we let

$$U_t\equiv U_{t\gamma}=g_\gamma(\varepsilon_t,\ldots,\varepsilon_{t-\gamma}) \tag{B.2}$$

for notational simplicity, in which $\gamma\equiv\gamma_T\to\infty$ and $\gamma/T\to 0$. $\overline{U}_t$ is then defined accordingly. Also, without loss of generality, let $T\equiv 2n\gamma$ denote the integer part, in which n stands for the number of blocks. Otherwise, we have to take into account the remaining terms, which are negligible for an obvious reason.

Denote the following σ-field:

$$\mathscr{F}_\gamma=\sigma(\underbrace{\varepsilon_{-\gamma+1},\ldots,\varepsilon_0}_{1^{st}\text{ block}},\underbrace{\varepsilon_{\gamma+1},\ldots\varepsilon_{2\gamma}}_{2^{nd}\text{ block}},\ldots,\underbrace{\varepsilon_{(2n-1)\gamma+1},\ldots,\varepsilon_{2n\gamma}}_{(n+1)^{th}\text{ block}}), \tag{B.3}$$

and let $E_{\mathscr{F}_\gamma}[\cdot]=E[\cdot\mid\mathscr{F}_\gamma]$ and $\Pr_{\mathscr{F}_\gamma}(\cdot)=\Pr(\cdot\mid\mathscr{F}_\gamma)$ respectively be the conditional expectation and the conditional probability induced by $\mathscr{F}_\gamma$. Also, with respect to $\mathscr{F}_t$ of Section A.2, we define

$$\mathscr{F}_t^*=\sigma(\varepsilon_t,\ldots,\varepsilon_1,\varepsilon_0',\varepsilon_{-1}',\ldots).$$

Decomposition of $S_{\mathbb{N}}$:

We are now ready to decompose $S_{\mathbb{N}}$. First let

$$S_{\mathbb{N}}=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(\overline{S}_{j|\mathbb{N}}+\widetilde{S}_{j|\mathbb{N}}),$$

where $\overline{S}_{j|\mathbb{N}}=\frac{1}{\sqrt{2\gamma}}\sum_{t=(2j-2)\gamma+1}^{(2j-1)\gamma}\overline{U}_t$ and $\widetilde{S}_{j|\mathbb{N}}=\frac{1}{\sqrt{2\gamma}}\sum_{t=(2j-1)\gamma+1}^{2j\gamma}\overline{U}_t$. By design, $\{\overline{S}_{j|\mathbb{N}}\}_{j=1}^n$ and $\{\widetilde{S}_{j|\mathbb{N}}\}_{j=1}^n$ are two sequences of independent variables, respectively.

Using (B.3), we can also decompose $S_{\mathbb{N}}$ into the following two parts:

$$S_{\mathbb{N}} = \overline{S}_{\mathbb{N}|\gamma} + \widetilde{S}_{\mathbb{N}|\gamma},$$

where $\overline{S}_{\mathbb{N}|\gamma} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\overline{U}_t - E[\overline{U}_t \mid \mathscr{F}_\gamma])$ and $\widetilde{S}_{\mathbb{N}|\gamma} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} E[\overline{U}_t \mid \mathscr{F}_\gamma]$.

Below, we further decompose $\overline{S}_{\mathbb{N}|\gamma}$ and $\widetilde{S}_{\mathbb{N}|\gamma}$. Then $\overline{S}_{\mathbb{N}|\gamma}$ can be written as follows:

$$\sqrt{T}\overline{S}_{\mathbb{N}|\gamma} = \sqrt{2\gamma} \sum_{j=1}^{n} \overline{V}_j, \quad \overline{V}_j = \frac{1}{\sqrt{2\gamma}} (\overline{V}_{j,\text{odd}} + \overline{V}_{j,\text{even}}),$$

where $\overline{V}_{j,\text{odd}} = \sum_{t=(2j-2)\gamma+1}^{(2j-1)\gamma} (\overline{U}_t - E[\overline{U}_t \mid \mathscr{F}_\gamma])$ and $\overline{V}_{j,\text{even}} = \sum_{t=(2j-1)\gamma+1}^{2j\gamma} (\overline{U}_t - E[\overline{U}_t \mid \mathscr{F}_\gamma])$. By construction, the blocks $\overline{V}_j$ with $j \in [n]$ are independent variables under the conditional probability measure $\Pr_{\mathscr{F}_\gamma}(\cdot)$.

Let $\widetilde{V}_0 = \frac{1}{\sqrt{2\gamma}} \sum_{t=1}^{\gamma} E[\overline{U}_t \mid \mathscr{F}_\gamma]$, $\widetilde{V}_n = \frac{1}{\sqrt{2\gamma}} \sum_{t=(2n-1)\gamma+1}^{2n\gamma} E[\overline{U}_t \mid \mathscr{F}_\gamma]$, and

$$\widetilde{V}_j = \frac{1}{\sqrt{2\gamma}} \sum_{t=(2j-1)\gamma+1}^{(2j+1)\gamma} E[\overline{U}_t \mid \mathscr{F}_\gamma]$$

for $j \in [n-1]$. Note that $\{\widetilde{V}_j\}_{j=0}^{n}$ is a sequence of independent variables under the probability measure $\Pr(\cdot)$. We then decompose $\widetilde{S}_{\mathbb{N}|\gamma}$ as follows:

$$\widetilde{S}_{\mathbb{N}|\gamma} = \frac{1}{\sqrt{n}} \sum_{j=0}^{n} \widetilde{V}_j.$$

Accordingly, we define the following coupled variables for $j \in [n]$:

$$V_{j,\text{odd}}^* = \frac{1}{\sqrt{2\gamma}} \sum_{t=(2j-2)\gamma+1}^{(2j-1)\gamma} \overline{U}_t^{(t-(2j-2)\gamma,*)} \quad \text{and} \quad V_{j,\text{even}}^* = \frac{1}{\sqrt{2\gamma}} \sum_{t=(2j-1)\gamma+1}^{2j\gamma} \overline{U}_t^{(t-(2j-1)\gamma,*)},$$

where by construction $V_{j,\text{odd}}^*$ is independent of $\mathscr{F}_\gamma$, and $V_{j,\text{even}}^*$ is independent of $V_{j,\text{odd}}^*$.

Moments:

We define the following second moments:

$$\overline{\sigma}_{j|\gamma}^2 = E_{\mathscr{F}_\gamma} |\overline{V}_j|^2, \quad \overline{\sigma}_j^2 = E|\overline{V}_j|^2, \quad \widetilde{\sigma}_j^2 = E|\widetilde{V}_j|^2.$$

Similarly, we let $\overline{\kappa}_{j|\gamma}^3$, $\overline{\kappa}_j^3$, and $\widetilde{\kappa}_j^3$ be the corresponding quantities for the third moments, and let $\overline{\tau}_{j|\gamma}^4$, $\overline{\tau}_j^4$, and $\widetilde{\tau}_j^4$ be the fourth moments.

By construction, we have $E[\overline{S}_{j|\mathbb{N}}^2] = E[\widetilde{S}_{j|\mathbb{N}}^2] := \widehat{\sigma}_j^2$. We let further $\sigma_{|\gamma}^2 = E_{\mathscr{F}_\gamma}[\overline{S}_{\mathbb{N}|\gamma}^2]$, and $\sigma^2 = E[\sigma_{|\gamma}^2]$. Simple algebra shows that

$$\sigma_{|\gamma}^2 = \frac{2\gamma}{T} \sum_{j=1}^{n} \overline{\sigma}_{j|\gamma}^2 \quad \text{and} \quad \sigma^2 = \frac{2\gamma}{T} \sum_{j=1}^{n} \overline{\sigma}_j^2.$$

Conditional approximations and distributions:

Let $F(\cdot)$ be any continuous distribution function and $\{\overline{Z}_j\}_{j\in[n]}$ be i.i.d. random variables and distributed according to $F(\cdot)$ such that

$$E(\overline{Z}_j)=0,\quad E(\overline{Z}_j^2)-\overline{\sigma}_j^2=O(\gamma^{-1}),\quad E(\overline{Z}_j^3)-\overline{\kappa}_j^3=O(\gamma^{-1}),$$
$$E(\overline{Z}_j^4)-\overline{\tau}_j^4=O(\gamma^{-1}).$$

The existence of $\overline{Z}_j$ is guaranteed by Lemmas B.4 and B.6 below. Let $\{\widetilde{Z}_j\}_{j\in[n]}$ be independent copy of $\{\overline{Z}_j\}_{j\in[n]}$.

Define

$$\Delta_{j,\gamma}(w)=\Pr(V_{j,\mathrm{odd}}^*>w)-\Pr(\overline{Z}_j>w),$$
$$\mathscr{U}_T=\int_{-c\sqrt{T}}^{c\sqrt{T}}|E(e^{ixS_{\mathbb{N}}})-E(e^{ix(\overline{Z}+\widetilde{Z})})|\frac{1}{|x|}\mathrm{d}x,$$

where c is a sufficiently large constant, $\overline{Z}=n^{-1/2}\sum_{j=1}^n\overline{Z}_j$, and $\widetilde{Z}=n^{-1/2}\sum_{j=1}^n\widetilde{Z}_j$.

Recursion step:

For $a>0$ and $b\in\mathbb{N}$ even, let $G_{a,b}$ be a real-valued random variable with density function

$$g_{a,b}(x)=c_b a\left|\frac{\sin ax}{ax}\right|^b$$

for some $c_b>0$ only depending on b. It is well-known that for even b the Fourier transform $\widehat{g}_{a,b}$ satisfies $\widehat{g}_{a,b}(t)=2\pi c_b u^{*b}[-a,a](t)$ if $|t|\le ab$ and $\widehat{g}_{a,b}(t)=0$ otherwise, where $u^{*b}[-a,a]$ denotes the b-fold convolution of the density of the uniform distribution on $[-a,a]$, that is $u[-a,a](t)=\frac{1}{2a}I(-a\le t\le a)$. For $b\ge 6$, let $\{H_t\}_{t=1}^T$ be a sequence of i.i.d. random variables with $H_t=_D G_{a,b}$ and independent of $\{\overline{U}_t\}$. Define

$$X_t^\diamond=\overline{U}_t+H_t-H_{t-1},$$
$$S_{\mathbb{N}}^\diamond=\frac{1}{\sqrt{T}}\sum_{t=1}^T X_t^\diamond=\frac{1}{\sqrt{T}}\sum_{t=1}^T\overline{U}_t+\frac{1}{\sqrt{T}}H_T-\frac{1}{\sqrt{T}}H_0,$$
$$\overline{S}_{j|\mathbb{N}}^\diamond=\frac{1}{\sqrt{2\gamma}}\sum_{t=(2j-2)\gamma+1}^{(2j-1)\gamma}\overline{U}_t+\frac{1}{\sqrt{2\gamma}}H_{(2j-1)\gamma}-\frac{1}{\sqrt{2\gamma}}H_{(2j-2)\gamma},$$
$$\widetilde{S}_{j|\mathbb{N}}^\diamond=\frac{1}{\sqrt{2\gamma}}\sum_{t=(2j-1)\gamma+1}^{2j\gamma}\overline{U}_t+\frac{1}{\sqrt{2\gamma}}H_{2j\gamma}-\frac{1}{\sqrt{2\gamma}}H_{(2j-1)\gamma}.$$

Accordingly, we have $s_{\mathbb{N}}^\diamond$, $\kappa_{\mathbb{N}}^\diamond$, $\left(\mathscr{T}_a^b\right)^\diamond$, $\mathscr{U}_T^\diamond$, $\Psi_{\mathbb{N}}^\diamond$, $\overline{Z}_j^\diamond$, $\widetilde{Z}_j^\diamond$, $\overline{Z}^\diamond$, $\widetilde{Z}^\diamond$ and the difference

$$\Delta_T^\diamond(w)=\Pr\left(S_{\mathbb{N}}^\diamond\le w\right)-\Psi_{\mathbb{N}}^\diamond(w).$$

By the formulae of $g_{a,b}$ and $\widehat{g}_{a,b}$, we have $E(H_t)=0$ and $E|H_t|^4<\infty$. Moreover, by independence we have

$$E(e^{ixS_{\mathbb{N}}^{\diamond}}) = E(e^{ixS_{\mathbb{N}}}) \cdot E(e^{ix\frac{H_T}{\sqrt{T}}}) \cdot E(e^{-ix\frac{H_0}{\sqrt{T}}}),$$

which, in connection with the definitions of $g_{a,b}(x)$ and $\widehat{g}_{a,b}(x)$, yields

$$E(e^{ixS_{\mathbb{N}}^{\diamond}}) = 0 \text{ for } |x| > \sqrt{T}|ab|.$$

For $j \in [n]$, define

$$\begin{aligned}
\sqrt{2\gamma}V_{j,\text{odd}}^{\diamond} &= \sqrt{2\gamma}V_{j,\text{odd}}^{*} + H_{(2j-1)\gamma+1} - H_{(2j-2)\gamma+1}, \\
\sqrt{2\gamma}V_{j,\text{even}}^{\diamond} &= \sqrt{2\gamma}V_{j,\text{even}}^{*} + H_{2j\gamma+1} - H_{(2j-1)\gamma+1}, \\
\Delta_{j,\gamma}^{\diamond}(w) &= \Pr(V_{j,\text{odd}}^{\diamond} > w) - \Pr(\overline{Z}_j^{\diamond} > w).
\end{aligned}$$

By definition, we have $\overline{S}_{j|\mathbb{N}}^{\diamond} =_D V_{j,\text{odd}}^{\diamond} =_D \widetilde{S}_{j|\mathbb{N}}^{\diamond} =_D V_{j,\text{even}}^{\diamond}$.

We now provide the following preliminary lemmas that have already been used in the proofs of Appendix A.

B.3 Preliminary Lemmas

In order to prove the main results in Lemma B.8, we will need to introduce Lemmas B.1–B.7 and then given their proofs in this appendix.

Lemma B.1. *For $a_1 > 0$, $a_2, w \in \mathbb{R}$, let $G(a_1, a_2, w) = \Psi\left(\frac{w}{a_1}\right) + \frac{1}{6}a_2\left(1 - \frac{w^2}{a_1^2}\right)\psi\left(\frac{w}{a_1}\right)$. Suppose that $\max_{1\le i\le 2}|a_i - b_i| \le y$. Then $\sup_{w\in\mathbb{R}}(w^2+1)|G(a_1,a_2,w) - G(b_1,b_2,w)| \le O(1)y$.*

Lemma B.2. *Under Assumption 1, the following results hold:*

1. $\|\overline{V}_{j,\text{even}}\|_\delta = O(1)$ *for each* $j \in [n]$;

2. $\|\overline{V}_j - V_{j,\text{odd}}^{*}\|_\delta = O(\gamma^{-1/2})$ *for each* $j \in [n]$;

3. $E(S_{\mathbb{N}}^3) = O\left(\frac{1}{\sqrt{T}}\right)$;

4. $E(S_{\mathbb{N}}^4) - 3[E(S_{\mathbb{N}}^2)]^2 = O\left(\frac{1}{T}\right)$.

Lemma B.3. *Let Assumption 1 hold. Suppose that $f(\cdot)$ is a three times differentiable function with $|f^{(s)}| \le M_f < \infty$ for $s = 0, 1, \ldots, 3$. Then the following results hold:*

1. $\|E_{\mathscr{F}_\gamma}(f(\overline{V}_j) - f(V_{j,\text{odd}}^{*})\|_1 = O(M_f\gamma^{-1})$ *for each* $j \in [n]$;

2. $\|f(\widetilde{V}_j) - f(V_{j,\text{even}}^{*})\|_1 = O(M_f\gamma^{-1})$ *for each* $j \in [n]$.

Let further $f(\cdot)$ be a fourth-degree polynomial with coefficients bounded by M_f. Then the following results hold:

3. $\|E_{\mathscr{F}_\gamma}(f(\overline{V}_j) - f(V_{j,\text{odd}}^{*}))\|_1 = O(M_f\gamma^{-1})$ *for each* $j \in [n]$;

4. $\|f(\widetilde{V}_j) - f(V^*_{j,\text{even}})\|_1 = O(M_f\gamma^{-1})$ *for each* $j \in [n]$.

Lemma B.4. *Under Assumption 1, the following results of the sample moments hold:*

1. $\|\overline{\sigma}^2_{j|\gamma} - \overline{\sigma}^2_j\|_{\delta/2} = \|\overline{\sigma}^2_{j|\gamma} - \widehat{\sigma}^2_j\|_{\delta/2} + O(\gamma^{-1}) = O(\gamma^{-1})$ *for each* $j \in [n]$*;*
2. $\|\overline{\sigma}^2_j - \widehat{\sigma}^2_j\|_{\delta/2} = O(\gamma^{-1})$ *for each* $j \in [n]$*;*
3. $\|\sigma^2_{|\gamma} - \sigma^2\|_{\delta/2} = O(T^{-1}n^{1/2})$*;*
4. $\|\overline{\kappa}^3_{j|\gamma} - E(\overline{S}^3_{j|\mathbb{N}})\|_1 = O(\gamma^{-1})$ *and* $\|\overline{\kappa}^3_{j|\gamma} - \overline{\kappa}^3_j\|_1 = O(\gamma^{-1})$ *for each* $j \in [n]$*;*
5. $E(S^3_{\mathbb{N}}) - E(\overline{S}^3_{\mathbb{N}|\gamma}) - E(\widetilde{S}^3_{\mathbb{N}|\gamma}) = O(T^{-1}n^{1/2})$ *and* $E(\overline{S}^3_{\mathbb{N}|\gamma}) + E(\widetilde{S}^3_{\mathbb{N}|\gamma}) - \left(\frac{2\gamma}{T}\right)^{3/2}\sum_{j=1}^n(\overline{\kappa}^3_j + \widetilde{\kappa}^3_j) = O(T^{-1}n^{1/2})$*;*
6. $\|\overline{\tau}^4_{j|\gamma} - E(\overline{S}^4_{j|\mathbb{N}})\|_1 = O(\gamma^{-1})$ *and* $\|\overline{\tau}^4_{j|\gamma} - \overline{\tau}^4_j\|_1 = O(\gamma^{-1})$ *for each* $j \in [n]$*;*
7. $\widetilde{\sigma}^2_j - \overline{\sigma}^2_j = O(\gamma^{-1})$ *for each* $j \in [n]$*;*
8. $\widetilde{\kappa}^3_j - \overline{\kappa}^3_j = O(\gamma^{-1})$ *for each* $j \in [n]$*;*
9. $\widetilde{\tau}^4_j - \overline{\tau}^4_j = O(\gamma^{-1})$ *for each* $j \in [n]$.

Lemma B.5. *Under Assumption 1, for all* $x > M\sqrt{T\log T}$ *with* M *being a large constant, we have* $\Pr(\sqrt{T}S_{\mathbb{N}} \geq x) = O(Tx^{-4})$.

Lemma B.6. *Let Assumption 1 hold. Then* $F(\cdot)$ *exists and can be chosen such that*

1. $\sup_{w\in\mathbb{R}}|F(w) - \Psi_{\gamma,j}(w)| = \sup_{w\in\mathbb{R}}|\Pr(\overline{Z}_j \leq w) - \Psi_{\gamma,j}(w)| = O(\gamma^{-1})$, *where*
$$\Psi_{\gamma,j}(w) := \Psi\left(\frac{w}{\widehat{\sigma}_j}\right) + \frac{1}{6}E(\overline{S}^3_{j|\mathbb{N}})\left(1 - \frac{w^2}{\widehat{\sigma}^2_j}\right)\psi\left(\frac{w}{\widehat{\sigma}_j}\right);$$
2. $\sup_{w\in\mathbb{R}}(w^2+1)|\Pr(\overline{Z} + \widetilde{Z} \leq w) - \Psi_{\mathbb{N}}(w)| = O(\gamma^{-1})$, *where*
$$\Psi_{\mathbb{N}}(w) := \Psi\left(\frac{w}{s_{\mathbb{N}}}\right) + \frac{1}{6}\kappa^3_{\mathbb{N}}\left(1 - \frac{w^2}{s^2_{\mathbb{N}}}\right)\psi\left(\frac{w}{s_{\mathbb{N}}}\right);$$
3. $E\left(|\overline{Z}_j|^3 I(|\overline{Z}_j| \geq \tau_T)\right) = O(\gamma^{-2})$, *where* $\tau_T \geq c_\tau\sqrt{\log T}$, *and* $c_\tau > 0$ *is sufficiently large.*

Lemma B.7. *Let Assumption 1 hold. Let* f *be a smooth function such that* $\sup_{x\in\mathbb{R}}|f^{(s)}(x)| \leq 1$ *for* $s = 0, 1, \ldots, 8$, *and* $x_n = x/\sqrt{n}$. *Then for* $\tau_T \geq c_\tau\sqrt{\log T}$ *with* $c_\tau > 0$ *being sufficiently large, we have*

1. $\|E_{\mathscr{F}_\gamma}(f(x_n\overline{V}^\diamond_j)) - E(f(x_n\overline{Z}^\diamond_j))\|_1$
$= (|x_n|^2 + \tau^5_T|x_n|^5)O(\gamma^{-1}) + \sup_{x\in\mathbb{R}}|\Delta^\diamond_{j,\gamma}(x)|O(\tau^5_T|x_n|^5 + \tau^6_T|x_n|^6)$*;*

2. $\|E_{\mathscr{F}_\gamma}(f(x_n\overline{V}_j)) - E(f(x_n\overline{Z}_j))\|_1$
$= (|x_n|^2 + \tau_T^5|x_n|^5)O(\gamma^{-1}) + \sup_{x\in\mathbb{R}} |\Delta^\diamond_{j,\gamma}(x)|O(\tau_T^5|x_n|^5 + \tau_T^6|x_n|^6)$;

3. $\|E_{\mathscr{F}_\gamma}(f(x_n\widetilde{V}^\diamond_j)) - E(f(x_n\widetilde{Z}^\diamond_j))\|_1$
$= (|x_n|^2 + \tau_T^5|x_n|^5)O(\gamma^{-1}) + \sup_{x\in\mathbb{R}} |\Delta^\diamond_{j,\gamma}(x)|O(\tau_T^5|x_n|^5 + \tau_T^6|x_n|^6)$;

4. $\|E_{\mathscr{F}_\gamma}(f(x_n\widetilde{V}_j)) - E(f(x_n\widetilde{Z}_j))\|_1$
$= (|x_n|^2 + \tau_T^5|x_n|^5)O(\gamma^{-1}) + \sup_{x\in\mathbb{R}} |\Delta^\diamond_{j,\gamma}(x)|O(\tau_T^5|x_n|^+ \tau_T^6|x_n|^6)$.

Lemma B.8. *Under Assumption 1, for $n \asymp (\log T)^m$ with $m \geq 5$,*

1. $\sup_{w\in\mathbb{R}} |\Delta^\diamond_T(w)| = O(n/T)$;

2. $\mathscr{U}_T = O(T^{-1}(\log T)^5)$ *with* $\mathscr{U}_T = \int_{-c\sqrt{T}}^{c\sqrt{T}} |E(e^{ixS_{\mathbb{N}}}) - E(e^{ix(\overline{Z}+\widetilde{Z})})|\frac{1}{|x|}\mathrm{d}x$.

B.4 Proofs of Preliminary Lemmas

Proof of Lemma B.1:

This is Lemma 9.14 in Jirak et al. (2021), so the proof is omitted. ■

Proof of Lemma B.2:

(1). Without loss of generality, we assume $j = 1$ and write

$$\begin{aligned}\|\overline{V}_{1,\text{even}}\|_\delta &\leq \sum_{t=\gamma+1}^{2\gamma} \left\|\overline{U}_t - E[\overline{U}_t \mid \mathscr{F}_\gamma]\right\|_\delta = \sum_{t=\gamma+1}^{2\gamma} \left\|E[\overline{U}_t - \overline{U}_t^{(t-\gamma,*)} \mid \mathscr{F}_\gamma, \mathscr{F}_t^*]\right\|_\delta \\ &\leq \sum_{t=\gamma+1}^{2\gamma} \|\overline{U}_t - \overline{U}_t^{(t-\gamma,*)}\|_\delta = \sum_{t=1}^{\gamma} \|\overline{U}_t - \overline{U}_t^*\|_\delta = \sum_{t=1}^{\gamma} \lambda^U_{t,\delta} = O(1),\end{aligned}$$

where the first inequality follows from the triangle inequality, the equality follows from $E[\overline{U}_t \mid \mathscr{F}_\gamma] =_D E[\overline{U}_t^{(t-\gamma,*)} \mid \mathscr{F}_\gamma, \mathscr{F}_t^*]$, the second inequality follows from the Jensen's inequality, and the second equality follows from the definitions of $\overline{U}_t$ and $\overline{U}_t^{(m,*)}$.

(2). Focus on $j = 1$ without loss of generality. Write

$$\begin{aligned}\sqrt{2\gamma}\|\overline{V}_1 - V^*_{1,\text{odd}}\|_\delta \leq \sqrt{2\gamma}\|\overline{V}_1 - V^*_{1,\text{odd}}\|_\delta &= \|\overline{V}_{1,\text{odd}} - V^*_{1,\text{odd}} + \overline{V}_{1,\text{even}}\|_\delta \\ &\leq \sum_{t=1}^{\gamma} \|\overline{U}_t^{(t,*)} - \overline{U}_t\|_\delta + \sum_{t=1}^{\gamma} \|E[\overline{U}_t \mid \mathscr{F}_\gamma]\|_\delta + \|\overline{V}_{1,\text{even}}\|_\delta \\ &= \sum_{t=1}^{\gamma} \|\overline{U}_t^* - \overline{U}_t\|_\delta + \sum_{t=1}^{\gamma} \|E[\overline{U}_t \mid \mathscr{F}_\gamma]\|_\delta + O(1) \\ &= \sum_{t=1}^{\gamma} \|E[\overline{U}_t \mid \mathscr{F}_\gamma]\|_\delta + O(1) = O(1),\end{aligned}$$

where the second equality follows from the first result of this lemma, the third equality follows from the definitions of $\overline{U}_t$ and $\overline{U}_t^{(t,*)}$, and the fourth equality follows from

$$\sum_{t=1}^{\gamma}\|E[\overline{U}_t \mid \mathscr{F}_\gamma]\|_\delta = \sum_{t=1}^{\gamma}\|E[\overline{U}_t - \overline{U}_t^* \mid \mathscr{F}_\gamma]\|_\delta \leq \sum_{t=1}^{\gamma}\|\overline{U}_t - \overline{U}_t^*\|_\delta = \sum_{t=1}^{\gamma}\lambda_{t,\delta}^U = O(1).$$

(3). For $k \leq s \leq t$, we write

$$\begin{aligned}
|E(\overline{U}_k\overline{U}_s\overline{U}_t)| &= |E[\overline{U}_k \cdot E(\overline{U}_s\overline{U}_t \mid \mathscr{F}_k)]| = |E[\overline{U}_k \cdot E(\overline{U}_s\overline{U}_t - \overline{U}_s^{(s-k,*)}\overline{U}_t^{(t-k,*)} \mid \mathscr{F}_k)]| \\
&\leq \|\overline{U}_k\|_3 \cdot \left\{E[|E(\overline{U}_s\overline{U}_t - \overline{U}_s^{(s-k,*)}\overline{U}_t^{(t-k,*)} \mid \mathscr{F}_k)|^{\frac{3}{2}}]\right\}^{\frac{2}{3}} \\
&\leq \|\overline{U}_k\|_3 \cdot \|E(\overline{U}_s\overline{U}_t - \overline{U}_s^{(s-k,*)}\overline{U}_t^{(t-k,*)} \mid \mathscr{F}_k)\|_3 \\
&\leq \|\overline{U}_k\|_3 \cdot \|\overline{U}_s\overline{U}_t - \overline{U}_s^{(s-k,*)}\overline{U}_t^{(t-k,*)}\|_3 \\
&\leq \|\overline{U}_k\|_3 \cdot (\|\overline{U}_s\|_3\|\overline{U}_t - \overline{U}_t^{(t-k,*)}\|_3 + \|\overline{U}_s - \overline{U}_s^{(s-k,*)}\|_3\|\overline{U}_t^{(t-k,*)}\|_3) \\
&= \|\overline{U}_k\|_3 \cdot (\|\overline{U}_s\|_3\|\overline{U}_t - \overline{U}_t^{(t-k,*)}\|_3 + \|\overline{U}_s - \overline{U}_s^{(s-k,*)}\|_3\|\overline{U}_t\|_3) \\
&= \|\overline{U}_0\|_3^2 \cdot (\lambda_{t-k,3}^U + \lambda_{s-k,3}^U),
\end{aligned}$$

where the second equality follows from $E(\overline{U}_k) = 0$ and $E(\overline{U}_s\overline{U}_t) = E(\overline{U}_s^{(s-k,*)}\overline{U}_t^{(t-k,*)} \mid \mathscr{F}_k)$, the first inequality follows from the Hölder inequality, the second inequality follows from the moments monotonicity, the third inequality follows from the Jensen's inequality, and the third and fourth equalities follow from the definition of $\overline{U}_t$.

Similarly, we have $|E(\overline{U}_k\overline{U}_s\overline{U}_t)| \leq \|\overline{U}_0\|_3^2\lambda_{t-s,3}^U$. Thus, we can write

$$\begin{aligned}
E(S_{\mathbb{N}}^3) &\leq O(1)\frac{1}{T^{3/2}}\sum_{1\leq k\leq s\leq t\leq T}|E(\overline{U}_k\overline{U}_s\overline{U}_t)| \leq O(1)\frac{1}{T^{3/2}}\sum_{k=1}^{T}\sum_{s=k}^{k+\gamma}\sum_{t=s}^{s+\gamma}|E(\overline{U}_k\overline{U}_s\overline{U}_t)| \\
&\leq O(1)\frac{1}{T^{3/2}}\sum_{k=1}^{T}\sum_{s=k}^{k+\gamma}\left(\sum_{t=s}^{s+s-k-1}|E(\overline{U}_k\overline{U}_s\overline{U}_t)| + \sum_{t=2s-k}^{s+\gamma}|E(\overline{U}_k\overline{U}_s\overline{U}_t)|\right) \\
&\leq O(1)\frac{1}{T^{3/2}}\sum_{k=1}^{T}\sum_{s=k}^{k+\gamma}\left(\sum_{t=s}^{2s-k-1}\lambda_{s-k,3}^U + \sum_{t=2s-k}^{s+\gamma}\lambda_{t-s,3}^U\right) \\
&= O(1)\frac{1}{T^{1/2}}\sum_{j=1}^{\gamma}j\lambda_{j,3}^U = O\left(\frac{1}{\sqrt{T}}\right).
\end{aligned} \tag{B.4}$$

(4). Expanding $E(S_{\mathbb{N}}^4)$, we have

$$\begin{aligned}
E(S_{\mathbb{N}}^4) \;=\; & \frac{1}{T^2}\sum_{t=1}^{T}E(\overline{U}_t^4) + \frac{4}{T^2}\sum_{1\leq t<s\leq T}E(\overline{U}_t^3\overline{U}_s) + \frac{4}{T^2}\sum_{1\leq t<s\leq T}E(\overline{U}_t\overline{U}_s^3) \\
& + \frac{12}{T^2}\sum_{1\leq t<s<k\leq T}E(\overline{U}_t^2\overline{U}_s\overline{U}_k) + \frac{12}{T^2}\sum_{1\leq t<s<k\leq T}E(\overline{U}_t\overline{U}_s^2\overline{U}_k) \\
& + \frac{12}{T^2}\sum_{1\leq t<s<k\leq T}E(\overline{U}_t\overline{U}_s\overline{U}_k^2) + \frac{6}{T^2}\sum_{1\leq t<s\leq T}E(\overline{U}_t^2\overline{U}_s^2) \\
& + \frac{24}{T^2}\sum_{1\leq t<s<k<l\leq T}E(\overline{U}_t\overline{U}_s\overline{U}_k\overline{U}_l) \\
:=\; & I_{T,1} + I_{T,2} + I_{T,3} + I_{T,4} + I_{T,5} + I_{T,6} + I_{T,7} + I_{T,8}.
\end{aligned}$$

The most difficult term to be dealing with is $I_{T,8}$ that we handle it first. Using similar arguments to the proof of the third result of this lemma, we have for $1 \le t < s < k < l \le T$

$$\|\overline{U}_t\overline{U}_s\overline{U}_k\overline{U}_l\|_1 \le O(1)\min(\lambda_{s-t,4}^U, \lambda_{l-s,4}^U + \lambda_{k-s,4}^U, \lambda_{l-k,4}^U).$$

Then using a procedure similar to (B.4) we can obtain

$$E|I_{T,8}| \le O(1)\frac{1}{T^2}\sum_{1\le t<s<k<l\le T}|E(\overline{U}_t\overline{U}_s\overline{U}_k\overline{U}_l)| \le O(1)\frac{1}{T}\sum_{j=1}^{\gamma} j^2\lambda_{j,4}^U = O\left(\frac{1}{T}\right).$$

Similarly, we have $E|I_{T,j}| = O\left(\frac{1}{T}\right)$ for $j = 1, 2, 3$ and

$$I_{T,4} + I_{T,5} + I_{T,6} - 18T^{-2}\sum_{t,s,v=1}^{T} E(\overline{U}_t^2)E(\overline{U}_s\overline{U}_v) = O\left(\frac{1}{T}\right).$$

For $I_{T,7}$, we have

$$I_{T,7} = 3\left(\frac{1}{T}\sum_{t=1}^{T} E(\overline{U}_t^2)\right)^2 + O\left(\frac{1}{T}\right).$$

Putting the above results together and in view of the expansion of $3[E(S_{\mathbb{N}}^2)]^2$, the result follows immediately. The proof is now completed. ■

Proof of Lemma B.3:

(1). Without loss of generality, let $j = 1$ and write

$$\begin{aligned}&\|E_{\mathscr{F}_\gamma}(f(\overline{V}_1) - f(V_{1,\text{odd}}^*)) - E_{\mathscr{F}_\gamma}(f^{(1)}(V_{1,\text{odd}}^*)(\overline{V}_1 - V_{1,\text{odd}}^*))\|_1\\ &\le M_f\|\overline{V}_1 - V_{1,\text{odd}}^*\|_2^2 = O(M_f\gamma^{-1}),\end{aligned}$$

where the first inequality follows from Taylor expansion, and the equality follows from Lemma B.2.2. In what follows, we show

$$\|E_{\mathscr{F}_\gamma}(f^{(1)}(V_{1,\text{odd}}^*)(\overline{V}_1 - V_{1,\text{odd}}^*))\|_1 = O(M_f\gamma^{-1}),$$

and the result then follows. Note further that we can decompose $f^{(1)}(V_{1,\text{odd}}^*)(\overline{V}_1 - V_{1,\text{odd}}^*)$ as follows:

$$\begin{aligned}&f^{(1)}(V_{1,\text{odd}}^*)(\overline{V}_1 - V_{1,\text{odd}}^*)\\ &= f^{(1)}(V_{1,\text{odd}}^*)(2\gamma)^{-1/2}\overline{V}_{1,\text{even}} + f^{(1)}(V_{1,\text{odd}})((2\gamma)^{-1/2}\overline{V}_{1,\text{odd}} - V_{1,\text{odd}}^*).\end{aligned} \tag{B.5}$$

Thus, we need only to examine the two terms on the right hand side of (B.5).

First, consider $f^{(1)}(V_{1,\text{odd}}^*)\overline{V}_{1,\text{even}}$. Define for $0 < m < \gamma$

$$V_{1,\text{odd}}^{(>m,*)} = \frac{1}{\sqrt{2\gamma}}\sum_{t=\gamma-m}^{\gamma}\overline{U}_t^* \quad \text{and} \quad V_{1,\text{odd}}^{(\le m,*)} = \frac{1}{\sqrt{2\gamma}}\sum_{t=1}^{\gamma-m-1}\overline{U}_t^*.$$

Then write

$$\begin{aligned}
&\|(f^{(1)}(V_{1,\text{odd}}^{*}) - f^{(1)}(V_{1,\text{odd}}^{(\le m,*)}) - f^{(2)}(V_{1,\text{odd}}^{(\le m,*)})V_{1,\text{odd}}^{(>m,*)})\overline{V}_{1,\text{even}}\|_1 \\
&\le O(1)M_f\|\overline{V}_{1,\text{even}}\|_3\|V_{1,\text{odd}}^{(>m,*)}\|_3^2 = O(\gamma^{-1}m),
\end{aligned}$$

where the inequality follows from Taylor expansion and Hölder inequality, and the equality follows from $\|\overline{V}_{1,\text{even}}\|_3 = O(1)$ by Lemma B.2.1, and $\|V_{1,\text{odd}}^{(>m,*)}\|_3^2 = O(\gamma^{-1}m)$ by Proposition A.1.1.

In order to investigate $f^{(1)}(V_{1,\text{odd}}^{(\le m,*)})\overline{V}_{1,\text{even}}$, we define

$$V_{1,\text{even}}^{(m,**)} = \sum_{t=\gamma+1}^{2\gamma}(\overline{U}_t - E[\overline{U}_t \mid \mathscr{F}_\gamma])^{(t-\gamma+m,**)},$$

and note that for $\gamma < t \le 2\gamma$ and $0 < m < \gamma$, we have

$$(E[\overline{U}_t \mid \mathscr{F}_\gamma])^{(t-\gamma+m,**)} = E[\overline{U}_t \mid \mathscr{F}_\gamma] = E[\overline{U}_t^{(t-\gamma+m,**)} \mid \mathscr{F}_\gamma]. \tag{B.6}$$

Then write

$$\begin{aligned}
&\|\overline{V}_{1,\text{even}} - V_{1,\text{even}}^{(m,**)}\|_\delta \\
&= \left\|\sum_{t=\gamma+1}^{2\gamma}(\overline{U}_t - E[\overline{U}_t \mid \mathscr{F}_\gamma]) - \sum_{t=\gamma+1}^{2\gamma}(\overline{U}_t - E[\overline{U}_t \mid \mathscr{F}_\gamma])^{(t-\gamma+m,**)}\right\|_\delta \\
&\le \left\|\sum_{t=\gamma+1}^{2\gamma}(\overline{U}_t - \overline{U}_t^{(t-\gamma+m,**)})\right\|_\delta \le O(1)\sum_{t=\gamma+1}^{2\gamma}\|\overline{U}_t - \overline{U}_t^{(t-\gamma+m,**)}\|_\delta \\
&= O(1)\sum_{t=1}^{\gamma-m}\|\overline{U}_{\gamma+t} - \overline{U}_{t+\gamma}^{(t+m,**)}\|_\delta \le O(1)\sum_{t=m}^{\gamma}\frac{t^2}{m^2}\|\overline{U}_t - \overline{U}_t^{*}\|_\delta = O(m^{-2}),
\end{aligned}$$

where the first inequality follows from (B.6) and the triangle inequality, the third inequality follows from the definition of $\overline{U}_t$ and the fact that $\frac{t}{m} > 1$, and the third equality follows from Assumption 1. Thus, we immediately obtain that

$$\|f^{(1)}(V_{1,\text{odd}}^{(\le m,*)})(\overline{V}_{1,\text{even}} - V_{1,\text{even}}^{(m,**)})\|_1 \le O(1)M_f\|\overline{V}_{1,\text{even}} - V_{1,\text{even}}^{(m,**)}\|_1 = O(M_f m^{-2}). \tag{B.7}$$

Since $V_{1,\text{odd}}^{(\le m,*)}$ and $V_{1,\text{even}}^{(m,**)}$ are independent under the probability measure $\Pr_{\mathscr{F}_\gamma}$, we have

$$E_{\mathscr{F}_\gamma}[f^{(1)}(V_{1,\text{odd}}^{(\le m,*)}) \cdot V_{1,\text{even}}^{(m,**)}] = 0. \tag{B.8}$$

Hence, by (B.7) and (B.8)

$$\|E_{\mathscr{F}_\gamma}[f^{(1)}(V_{1,\text{odd}}^{(\le m,*)}) \cdot \overline{V}_{1,\text{even}}]\|_1 = O(M_f m^{-2}).$$

Since $V_{1,\text{odd}}^{(\le m,*)}$ and $V_{1,\text{odd}}^{(>m,**)}V_{1,\text{even}}^{(m,**)}$ are independent under the probability measure $\Pr_{\mathscr{F}_\gamma}$, we obtain that

$$
\begin{aligned}
&\sqrt{2\gamma}\|E_{\mathscr{F}_\gamma}[f^{(2)}(V_{1,\text{odd}}^{(\le m,*)})V_{1,\text{odd}}^{(>m,**)}V_{1,\text{even}}^{(m,**)}]\|_1 \\
&\le O(1)M_f\sqrt{2\gamma}\|E_{\mathscr{F}_\gamma}[V_{1,\text{odd}}^{(>m,**)}V_{1,\text{even}}^{(m,**)}]\|_1 \\
&\le O(1)M_f\left\|\sum_{t=\gamma-m}^{\gamma}\sum_{k=\gamma+1}^{2\gamma}E_{\mathscr{F}_\gamma}[\overline{U}_t^{(t-\gamma+m,**)}\overline{U}_k^{(k-\gamma+m,**)}]\right\|_1 .
\end{aligned}
$$

For $\gamma-m\le t\le\gamma$, $\overline{U}_t^{(t-\gamma+m,**)}$ is independent of $\mathscr{F}_\gamma$, so we have

$$
E_{\mathscr{F}_\gamma}[\overline{U}_t^{(t-\gamma+m,**)}]=E(\overline{U}_t)=0. \tag{B.9}
$$

Also, by the conditional independence of $\overline{U}_t^{(t-\gamma+m,**)}$ and $\overline{U}_k^{(k-t,**)}$ for $\gamma-m\le t\le\gamma$, $\gamma+1\le k\le 2\gamma$, we have

$$
\begin{aligned}
&\sum_{t=\gamma-m}^{\gamma}\sum_{k=\gamma+1}^{2\gamma}E_{\mathscr{F}_\gamma}[\overline{U}_t^{(t-\gamma+m,**)}\overline{U}_k^{(k-t,**)}] \\
&=\sum_{t=\gamma-m}^{\gamma}\sum_{k=\gamma+1}^{2\gamma}E_{\mathscr{F}_\gamma}[\overline{U}_t^{(t-\gamma+m,**)}]E_{\mathscr{F}_\gamma}[\overline{U}_k^{(k-t,**)}]=0,
\end{aligned}
\tag{B.10}
$$

where the last equality follows from (B.9). Then we write

$$
\begin{aligned}
&\sqrt{2\gamma}\|E_{\mathscr{F}_\gamma}[V_{1,\text{odd}}^{(>m,**)}V_{1,\text{even}}^{(m,**)}]\|_1 \\
&=\left\|\sum_{t=\gamma-m}^{\gamma}\sum_{k=\gamma+1}^{2\gamma}E_{\mathscr{F}_\gamma}[\overline{U}_t^{(t-\gamma+m,**)}]\cdot(\overline{U}_k^{(k-\gamma+m,**)}-\overline{U}_k^{(k-t,**)})\right\|_1 \\
&\le\sum_{t=\gamma-m}^{\gamma}\sum_{k=\gamma+1}^{2\gamma}\|\overline{U}_t\|_2\|\overline{U}_k^{(k-\gamma+m,**)}-\overline{U}_k^{(k-t,**)}\|_2 \\
&=\sum_{t=0}^{m}\sum_{k=1}^{\gamma}\|\overline{U}_t\|_2\|\overline{U}_k-\overline{U}_k^{(k+t,*)}\|_2\le O(1)\sum_{k=1}^{\gamma}k\|\overline{U}_k-\overline{U}_k^{*}\|_2,
\end{aligned}
$$

where the first equality follows from (B.10), and the first inequality follows from Cauchy-Schwarz inequality. Finally, we can conclude that

$$
\|E_{\mathscr{F}_\gamma}[f^{(2)}(V_{1,\text{odd}}^{(\le m,*)})V_{1,\text{odd}}^{(>m,*)}\overline{V}_{1,\text{even}}]\|_1=M_fO(\gamma^{-1/2}+m^{-2})=M_fO(\gamma^{-1/2}),
$$

where the last equality follows from letting $m=\gamma^{1/3}$.

Second, using a similar (but simpler) strategy, we can show that

$$
\|f^{(1)}(V_{1,\text{odd}}^{*})((2\gamma)^{-1/2}\overline{V}_{1,\text{odd}}-V_{1,\text{odd}}^{*})\|_1=O(M_f\gamma^{-1}).
$$

Collecting the above results, the proof of the first result is now completed.

(2). Again, without loss of generality, let $j=1$, and the proof of the second result can be done in a way similar to that for the first result. We omit the details herewith.

(3)-(4). The proofs of these two results are much similar to those for the first two results of this lemma. The only difference is that we use Hölder inequality instead of the bounded derivatives whenever necessary. ■

Proof of Lemma B.4:

(1)-(2). Without loss of generality, let $j = 1$. We first establish that $\|\overline{\sigma}_{j|\gamma}^2 - \widehat{\sigma}_j^2\|_{\delta/2} = O_P(1/\gamma)$. Write

$$
\begin{aligned}
2\gamma(\overline{\sigma}_{j|\gamma}^2 - \widehat{\sigma}_j^2) &= E_{\mathscr{F}_\gamma}\left([\overline{V}_{1,\text{odd}} + \overline{V}_{1,\text{even}}]^2\right) - 2\gamma\widehat{\sigma}_j^2 \\
&= E_{\mathscr{F}_\gamma}\left([\sum_{k=1}^{\gamma}(\overline{U}_k^* + (\overline{U}_k - \overline{U}_k^*) - E[\overline{U}_k \mid \mathscr{F}_\gamma]) + \overline{V}_{1,\text{even}}]^2\right) - 2\gamma\widehat{\sigma}_j^2.
\end{aligned}
$$

By squaring out the first term, we have a sum of squared terms and a sum of interaction terms. Here, we treat the interaction terms first:

$$
\begin{aligned}
&2\sum_{t=1}^{\gamma}\sum_{s=1}^{\gamma} E_{\mathscr{F}_\gamma}\left[\overline{U}_t^*(\overline{U}_s - \overline{U}_s^*) - \overline{U}_t^* E[\overline{U}_s \mid \mathscr{F}_\gamma] - E[\overline{U}_t \mid \mathscr{F}_\gamma](\overline{U}_s - \overline{U}_s^*)\right] \\
&+2\sum_{t=1}^{\gamma} E_{\mathscr{F}_\gamma}\left[\overline{V}_{1,\text{even}}\overline{U}_t^* + \overline{V}_{1,\text{even}}(\overline{U}_t - \overline{U}_t^*) - \overline{V}_{1,\text{even}}E[\overline{U}_t \mid \mathscr{F}_\gamma]\right] \\
:=\ &2I_{\gamma,1} - 2I_{\gamma,2} - 2I_{\gamma,3} + 2I_{\gamma,4} + 2I_{\gamma,5} - 2I_{\gamma,6}.
\end{aligned}
$$

Consider $I_{\gamma,1}$, and write

$$
\begin{aligned}
I_{\gamma,1} &= \sum_{s=1}^{\gamma}\sum_{t=s}^{\gamma} E\left[\overline{U}_t^*(\overline{U}_s - \overline{U}_s^*) \mid \mathscr{F}_\gamma\right] + \sum_{s=1}^{\gamma}\sum_{t=1}^{s-1} E\left[\overline{U}_t^*(\overline{U}_s - \overline{U}_s^*) \mid \mathscr{F}_\gamma\right] \\
&= \sum_{s=1}^{\gamma}\sum_{t=s}^{\gamma} E\left[(\overline{U}_s - \overline{U}_s^*)E(\overline{U}_t^* \mid \mathscr{F}_s, \mathscr{F}_s^*) \mid \mathscr{F}_\gamma\right] + \sum_{s=1}^{\gamma}\sum_{t=1}^{s-1} E\left[\overline{U}_t^*(\overline{U}_s - \overline{U}_s^*) \mid \mathscr{F}_\gamma\right].
\end{aligned}
$$

Note that for $s \leq t \leq \gamma$

$$
E_{\mathscr{F}_\gamma}(\overline{U}_t^* \mid \mathscr{F}_s, \mathscr{F}_s^*) =_D E(\overline{U}_t \mid \mathscr{F}_s) = E(\overline{U}_t - \overline{U}_t^{(t-s,*)} \mid \mathscr{F}_s),
$$

where the equality follows from the fact that $E(\overline{U}_t^{(t-s,*)} \mid \mathscr{F}_s) = 0$. Thus, by Cauchy-Schwarz inequality and Jensen's inequality, we have

$$
\begin{aligned}
\|I_{\gamma,1}\|_{\delta/2} &\leq O(1)\sum_{s=1}^{\gamma}\sum_{t=s}^{\gamma}\|\overline{U}_s - \overline{U}_s^*\|_\delta\|\overline{U}_t - \overline{U}_t^{(t-s,*)}\|_\delta \\
&\quad +O(1)\sum_{s=1}^{\gamma}\sum_{t=1}^{s-1}\|\overline{U}_s - \overline{U}_s^*\|_\delta\|\overline{U}_t^*\|_\delta = O(1),
\end{aligned}
$$

where the equality follows from Assumption 1.

Consider $I_{\gamma,2}$. Since $E[\overline{U}_t^* \mid \mathscr{F}_\gamma] = E[\overline{U}_t] = 0$ for $1 \leq t \leq \gamma$, we have $I_{\gamma,2} = 0$.

Consider $I_{\gamma,3}$. By Jensen's inequality, we have

$$\|E[\overline{U}_t \mid \mathscr{F}_\gamma]\|_\delta = \|E[\overline{U}_t - \overline{U}_t^* \mid \mathscr{F}_\gamma]\|_\delta \le \|\overline{U}_t - \overline{U}_t^*\|_\delta.$$

Thus, we can obtain that

$$\|I_{\gamma,3}\|_{\delta/2} \le O(1)\left(\sum_{t=1}^{\gamma}\|\overline{U}_t - \overline{U}_t^*\|_\delta\right)^2 = O(1).$$

Consider $I_{\gamma,4}$. Note that $\overline{U}_t$ and $\overline{U}_s^{(s-t,*)}$ are independent for $1 \le t \le \gamma$ and $\gamma+1 \le s \le 2\gamma$, and $E[\overline{U}_t^* \mid \mathscr{F}_\gamma] = 0$. Thus we can write

$$\sum_{t=1}^{\gamma} E_{\mathscr{F}_\gamma}\left[\overline{V}_{1,\text{even}}\overline{U}_t^*\right] = \sum_{t=1}^{\gamma}\sum_{s=\gamma+1}^{2\gamma} E\left[(\overline{U}_s - \overline{U}_s^{(s-t,*)})\overline{U}_t^* \mid \mathscr{F}_\gamma\right].$$

By Cauchy-Schwarz inequality and Jensen's inequality,

$$\|I_{\gamma,4}\|_{\delta/2} \le O(1)\sum_{t=1}^{\gamma}\sum_{s=\gamma+1}^{2\gamma}\|\overline{U}_t^*\|_\delta\|\overline{U}_s - \overline{U}_s^{(s-t,*)}\|_\delta \le O(1)\sum_{t=1}^{\infty} t\|\overline{U}_t - \overline{U}_t^*\|_\delta < \infty.$$

Consider $I_{\gamma,5}$. By using Cauchy-Schwarz inequality, Jensen's inequality and Lemma B.2.1, we have

$$\|I_{\gamma,5}\|_{\delta/2} \le O(1)\sum_{t=1}^{\gamma}\|\overline{U}_t - \overline{U}_t^*\|_\delta = O(1).$$

Similarly, we have $I_{\gamma,6} = O(1)$.

We next deal with the squared terms:

$$\begin{aligned}
&\sum_{t=1}^{\gamma}\sum_{s=1}^{\gamma} E_{\mathscr{F}_\gamma}\left[\overline{U}_t^*\overline{U}_s^* + (\overline{U}_t - \overline{U}_t^*)(\overline{U}_s - \overline{U}_s^*)\right] \\
&-\sum_{t=1}^{\gamma}\sum_{s=1}^{\gamma} E_{\mathscr{F}_\gamma}(\overline{U}_t)E_{\mathscr{F}_\gamma}(\overline{U}_s) + E_{\mathscr{F}_\gamma}[\overline{V}_{1,\text{even}}^2] \\
:=\ & I_{\gamma,7} + I_{\gamma,8} + I_{\gamma,9} + I_{\gamma,10}.
\end{aligned}$$

Consider $I_{\gamma,7}$. For $1 \le t, s \le \gamma$, since $E_{\mathscr{F}_\gamma}[\overline{U}_t^*\overline{U}_s^*] = E[\overline{U}_t^*\overline{U}_s^*] = E[\overline{U}_t\overline{U}_s]$, we have $I_{\gamma,7} = 2\gamma\widehat{\sigma}_1^2$. Similar to the above development, we can show that $\|I_{\gamma,8}\|_{\delta/2} = O(1)$, $\|I_{\gamma,9}\|_{\delta/2} = O(1)$ and $\|I_{\gamma,10}\|_{\delta/2} = O(1)$. Hence, we have proved that $\|\overline{\sigma}_{j|\gamma}^2 - \widehat{\sigma}_j^2\|_{\delta/2} = O(\gamma^{-1})$. In addition, from the above arguments, we can easily establish $|\overline{\sigma}_j^2 - \widehat{\sigma}_j^2| = O(\gamma^{-1})$. Hence, the first two results follow.

(3). Let $\mathcal{I} = \{1,3,5,\ldots\}$ and $\mathcal{J} = \{2,4,6,\ldots\}$ such that $\mathcal{I}\cup\mathcal{J} = \{1,2,\ldots,n\}$, and write

$$\|\sum_{j=1}^{n}(\overline{\sigma}_{j|\gamma}^2 - \overline{\sigma}_j^2)\|_{\delta/2} \le \|\sum_{j\in\mathcal{I}}(\overline{\sigma}_{j|\gamma}^2 - \overline{\sigma}_j^2)\|_{\delta/2} + \|\sum_{j\in\mathcal{J}}(\overline{\sigma}_{j|\gamma}^2 - \overline{\sigma}_j^2)\|_{\delta/2}.$$

Note that $\{(\overline{\sigma}_{j|\gamma}^2 - \overline{\sigma}_j^2)\}_{j\in\mathcal{I}}$ is a sequence of independent random variables under the probability measure Pr, and the same is true for $\{(\overline{\sigma}_{j|\gamma}^2 - \overline{\sigma}_j^2)\}_{j\in\mathcal{J}}$. By Proposition A.1.1, we have

$$\|\sum_{j=1}^{n}(\overline{\sigma}_{j|\gamma}^2-\overline{\sigma}_j^2)\|_{\delta/2}\leq n^{1/2}\|\overline{\sigma}_{j|\gamma}^2-\overline{\sigma}_j^2\|_{\delta/2}.$$

Hence, the result follows from the first result of this lemma.

(4). Note that $E_{\mathscr{F}_\gamma}(V_{j,\mathrm{odd}}^{*3})=E(V_{j,\mathrm{odd}}^{*3})=E(\overline{S}_{j|\mathbb{N}}^3)$. By Lemma B.3.3, we have

$$\|\overline{\kappa}_{j|\gamma}^3-E(\overline{S}_{j|\mathbb{N}}^3)\|_1=O(\gamma^{-1}).$$

Similarly, we have $\|\overline{\kappa}_{j|\gamma}^3-\overline{\kappa}_j^3\|_1=O(\gamma^{-1})$.

(5). We first prove the first equality. Since $E_{\mathscr{F}_\gamma}[\overline{S}_{\mathbb{N}|\gamma}]=0$, using the iterated law of expectation, we have

$$E(\overline{S}_{\mathbb{N}}^3)=E(\overline{S}_{\mathbb{N}|\gamma}^3)+E(\widetilde{S}_{\mathbb{N}|\gamma}^3)+3E(\widetilde{S}_{\mathbb{N}|\gamma}E_{\mathscr{F}_\gamma}(\overline{S}_{\mathbb{N}|\gamma}^2)).$$

It then suffices to show $E(\widetilde{S}_{\mathbb{N}|\gamma}E_{\mathscr{F}_\gamma}(\overline{S}_{\mathbb{N}|\gamma}^2))=O(T^{-1}n^{1/2})$. Since $E[\widetilde{S}_{\mathbb{N}|\gamma}]=0$, we have

$$|E(\widetilde{S}_{\mathbb{N}|\gamma}E_{\mathscr{F}_\gamma}(\overline{S}_{\mathbb{N}|\gamma}^2))|=|E(\widetilde{S}_{\mathbb{N}|\gamma}[E_{\mathscr{F}_\gamma}(\overline{S}_{\mathbb{N}|\gamma}^2)-E(\overline{S}_{\mathbb{N}|\gamma}^2)])|.$$

By Hölder's inequality and the third result of this lemma, we have $|E(\widetilde{S}_{\mathbb{N}|\gamma}E_{\mathscr{F}_\gamma}(\overline{S}_{\mathbb{N}|\gamma}^2))|=O(T^{-1}n^{1/2})$.

We next prove the second equality. Since $\{V_{j,\mathrm{odd}}^*\}_{j=1}^n$ and $\{V_{j,\mathrm{even}}^*\}_{j=1}^n$ are receptively $\Pr_{\mathscr{F}_\gamma}$-independent and Pr-independent, we have the second equality.

(6)-(9). The proofs are very much similar to those of the first four results of this lemma, so we omit the details. ■

Proof of Lemma B.5:

We define a few variables to facilitate development. Define $\overline{U}_{i,m}=E[\overline{U}_i\mid\varepsilon_i,\dots,\varepsilon_{i-m}]$, $S_t=\sum_{i=1}^t\overline{U}_i$ and $S_{t,m}=\sum_{i=1}^t\overline{U}_{i,m}$. Let x_m, $m=1,\dots,\gamma$ be a positive sequence such that $\sum_{m=1}^{\gamma}x_m\leq 1$. Hence, we can rewrite $\overline{U}_t$ as

$$\overline{U}_t=\overline{U}_t-\overline{U}_{t,\gamma}+\sum_{m=1}^{\gamma}(\overline{U}_{t,m}-\overline{U}_{t,m-1})+\overline{U}_{t,0}.$$

Define $X_{t,m}=\sum_{i=1}^t(\overline{U}_{i,m}-\overline{U}_{i,m-1})$ and thus $S_{t,\gamma}-S_{t,0}=\sum_{i=1}^t(\overline{U}_{i,\gamma}-\overline{U}_{i,0})=\sum_{m=1}^{\gamma}X_{t,m}$. Let $\overline{X}_{T,m}=\max_{1\leq t\leq T}|X_{t,m}|$. For each $1\leq m\leq\gamma$, let $Y_{t,m}=\sum_{i=1+(t-1)m}^{\min\{tm,T\}}(\overline{U}_{i,m}-\overline{U}_{i,m-1})$, in which $1\leq t\leq l$ and $l=\lfloor T/m\rfloor+1$. Define $\lfloor t\rceil_m:=\lfloor t/m\rfloor m$.

We are now ready to start the proof:

$$\begin{aligned}\Pr\left(\overline{X}_{T,m}\geq 3x_jx\right)&\leq\Pr\left(\max_{1\leq t\leq T}|X_{\lfloor t\rceil_m,m}|\geq 2x_jx\right)+\Pr\left(\max_{1\leq t\leq T}|X_{\lfloor t\rceil_m,m}-X_{t,m}|\geq x_jx\right)\\&\leq\Pr\left(\max_{1\leq s\leq l}\left|\sum_{t=1}^{s}(1+(-1)^t)/2\times Y_{t,m}\right|\geq x_jx\right)+\Pr\left(\max_{1\leq s\leq l}\left|\sum_{t=1}^{s}(1-(-1)^t)/2\times Y_{t,m}\right|\geq x_jx\right)\end{aligned}$$

$$+\sum_{t=1}^{l}\Pr\left(\max_{1+(t-1)m\le j\le \min\{tm,T\}}\left|\sum_{i=1+(t-1)m}^{j}(\overline{U}_{i,m}-\overline{U}_{i,m-1})\right|\ge x_j x\right)$$
$$:= I_{m,1}+I_{m,2}+I_{m,3}.$$

Since $Y_{2,m}, Y_{4,m}, \ldots$ are independent, by the classical Nagaev inequality for independent random variables (Corollary 1.8 of Nagaev, 1979), we have

$$I_{m,1}\le\left(1+\frac{2}{4}\right)^4\frac{\sum_{s=1}^{l}E\left(Y_{s,m}^4\right)}{x^4}+2\exp\left(-\frac{2x^2}{e^4(4+2)^2\sum_{s=1}^{l}E\left(Y_{s,m}^2\right)}\right).$$

In addition, since $\{\overline{U}_{k,m}-\overline{U}_{k,m-1}\}_{k=1}^{T}$ are martingale differences with respect to $\sigma(\varepsilon_{k-m},\varepsilon_{k-m-1},\ldots)$, by Burkholder's inequality, we have

$$\left[E\left(Y_{s,m}^4\right)\right]^{1/4}\le O(1)\left(E\left[\sum_{i=1+(t-1)m}^{\min\{tm,T\}}|\overline{U}_{i,m}-\overline{U}_{i,m-1}|^2\right]^2\right)^{1/4}$$
$$\le O(1)\left(\sum_{i=1+(t-1)m}^{\min\{tm,T\}}\|\overline{U}_{i,m}-\overline{U}_{i,m-1}\|_4^2\right)^{1/2}=O(1)\theta_{m,4}^{U}.$$

and by $(\sum_{i=1}^{m}|a_i|)^4\le m^{4-1}\sum_{i=1}^{m}|a_i|^4$ and the conditions on x, we have $I_{m,1}=O(1)\frac{T}{x^4}(m)^{4/2-1}\theta_{m,4}^{U,4}$. A similar bound holds for $I_{m,2}$.

For $I_{m,3}$, by Burkholder's inequality and Doob inequality, we have

$$E\left(\max_{1+(t-1)m\le j\le\min\{tm,T\}}|\sum_{i=1+(t-1)m}^{j}(\overline{U}_{i,m}-\overline{U}_{i,m-1})|^4\right)$$
$$\le 2^{4-1}E\left(|Y_{t,m}|^4+\max_{1+(t-1)m\le j\le\min\{tm,T\}}|\sum_{i=j}^{\min\{tm,T\}}(\overline{U}_{i,m}-\overline{U}_{i,m-1})|^4\right)$$
$$\le 2^{4-1}E\left(Y_{t,m}^4\right)+2^{4-1}(\frac{4}{4-1})^4E\left(Y_{t,m}^4\right).$$

Hence, we have $\Pr\left(\overline{X}_{T,m}\ge 3x_jx\right)=(m)^{4/2-1}\theta_{m,4}^{U,4}O(\frac{T}{x^4})$.

By the above derivations and the classical Nagaev inequality for independent random variables again, we have

$$\Pr\left(\max_{1\le t\le T}|S_t|\ge 5x\right)=\Pr\left(\max_{1\le t\le T}|S_t-S_{t,\gamma}+S_{t,\gamma}-S_{t,0}+S_{t,0}|\ge 5x\right)$$
$$\le\Pr\left(\max_{1\le t\le T}|S_t-S_{t,\gamma}|\ge x\right)+\sum_{m=1}^{\gamma}\Pr\left(X_{T,m}\ge 3x_mx\right)+\Pr\left(\max_{1\le t\le T}|S_{t,0}|\ge x\right)$$
$$\le 0+\sum_{m=1}^{\gamma}(m)^{4/2-1}\theta_{m,4}^{U,4}O\left(\frac{T}{x^4}\right)+O\left(\frac{T}{x^4}\right)=O\left(\frac{T}{x^4}\right).$$

The proof is now completed. ■

Proof of Lemma B.6:

(1). By Lemma B.2.3, we have $\overline{\kappa}_j^3 = O(1/\sqrt{\gamma})$. Let $\{A_{j,t}\}_{t=1}^{\gamma}$ be i.i.d. random variables with continuous distribution function $F_j(\cdot)$ such that[4] $E(A_{j,t}) = 0$, $E(A_{j,t}^2) = \overline{\sigma}_j^2$, $E(A_{j,t}^3) = \sqrt{\gamma}\overline{\kappa}_j^3$ and $E|A_{j,t}^8| = O(1)$. It follows that the first three moments of $\gamma^{-1/2}\sum_{t=1}^{\gamma} A_{j,t}$ coincide with 0, $\overline{\sigma}_j^2$ and $\overline{\kappa}_j^3$. In addition, by Lemma B.2.4 and Lemma B.4.6, we have

$$E\left(|A_{j,t}|^4\right) - \overline{\tau}_j^4 = O(1/\gamma),$$
$$E\left(|A_{j,t}|^4\right) - E(S_{j|\mathbb{N}}^4) = O(1/\gamma).$$

Let $\overline{Z}_j =_D \gamma^{-1/2}\sum_{t=1}^{\gamma} A_{j,t}$, we can conclude that

$$E(\overline{Z}_j) = 0, \quad E(\overline{Z}_j^2) - \overline{\sigma}_j^2 = O(\gamma^{-1}), \quad E(\overline{Z}_j^3) - \overline{\kappa}_j^3 = O(\gamma^{-1}), \quad E(\overline{Z}_j^4) - \overline{\tau}_j^4 = O(\gamma^{-1}).$$

Since the first four moments of $\overline{Z}_j$ (only three are necessary here) match those of $S_{j|\mathbb{N}}$ up to an error term of order $O(\gamma^{-1})$ by Lemmas B.4.2, B.4.4 and B.4.6, then the result follows directly from the classic Edgeworth expansion and Lemma B.1.

(2). By Lemma B.4, the first four moments of $\overline{Z} + \widetilde{Z}$ match those of $S_{\mathbb{N}}$ up to an error of order $O_P(\gamma^{-1})$, so the second result can be justified in the same fashion as for the first result of this lemma.

(3). By the construction of $\overline{Z}_j$ in the proof of part (1), and using the classical Fuk-Nagaev inequality (which is a simple version of Lemma B.5), the proof is the same as in the proof of Lemma B.7, so we omit it here for the time being. ■

Proof of Lemma B.7:

Before proving this lemma, we show that

$$\begin{aligned}&\|E_{\mathscr{F}_\gamma}(f(x_n\overline{V}_j)) - E(f(x_n\overline{Z}_j))\|_1 \\ &= (|x_n|^2 + \tau_T^5|x_n|^5)O(\gamma^{-1}) + \sup_{x\in\mathbb{R}}|\Delta_{j,\gamma}(x)|O(\tau_T^5|x_n|^5 + \tau_T^6|x_n|^6).\end{aligned} \tag{B.11}$$

Then the results of this lemma then follow either immediately or in the same manner.

By the following Taylor expansion:

$$f(x+h) - f(x) = \sum_{j=1}^{s+1}\frac{f^{(j)}(x)h^j}{j!} + \frac{h^{s+1}}{s!}\int_0^1 (1-t)^s(f^{(s+1)}(th+x) - f^{(s+1)}(x))\mathrm{d}t$$

and $E_{\mathscr{F}_\gamma}(\overline{V}_j) = E(\overline{V}_j) = 0$, we have

$$\begin{aligned}&E_{\mathscr{F}_\gamma}(f(x_n\overline{V}_j)) - E(f(x_n\overline{Z}_j)) \\ &= \frac{x_n^2}{2}(\overline{\sigma}_{j|\gamma}^2 - \overline{\sigma}_j^2)f^{(2)}(0) + \frac{x_n^3}{6}(\overline{\kappa}_{j|\gamma}^3 - \overline{\kappa}_j^3)f^{(3)}(0)\end{aligned}$$

[4]Note that the existence of $E|A_{j,t}^8|$ is only needed in the proof for the third result of this lemma.

$$
\begin{aligned}
&+\frac{1}{2}\int_0^1 (1-t)^2 E_{\mathscr{F}_\gamma}\left((x_n\overline{V}_j)^3(f^{(3)}(tx_n\overline{V}_j)-f^{(3)}(0))\right)\mathrm{d}t \\
&-\frac{1}{2}\int_0^1 (1-t)^2 E\left((x_n\overline{Z}_j)^3(f^{(3)}(tx_n\overline{Z}_j)-f^{(3)}(0))\right)\mathrm{d}t \\
:=\ & I_1+I_2+I_3,
\end{aligned}
$$

where the definitions of I_1 to I_3 are obvious. By Lemma B.4, we have $E|I_1| = O(x_n^2\gamma^{-1}+|x_n|^3\gamma^{-1})$. Thus, we just need to focus on I_2 and I_3 below.

Let $h_T(x)$ be a three times continuously differentiable function such that $h_T(x)=1$ if $|x|\le \tau_t/2$, $h_T(x)=0$ if $|x|\ge\tau_t$, and $h_T^{(s)}(x)\le M$ for $s\in\{0,1,2,3\}$. Note that for a random variable X and $q\ge 1$, we have

$$
E\left[|X|^q I(|X|\ge\tau_T)\right]\le q\tau_T^q\Pr(|X|\ge\tau_T)+q\int_{\tau_T}^\infty x^{q-1}\Pr(|X|\ge x)\mathrm{d}x. \tag{B.12}
$$

Hence, by Lemma B.5,

$$
\begin{aligned}
&E[|\overline{V}_j|^3(1-h_T(\overline{V}_j))]\le E[|\overline{V}_j|^3 I(|\overline{V}_j|\ge\tau_T/2)] = O(\tau_T^{3-4}\gamma^{1-4/2})+O(\gamma^{1-4/2})\int_{\tau_T/2}^\infty x^{2-4}\mathrm{d}x \\
&=O(\tau_T^{-1}\gamma^{-1}).
\end{aligned}
$$

Note that by $|f^{(3)}(x)|\le 1$ and Jensen's inequality, we have

$$
\begin{aligned}
&\|E_{\mathscr{F}_\gamma}\left[(x_n\overline{V}_j)^3(f^{(3)}(tx_n\overline{V}_j)-f^{(3)}(0))(1-h_T(\overline{V}_j))\right]\|_1 \\
\le\ & O(1)|x_n^3|\cdot\||\overline{V}_j|^3(1-h_T(\overline{V}_j))\|_1 = O(|x_n^3|\tau_T^{-1}\gamma^{-1}) = o(|x_n^3|\gamma^{-1}).
\end{aligned}
$$

Using Taylor expansion again, we have

$$
\begin{aligned}
&E_{\mathscr{F}_\gamma}\left[(x_n\overline{V}_j)^3(f^{(3)}(tx_n\overline{V}_j)-f^{(3)}(0))h_T(\overline{V}_j)\right] \\
=\ & tx_n^4\int_0^1(1-s)E_{\mathscr{F}_\gamma}\left[\overline{V}_j^4 f^{(4)}(stx_n\overline{V}_j)h_T(\overline{V}_j)\right]\mathrm{d}s.
\end{aligned}
$$

Let $g(x)=x^4 f^{(4)}(stx_nx)h_T(x)$. By Taylor expansion, we have

$$
g(x)=h_T(x)\left(x^4f^{(4)}(0)+stx_nx^5f^{(5)}(0)+stx_nx^5\int_0^1(1-u)[f^{(5)}(stux_nx)-f^{(5)}(0)]\mathrm{d}u\right).
$$

Let $g_2(x)=h_T(x)x^5\int_0^1(1-u)[f^{(5)}(stux_nx)-f^{(5)}(0)]\mathrm{d}u$. As the derivatives of $f(\cdot)$ are uniformly bounded and by the definition of $h_T(x)$, we have $|g_2^{(s)}(x)|\le O(1)\tau_T^5(1+|x_n^3|)$ for $s\in\{0,1,2,3\}$. Then by Lemma B.3.1, we have

$$
\|E_{\mathscr{F}_\gamma}(stx_n^5g_2(\overline{V}_j)-stx_n^5g_2(V_{j,\mathrm{odd}}^*))\|_1 = O(\gamma^{-1}\tau_T^5(|x_n^5|+|x_n^8|)).
$$

As $g_2(0)=0$ and for any random variable Y and differentiable function f

$$E[f(Y)-f(0)] = \int_0^\infty f^{(1)}(y)\Pr(Y\geq y)\mathrm{d}y - \int_{-\infty}^0 f^{(1)}(y)\Pr(Y\leq y)\mathrm{d}y,$$

and $h_T(x)$, $h_T^{(1)}(x)$ and $g_2^{(1)}(x)$ equal to zero for $|x|>\tau_T$, we obtain that

$$\begin{aligned} E_{\mathscr{F}_\gamma}(g_2(V^*_{j,\text{odd}})) &= \int_0^{\tau_T} g_2^{(1)}(x)\Pr_{\mathscr{F}_\gamma}(V^*_{j,\text{odd}}\geq x)\mathrm{d}x - \int_{-\tau_T}^0 g_2^{(1)}(x)\Pr_{\mathscr{F}_\gamma}(V^*_{j,\text{odd}}\leq x)\mathrm{d}x \\ &= \int_0^{\tau_T} g_2^{(1)}(x)\Pr(V^*_{j,\text{odd}}\geq x)\mathrm{d}x - \int_{-\tau_T}^0 g_2^{(1)}(x)\Pr(V^*_{j,\text{odd}}\leq x)\mathrm{d}x, \end{aligned}$$

where we use the fact that $V^*_{j,\text{odd}}$ is independent of $\mathscr{F}_\gamma$. In addition, we have

$$\left\|\int_0^{\tau_T} g_2^{(1)}(x)\left[\Pr(V^*_{j,\text{odd}}\geq x)-\Pr(\overline{Z}_j\geq x)\right]\mathrm{d}x\right\|_1 \leq O(1)\sup_{x\in\mathbb{R}}|\Delta_{j,\gamma}(x)|\left(\tau_T^5+\tau_T^6|x_n|\right).$$

Similarly, we have

$$\left\|\int_{-\tau_T}^0 g_2^{(1)}(x)\left[\Pr(V^*_{j,\text{odd}}\leq x)-\Pr(\overline{Z}_j\leq x)\right]\mathrm{d}x\right\|_1 \leq O(1)\sup_{x\in\mathbb{R}}|\Delta_{j,\gamma}(x)|\left(\tau_T^5+\tau_T^6|x_n|\right).$$

Combing the above derivations, we obtain that

$$\begin{aligned} &\|E_{\mathscr{F}_\gamma}(stx_n^5 g_2(\overline{V}_j) - stx_n^5 g_2(\overline{Z}_j))\|_1 \\ &= O(\gamma^{-1}\tau_T^5(|x_n^5|+|x_n^8|)) + O(1)\sup_{x\in\mathbb{R}}|\Delta_{j,\gamma}(x)|\left(\tau_T^5|x_n|^5+\tau_T^6|x_n|^6\right). \end{aligned}$$

Using (B.12) and similar arguments to the above, we obtain

$$\|E_{\mathscr{F}_\gamma}\left[(x_n\overline{V}_j)^4(1-h_T(\overline{V}_j))\right]\|_1 = o(|x_n^4|\gamma^{-1}).$$

By Lemma B.4.6, we have $\|E_{\mathscr{F}_\gamma}(\overline{V}_j^4-\overline{Z}_j^4)\|_1 = O(\gamma^{-1})$. Hence, putting every piece together, we have

$$\begin{aligned} &\left\|E_{\mathscr{F}_\gamma}\left[(x_n\overline{V}_j)^3(f^{(3)}(tx_n\overline{V}_j)-f^{(3)}(0))\right] - t\int_0^1(1-s)E_{\mathscr{F}_\gamma}\left[x_n^4\overline{Z}_j^4 f^{(4)}(stx_n\overline{Z}_j)h_T(\overline{Z}_j)\right]\mathrm{d}s\right\|_1 \\ &= O(|x_n^3|\tau_T^{-1}\gamma^{-1}) + o(|x_n^4|\gamma^{-1}) + O(\gamma^{-1}\tau_T^5(|x_n^5|+|x_n^8|)) + O(1)\sup_{x\in\mathbb{R}}|\Delta_{j,\gamma}(x)|\left(\tau_T^5|x_n|^5+\tau_T^6|x_n|^6\right). \end{aligned}$$

Using the same arguments as the above and in the proof of Lemma B.6.3, we have

$$\begin{aligned} &\left\|E_{\mathscr{F}_\gamma}\left[(x_n\overline{Z}_j)^3(f^{(3)}(tx_n\overline{Z}_j)-f^{(3)}(0))\right] - t\int_0^1(1-s)E_{\mathscr{F}_\gamma}\left[x_n^4\overline{Z}_j^4 f^{(4)}(stx_n\overline{Z}_j)h_T(\overline{Z}_j)\right]\mathrm{d}s\right\|_1 \\ &= O(|x_n^3|\gamma^{-2}). \end{aligned}$$

Finally, we obtain

$$\begin{aligned} &\left\|E_{\mathscr{F}_\gamma}(f(x_n\overline{V}_j)) - E(f(x_n\overline{Z}_j))\right\|_1 \\ &= (|x_n|^2+\tau_T^5|x_n|^5)O(\gamma^{-1}) + \sup_{x\in\mathbb{R}}|\Delta_{j,\gamma}(x)|O(\tau_T^5|x_n|^5+\tau_T^6|x_n|^6), \end{aligned}$$

which completes the proof of (B.11).

(1). The proof is identical to the development of (B.11).

(2). By Taylor expansion, we have

$$f(x_n V^{\diamond}_{j,\text{odd}}) = f(x_n V^{*}_{j,\text{odd}}) + \frac{x_n}{\sqrt{2\gamma}} f^{(1)}(x_n V^{*}_{j,\text{odd}})(H_{(2j-1)\gamma+1} - H_{(2j-2)\gamma+1})$$
$$+ \frac{x_n^2}{4\gamma} f^{(2)}(x_n V^{*}_{j,\text{odd}})(H_{(2j-1)\gamma+1} - H_{(2j-2)\gamma+1})^2 + o\left(\frac{x_n^2}{4\gamma}(H_{(2j-1)\gamma+1} - H_{(2j-2)\gamma+1})^2\right)$$

and thus by using $E(H_t) = 0$ and $f^{(s)}$ is bounded,

$$\|E(f(x_n V^{\diamond}_{j,\text{odd}})) - E(f(x_n V^{*}_{j,\text{odd}}))\|_1 \leq O(1)\frac{x_n^2}{\gamma}.$$

Then, by triangle inequality and Lemma B.3.1, we have

$$\begin{aligned}
&\|E_{\mathscr{F}_\gamma}(f(x_n \overline{V}_j)) - E(f(x_n V^{\diamond}_{j,\text{odd}}))\|_1 \\
&\leq \|E_{\mathscr{F}_\gamma}(f(x_n \overline{V}_j)) - E(f(x_n V^{*}_{j,\text{odd}}))\|_1 + \|E(f(x_n V^{\diamond}_{j,\text{odd}})) - E(f(x_n V^{*}_{j,\text{odd}}))\|_1 \\
&= O(x_n^2/\gamma).
\end{aligned}$$

The above bound also applies to $\overline{Z}_j$. Hence, the second result follows directly from the first result.

(3)-(4). The proofs of these two results are almost identical to those for the first two results of this lemma. ■

Proof of Lemma B.8:

Before proving the results of this lemma, we show that

$$\sup_{w\in\mathbb{R}} |\Delta^{\diamond}_T(w)| = O(n/T) + (\log T)^{5/2} n^{-3/2} \sup_{w\in\mathbb{R}} |\Delta^{\diamond}_{j,\gamma}(w)|. \tag{B.13}$$

Then the two results of this lemma follow immediately.

By Berry's smoothing inequality (Lemma 2, XVI.3 in Feller, 1970), for $T^+ \geq c\sqrt{T}$, we obtain

$$\sup_{w\in\mathbb{R}} |\Delta^{\diamond}_T(w)| \leq O(1)\left(\mathscr{U}^{\diamond}_T + \sup_{w\in\mathbb{R}}(\mathscr{T}^{T+}_{c\sqrt{T}})^{\diamond}(w) + (T^+)^{-1}\right).$$

Selecting $a > 0$ of the density function of H_t such that $c > ab$ and using $E(e^{ixS^{\diamond}_{\mathbb{N}}}) = 0$ for $|x| > \sqrt{T}|ab|$, we have $\sup_{w\in\mathbb{R}}(\mathscr{T}^{T+}_{c\sqrt{T}})^{\diamond}(w) = 0$. In addition, by setting $T^+ = +\infty$, we only need to focus on $\mathscr{U}^{\diamond}_T$.

By $|e^{ix}| = 1$, $\overline{S}^{\diamond}_{j|\mathbb{N}} =_D V^{\diamond}_{j,\text{odd}}$, $\widetilde{S}^{\diamond}_{j|\mathbb{N}} =_D V^{\diamond}_{j,\text{even}}$ and the properties of conditional expectation, we have

$$\left|E(e^{ixS^{\diamond}_{\mathbb{N}}}) - E(e^{ix(\overline{Z}^{\diamond}+\widetilde{Z}^{\diamond})})\right|$$

$$= \left| E\left[E_{\mathscr{F}_\gamma}\left[e^{ix\sum_{j=1}^n \overline{S}_{j|\mathbb{N}}^\diamond/\sqrt{n}} \right] e^{ix\sum_{j=1}^n \widetilde{S}_{j|\mathbb{N}}^\diamond/\sqrt{n}} \right] - E(e^{ix\overline{Z}^\diamond})E(e^{ix\widetilde{Z}^\diamond}) \right|$$
$$\leq \|E_{\mathscr{F}_\gamma}\left[e^{ix\sum_{j=1}^n \overline{S}_{j|\mathbb{N}}^\diamond/\sqrt{n}} \right] - E(e^{ix\overline{Z}^\diamond})\|_1 + |E\left[e^{ix\sum_{j=1}^n \widetilde{S}_{j|\mathbb{N}}^\diamond/\sqrt{n}} \right] - E(e^{ix\widetilde{Z}^\diamond})|$$
$$= |E\left[e^{ix\sum_{j=1}^n V_{j,\mathrm{odd}}^\diamond/\sqrt{n}} \right] - E(e^{ix\overline{Z}^\diamond})| + |E\left[e^{ix\sum_{j=1}^n V_{j,\mathrm{even}}^\diamond/\sqrt{n}} \right] - E(e^{ix\widetilde{Z}^\diamond})|$$
$$:= A_\gamma(x) + B_\gamma(x).$$

Consider $A_\gamma(x)$. Define $\phi_j(x) = E(e^{ixV_{j,\mathrm{odd}}^\diamond})$, $\psi_j(x) = E(e^{ix\overline{Z}_j^\diamond})$ and $x_n = x/\sqrt{n}$. Since $\{V_{j,\mathrm{odd}}^\diamond\}$ is a sequence of random variables under Pr, $|\phi_j(x)| \leq 1$, $|\psi_j(x)| \leq 1$, and by using the equality

$$\prod_{i=1}^n a_j - \prod_{i=1}^n b_j = \sum_{i=1}^n \left(\prod_{j=1}^{i-1} b_j \right) (a_i - b_i) \left(\prod_{j=i+1}^n a_j \right),$$

we have

$$A_\gamma(x) = \left| \prod_{i=1}^n \phi_j(x_n) - \prod_{i=1}^n \psi_j(x_n) \right|$$
$$\leq \sum_{i=1}^n \left| \prod_{j=1}^{i-1} \psi_j(x_n) \right| \cdot |\phi_i(x_n) - \psi_i(x_n)| \cdot \left| \prod_{j=i+1}^n \phi_j(x_n) \right|.$$

By using Eqn (5.8) and (5.9) in Feller (1970), XVI.5, we have $\phi_j(x_n) = O(e^{-x^2/n})$, $\psi_j(x_n) = O(e^{-x^2/n})$ and $|e^{-x^2/n}|^{n-1} \leq O(1)e^{-x^2}$. Hence, by Lemma B.7.1 and the condition on n, we have

$$\int_{-c\sqrt{T}}^{c\sqrt{T}} \frac{|A_\gamma(x)|}{|x|} \mathrm{d}x \leq O(\gamma^{-1}) + (\log T)^{5/2} n^{-3/2} \sup_{w\in\mathbb{R}} |\Delta_{j,\gamma}^\diamond(w)|.$$

Similarly, using Lemma B.7.3

$$\int_{-c\sqrt{T}}^{c\sqrt{T}} \frac{|B_\gamma(x)|}{|x|} \mathrm{d}x \leq O(\gamma^{-1}) + (\log T)^{5/2} n^{-3/2} \sup_{w\in\mathbb{R}} |\Delta_{j,\gamma}^\diamond(w)|.$$

Hence, the proof of (B.13) is completed.

(1). Recall that we have let $T = 2n\gamma$. Let further that $n = C(\log T)^5$ and $C > 0$ large enough. By using (B.13) recursively, we have

$$\sup_{w\in\mathbb{R}} |\Delta_T^\diamond(w)| \leq O(1)T^{-1} n \sum_{k=0}^{\gamma+1} \left[(\log T)^5 n^{-3/2} \right]^k (2n)^k + O(1)((\log T)^5 n^{-3/2})^\gamma = O(n/T).$$

The proof of the first result is now completed.

(2). The proof is similar to that of (B.13). The difference is that we use Lemmas B.7.2 and B.7.4. It yields that

$$\mathscr{U}_T \leq O(n/T) + (\log T)^{5/2} n^{-3/2} \sup_{w\in\mathbb{R}} |\Delta_{j,\gamma}^\diamond(w)|.$$

The result follows directly from the first result of this lemma. ■