6-2021

Three essays on panel and factor models

Ji FENG
Singapore Management University

THREE ESSAYS ON PANEL AND FACTOR MODELS

JI FENG

SINGAPORE MANAGEMENT UNIVERSITY

2021

THREE ESSAYS ON PANEL AND FACTOR MODELS

JI FENG

A DISSERTATION

IN

ECONOMICS

Presented to the Singapore Management University in Partial Fulfillment

of the Requirement for the Degreee of PhD in Economics

2021

Supervisor of Dissertation

PhD in Economics, Programme Director

Three Essays on Panel and Factor Models

by

Ji FENG

Submitted to School of Economics in partial fulfillment

of the requirements for the Degree of Doctor of Philosophy in Economics

Dissertation Committee

Liangjun Su (Supervisor/Chair)
C.V. Starr Chair Professor of Economics
Tsinghua University

Jun Tu (Co-Chair)
Associate Professor of Finance
Singapore Management University

Sainan Jin
Professor of Economics
Singapore Management University

Yichong Zhang
Assistant Professor of Economics
Singapore Management University

Qu Feng
Associate Professor of Economics
National Technological University

Singapore Management University

2021

The information below is a further clarification of the committee composition:

Liangjun Su

Associate Professor of Economics

School of Economics

Singapore Management University

1 Jul 2008 to 30 Jun 2012

Lee Kong Chian Professor of Economics

School of Economics

Singapore Management University

1 Jul 2012 to Dec 2020

C.V. Starr Chair Professor of Economics

School of Economics and Management

Tsinghua University

From Jul 2020 onwards

Abstract

Three Essays on Panel and Factor Models

Ji Feng

The dissertation includes three chapters on panel and factor models. In the first chapter, we introduce a two-way linear random coefficient panel data models with fixed effects and the cross-sectional dependence. We follow the idea of the within-group fixed effects estimator to estimate parameters of interests. We establish the limiting distributions of the estimates and also propose the two-way heterogeneity bias test to check the desirability of the estimation strategy. The specification tests then are constructed to examine the existence of the slope heterogeneity and time-varyingness. We study the asymptotic properties of the specification tests and employ two bootstrap schemes to rectify the downward size distortion of the specification tests. We apply the specification tests to reveal the heterogenous relationship between the unemployment rate and youth labor rate in the working-age population. In the second chapter, we devise a simple but effective procedure to test bubbles in the idiosyncratic components in the presence of nonstationary or mildly explosive factors in common components in panel factor models. We study the asymptotic properties of our test. We also propose a wild bootstrap procedure to improve the finite sample performance of our test. As an illustrative example, we consider testing the bubbles in the idiosyncratic components of cryptocurrency prices. In the third chapter, we propose the tests constructed from estimated common factors for detecting bubbles in unobserved common factors when the idiosyncratic components follow a unit-root or local-to-unity process. We study the asymptotic properties of our proposed tests. We show that our proposed tests have non-trivial power to detect those bubbles in unobserved common factors under the alternative of local-to-unity. To implement our proposed tests, we propose to use the dependent wild bootstrap method to simulate the critical values in practice.

Table of Contents

Acknowledgments

First and foremost, I would like to sincerely thank Professor Liangjun Su for being my supervisor for the past four years. I feel extremely fortunate to learn from him and continue to benefit from his guidance and consistent encouragements. I am deeply thankful and indebted to him for sharing his expertise and knowledge with me throughout my Ph.D. study. His diligence and passion for research is inspiring to me and teach me how to be a qualified researcher. It has been a great privilege and honor to work and study under his supervision.

I would like to give special thanks to my committee members Professor Sainan Jin, Professor Jun Tu, Professor Yichong Zhang and Professor Qu Feng for their patience and support for reviewing my dissertation. My gratitude extents to Professor Yichong Zhang for taking time out of his schedule to share his insights and knowledge.

I also want to thank all participants of the econometrics reading group, namely, Wuyi Wang, Wenxin Huang, Ke Miao, Xin Zheng, Yin Xia and Shuyao Ke. It has been a pleasure studying with them and getting to know one another.

I owe everything for my parents for their love and upbringing, their companionship is invaluable for me. Last but not least, I appreciate Bingqian Lin for her understanding, support, sacrifices and love. They has given me a solid foundation to shape my future.

Chapter 1

Estimation and Specification Tests of Random Coefficient Panel Data Models with Cross-Sectional Dependence

1.1 Introduction

Heterogeneity and time-varyingness are universal phenomena in the study of economic and financial modeling. We can integrate such kind of heterogeneous effects by the panel data models with variable coefficients. There is a long history of modeling the slope heterogeneity and time-varyingness by random coeffcients models, see e.g, Hildreth and Houck (1968), Swamy (1970), and Hsiao (1974). In particular, Hsiao (1974) proposes to accommodate slope heterogeneity and time-varyingness simultaneously as random coefficients in panel data models. However, Hsiao (1974) excludes the fixed effects and does not study the inferential theory on the realizations of random coefficients. Besides, their test rely upon the stringent conditions such the normality of error terms and random coefficients. Such kind of restrictions may narrow the applications of two-way linear random coefficients panel data models. For a recent and systematic introduction to the panel data models with variable coefficients, we refer readers to Chapter 6 in Hsiao (2014).

This paper seeks to contribute as the good complement of studies on random coefficients panel data models by extending the classical setups. In particular, we introduces a two-way linear random coefficients panel data models with both time- and individual- specific fixed effects in the presence of cross-sectional dependence. We take advantages of the two-way linear random coefficients to capture the slope heterogeneity and time-varyingness, and include the time- and individual- fixed effects to account for unobserved heterogeneity that can be arbitrarily correlated with explanatory variables.

For estimation methodology, we follow the idea of the within-group estimator to obtain the estimates of parameters of interest. Importantly, we allow for the weak cross sectional dependence among observations and develop the inferential theory for these estimators under the large N and large T settings with mild conditions. We also propose the two-way heterogeneity bias test to check the desirability of our estimation strategy, and study its asymptotic properties. furthermore, we construct the max-type test statistics to examine the existence of the slope heterogeneity and time-varyingness in our model, and study the null distribution and asymptotic local power properties of the these specification test statistics. Our Monte Carlo experiments reveal that the specification tests are too conservative under the null based on the asymptotic critical values. Thus, we suggest two bootstrap schemes to correct the size of our specification tests, simulation results show that the bootstrap implementations of our specification tests have proper size and decent power against the alternatives in various scenarios.

As an illustrative application, we apply the specification tests to study the relationship between the unemployment rate and the youth labor rate in working-age population using panel data covering 171 countries at most and 23 years from 1991 to 2013.

The two-way linear random coefficients panel data models we consider in this paper links to the prevailing studies on panel data models as follows.

When there is no slope time-varyingness, our model relates to studies on heterogeneous panel data models. In particular, the correlated random-coefficient (hereafter, CRC) panel data models usually allows for the heterogeneous slopes are random and can be correlated the covariates, which is also taken into account in our model for both heterogenebous and time-varying slopes. For recent studies on the CRC penal data models, see, e.g., Gao et al. (2015), Graham et al. (2018). Besides, we can also adopt the semi- or non- parametric methods to portray the slope heterogeneity and time-varyingness in our model. There is a rapid growing literature on the the semi- or non- parametric methods for panel data models in the recent past. For example, Boneva et al. (2015) introduce the panel data models with both time- and individual- specific fixed effects that account for heterogeneous nonparametric covariate effects, and provide the estimation methodology. Gao et al. (2020) propose several estimators and the homogeneity test for the semiparametric panel data models that allows for slope heterogeneity and individual-specific trending function. To address the possible existence of slope heterogeneity, many studies also focus on offering the testing procedures, see e.g., Pesaran and Yamagata (2008), Su and Chen (2013), Juhl and Lugovskyy (2014) Ando and Bai (2015).

When there is no slope heterogeneity, our model admits time-varying panel data models as the special case. There are a large amount of researches on time-varying coefficients panel data models since the seminal work by Robinson (1989). Recently, Zhu (2017) presents the estimation methodology and inferential theory for high-dimensional time-varying panel data models with interactive effects while Chen and Huang (2018) consider a semiparametric panel data model to account for the smooth slope time-varyingness and propose the generalized Hausman tests to detect smooth structural changes.

More importantly, on the theoretical side, this paper directly contributes to studies on the panel data models with the variable slope in both the time- and individual- dimensions. The study of asymptotic properties for such as panel data models will inevitably be more challenging than that for homogeneous, heterogeneous or time-varying panel data models. Despite the tricky piece of work, there are also growing literature on this direction recently. For example, Su et al. (2017) study a heterogeneous time-varying coefficients panel data model with the latent group structure, and Chernozhukov et al. (2018) bring up so-called post-SVD inference based on the nuclear norm regularization for panel data models with variable slopes in both individual- and time- dimensions.

Keane and Neal (2020) employ a similar two-way random coefficients panel models as our model in this paper for their empirical studies, however, they does not include the fixed effects in the model; moreover, their inferential theory is developed under more stringent conditions and misleading. For the very recent, Lu and Su (2021) propose the generalized fixed effects estimator for the panel data models with the variable slope in both the time- and individual- dimensions with two-way fixed effects to achieve the uniform inference.

The remainder of the paper proceeds as follows. In Section 1.2, we describe the specification of linear random coefficients panel data models with fixed effects and propose the estimation methodology for the parameters of interests. Section 1.3 establishes the asymptotic distribution of the estimators and presents the two-way heterogeneity bias test. In Section 1.4, we propose the specification tests, and study the limiting null distribution and asymptotic local power properties of the specification test statistics. We conduct Monte Carlo experiments to evaluate the finite-sample performances of the specification tests in Section 1.5. Section 1.6 provides an empirical study to highlight the usefulness of the specification tests. Section 1.7 concludes. All proofs are relegated to the Appendix A.

NOTATION. For an $m \times n$ real matrix A, we denote its transpose as A', its trace as trA, its Frobenius norm as $\|A\|$ $(\equiv [tr\,(A'A)]^{1/2})$, and its spectral norm as $\|A\|_{\text{sp}}$ $(\equiv [\varphi_1(AA')]^{1/2})$, where $\equiv$ signifies a defintional relationship and $\varphi_k(\cdot)$ denotes the k-th largest eigenvalue of a real symmetric matrix A by counting eigenvalues of multiplicity multiple times. We also use $\varphi_{\min}$ and $\varphi_{\max}$ to stand for the minimum and maximum eigenvalues of a symmetric real matrix. Let diag$(a_1, \ldots, a_m)$ represent a $m \times m$ diagonal matrix with entries $a_1, \ldots, a_m$ on its diagonal. We write $A \asymp B$ if there exist some finite positive constants m and M such that $m|A| \leq B \leq M|A|$. We also use M to stand for a generic large positive constant that may vary across lines. For generic random variable v_{it}, let $\bar{v}_{\cdot t} \equiv \frac{1}{N}\sum_{i=1}^{N} v_{it}$, $\bar{v}_{i\cdot} \equiv \frac{1}{T}\sum_{t=1}^{T} v_{it}$, and $\bar{v}_{\cdot\cdot} \equiv \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}$. Define

$$\ddot{v}_{it} \equiv v_{it} - \bar{v}_{i\cdot} - \bar{v}_{\cdot t} + \bar{v}_{\cdot\cdot},\ \ \breve{v}_{it} \equiv v_{it} - \bar{v}_{\cdot t},\ \text{and}\ \tilde{v}_{it} \equiv v_{it} - \bar{v}_{i\cdot}.$$

Let $v_{it}^* \equiv v_{it} - E(v_{it})$. Let $\delta_{NT} \equiv \min\{\sqrt{N}, \sqrt{T}\}$. For brevity, we also use $\sum_i$ and $\sum_{i,j}$ to denote $\sum_{i=1}^N$ and $\sum_{i=1}^N \sum_{j=1}^N$, respectively, and $\sum_t$ and $\sum_{t,s}$ to denote $\sum_{t=1}^T$ and $\sum_{t=1}^T \sum_{s=1}^T$ respectively. The operator $\overset{P}{\to}$ denotes convergence in probability, $\overset{d}{\to}$ convergence in distribution, and $plim$ probability limit. We use $\mathbb{I}(\cdot)$ to denote the usual indicator function. Let $a \wedge b \equiv \min\{a, b\}$. We use $(N,T) \to \infty$ to denote that N and T pass to infinity jointly.

1.2 The Model and Estimators

In this section we introduce the model and estimates.

1.2.1 The model

We consider the following model:

$$y_{it} = x_{it}'(\beta + \lambda_i + \gamma_t) + \eta_i + \omega_t + u_{it}, \tag{1.2 .1}$$

where $i = 1, ..., N$, $t = 1, ..., T$, y_{it} is a sclar dependent variable, x_{it} is a $K \times 1$ vector of covariates, β, λ_i and γ_t are $K \times 1$ vectors of unknown parameters, η_i and ω_t stand for individual- and time- specific fixed effects (FE), respectively, and u_{it} is the idiosyncratic error term. The sequences $\{x_{it}, u_{it}\}$ are allowed to be serially and cross-sectionally dependent. We observe $\{y_{it}, x_{it}\}$ but not the other variables in the model. Like Lu and Su (2021), λ_i and γ_t characterize the individual and time heterogeneity of the slope coefficients, whereas η_i and ω_t has the usual individual and time fixed effects interpretation as both of them are allowed to be correlated with the regressor x_{it}. The central interest is the estimation and inference of $\{\beta, \lambda_i, \gamma_t\}$ under some weak conditions.

Lu and Su (2021) consider the estimation of the parameters in the slope coefficient, namely, β, λ_i and γ_t together with the parameters in the intercept term, i.e., η_i and ω_t jointly. Since they allow all both (λ_i, γ_t) and (η_i, ω_t) to be correlated with the regressor x_{it}, their call the estimates as a *generalized fixed effects* (GFE) estiamtes. The estimation requires the inversion of $(N + T - 1)\, K$-

dimension square matrix that can be computationally demanding when $(N+T-1)K$ is large or a bootstrap procedure is needed for the inference. In this paper, we consider a simple estimation procedure that does not require the inversion of a high dimensional matrix but some additional restrictive assumptions.

1.2.2 Estimation

Below we consider first the estimation of β, and then that of λ_i and γ_t.

To estimate β, we consider the two-way within-group transformation to eliminate the individual and time fixed effects η_i and ω_t in (1.2 .1) to obtain

$$\ddot{y}_{it} \equiv \ddot{x}_{it}'\beta + \ddot{u}_{it} + \ddot{\theta}_{it} + \ddot{\varphi}_{it}, \tag{1.2 .2}$$

where $\theta_{it} = x_{it}'\lambda_i$, $\varphi_{it} \equiv x_{it}'\gamma_t$, and, e.g., $\ddot{\theta}_{it} = \theta_{it} - \bar{\theta}_{i\cdot} - \bar{\theta}_{\cdot t} + \bar{\theta}_{\cdot\cdot}$. We will show that under some conditions, we can $\ddot{u}_{it} + \ddot{\theta}_{it} + \ddot{\varphi}_{it}$ as the composite errror terms in (1.2 .2) so that we can regress $\ddot{y}_{it}$ on $\ddot{x}_{it}$ to obtain the two-way FE estimate of β as follows:

$$\widehat{\beta} = \left(\frac{1}{NT}\sum_i\sum_t \ddot{x}_{it}\ddot{x}_{it}'\right)^{-1}\left[\frac{1}{NT}\sum_i\sum_t \ddot{x}_{it}\ddot{y}_{it}\right]. \tag{1.2 .3}$$

Obviously, $\widehat{\beta}$ is a consistent estimator under the key condition that $E\left[\ddot{x}_{it}\left(\ddot{u}_{it} + \ddot{\theta}_{it} + \ddot{\varphi}_{it}\right)\right] = 0.$

Given $\widehat{\beta}$ in (1.2 .3), let $\widehat{\ddot{y}}_{it} \equiv \ddot{y}_{it} - \ddot{x}_{it}'\widehat{\beta}$. Then using the fact that

$$\begin{aligned}
\ddot{\theta}_{it} &= (\theta_{it} - \bar{\theta}_{i\cdot}) - (\bar{\theta}_{\cdot t} - \bar{\theta}_{\cdot\cdot}) = (x_{it} - \bar{x}_{i\cdot})'\lambda_i - \frac{1}{N}\sum_l (x_{lt} - \bar{x}_{l\cdot})'\lambda_l \\
&= \tilde{x}_{it}'\left(\frac{N-1}{N}\lambda_i\right) - \frac{1}{N}\sum_{l\neq i}\tilde{x}_{lt}'\lambda_l
\end{aligned}$$

we have

$$\begin{aligned}\widehat{\ddot{y}}_{it} &= \ddot{x}_{it}'(\beta - \widehat{\beta}) + \ddot{u}_{it} + \ddot{\theta}_{it} + \ddot{\varphi}_{it} = \tilde{x}_{it}'\left(\frac{N-1}{N}\lambda_i\right) + e_{it} \\ &= \tilde{x}_{it}'\tilde{\lambda}_i + e_{it}, \end{aligned} \tag{I.2 .4}$$

where $\tilde{\lambda}_i \equiv \frac{N-1}{N}\lambda_i$ and $e_{it} \equiv \ddot{x}_{it}'(\beta - \widehat{\beta}) + \ddot{\varphi}_{it} + \ddot{u}_{it} - N^{-1}\sum_{l\neq i}\tilde{x}_{lt}'\lambda_l$. Based on (I.2 .4), we propose to estimate $\tilde{\lambda}_i$ by running the time series OLS regression of $\widehat{\ddot{y}}_{it}$ on $\tilde{x}_{it}$ for each i to obtain

$$\widehat{\tilde{\lambda}}_i = \left(\frac{1}{T}\sum_t \tilde{x}_{it}\tilde{x}_{it}'\right)^{-1}\frac{1}{T}\sum_t \tilde{x}_{it}\widehat{\ddot{y}}_{it}. \tag{I.2 .5}$$

Since $\widehat{\tilde{\lambda}}_i$ is estimating $\frac{N-1}{N}\lambda_i$, we propose to estimate λ_i by $\widehat{\lambda}_i = \frac{N}{N-1}\widehat{\tilde{\lambda}}_i$.

Similarly, noting that

$$\begin{aligned}\ddot{\varphi}_{it} &= (\varphi_{it} - \bar{\varphi}_{\cdot t}) - (\bar{\varphi}_{i\cdot} - \bar{\varphi}_{\cdot\cdot}) = (x_{it} - \bar{x}_{\cdot t})'\gamma_t - \frac{1}{T}\sum_r (x_{ir} - \bar{x}_{\cdot r})'\gamma_t \\ &= \breve{x}_{it}'\left(\frac{T-1}{T}\lambda_t\right) - \frac{1}{N}\sum_{r\neq t}\breve{x}_{ir}'\gamma_r,\end{aligned}$$

we have

$$\widehat{\ddot{y}}_{it} = \breve{x}_{it}'\left(\frac{T-1}{T}\gamma_t\right) + \nu_{it} = \breve{x}_{it}'\breve{\gamma}_t + \nu_{it}, \tag{I.2 .6}$$

where $\breve{\gamma}_t \equiv \frac{T-1}{T}\gamma_t$ and $\nu_{it} = \ddot{x}_{it}'(\beta - \widehat{\beta}) + \ddot{\theta}_{it} + \ddot{u}_{it} - T^{-1}\sum_{r\neq t}\breve{x}_{ir}'\gamma_r$. Based on (I.2 .6), we run the cross-section OLS regression of $\widehat{\ddot{y}}_{it}$ on $\breve{x}_{it}$ for each t to obtain

$$\widehat{\breve{\gamma}}_t = \left(\frac{1}{N}\sum_i \breve{x}_{it}\breve{x}_{it}'\right)^{-1}\frac{1}{N}\sum_i \breve{x}_{it}\widehat{\ddot{y}}_{it}. \tag{I.2 .7}$$

Since $\widehat{\breve{\gamma}}_t$ estimates $\frac{T-1}{T}\gamma_t$, the estimator of γ_t is given by $\widehat{\gamma}_t = \frac{T}{T-1}\widehat{\breve{\gamma}}_t$.

1.3 Asymptotic Results

In this section, we first present a set of basic assumptions that underly our asymptotic analysis. Then we establish the asymptotic properties of the estimators proposed in the last section.

1.3.1 Assumptions

Let $v_{it} = \{x_{it}^*, \lambda_i' x_{it}^*, x_{it}^{*\prime}\gamma_t\}$. Let $x_t = (x_{1t}, ..., x_{Nt})'$ and $X = (x_1', \ldots, x_T')'$. We make the following assumptions.

Assumption A1.1. *For every (i, j, t, s), the following holds:*

(i) The process $v_t = \{v_{1t}', ..., v_{Nt}'\}'$ is stationary and α-mixing, and for each i, let $\alpha_i(|t-s|)$ be the α -mixing coefficient between v_{it} and v_{is}, $\alpha_{ij}(|t-s|)$ be the α -mixing coefficient between v_{it} and v_{js} such that $\max\limits_{1\leq i\leq N} \alpha_i(|t-s|) < \alpha(|t-s|)$ and $\max\limits_{1\leq i,j\leq N} \alpha_{ij}(|t-s|) < \alpha(|t-s|)$ uniformly such that it satisfies $\alpha(\tau) \leq M\tau^{-\kappa}$ for some $C > 0$ and $\kappa > 0$, where κ is some positive constant that depends on $\delta_1 > 0$ and $\delta_2 > 0$ such that $\max\limits_i \sum_{j=1}^N \sum_{\tau=1}^T \alpha_{ij}(\tau)^{\delta_1/(4+\delta_1)} = O(1)$ and $\max\limits_i \sum_{j=1}^N \alpha_{ij}(0)^{\delta_1/(4+\delta_1)} = O(1)$, for the same δ_1, $E\|v_{it}\|^{4+\delta_1+\epsilon_1} < \infty$ uniformly in i and t, where ϵ_1 is positive and can be small enough. Furthermore, assume that $E\|\sum_i v_{it}\|^q = O\left(N^{q/2}\right)$ for some $q > 2$.

(ii) $E(x_{it}) = \mu$, $Var(x_{it}) = \Sigma_x$ such that $m < s_{min}(\Sigma_x) \leq s_{max}(\Sigma_x) < M$, and $Cov(x_{it}, x_{is}) = \Gamma_{t-s}$.

(iii) (a) $E(\lambda_i \mid X) = 0$, (b) $Var\left(N^{-1/2}\sum_i \lambda_i\right) = O(1)$,and for some positive constant m, $E\|\lambda_i\|_2^{2+m} < \infty$ for each i.

(iv) (a) $E(\gamma_t \mid X) = 0$, (b) $Var\left(T^{-1/2}\sum_t \gamma_t\right) = O(1)$, and for some positive constant m, $E\|\gamma_t\|_2^{2+m} < \infty$ for each t.

(v) $Cov(\lambda_j, \gamma_s \mid X) = 0$ holds for any (j, s).

Assumption A1.2. *For every (i, j, t, s), the following holds:*

(i) The process $u_t = \{u_{1t}, ..., u_{Nt}\}^T$ is stationary and α-mixing with zero mean, and $\alpha_{u,i}(|t-s|)$ be the α-mixing coefficient between u_{it} and u_{is} such that $\max_{1\le i\le N} \alpha_{u,i}(|t-s|) < \alpha(|t-s|)$ and $\max_{1\le i,j\le N} \alpha_{u,ij}(|t-s|) < \alpha(|t-s|)$, where $\alpha(.)$ is the same one defined in Assumption A1.1 above. And for the same $\delta_2 > 0$ in Assumption A1.1 above, $E\|u_{it}\|^{4+\delta_2+\epsilon_2} < \infty$ uniformly in i and t, where ϵ_2 is positive and can be small enough. Furthermore, assume that $E\|\sum_i u_{it}\|^q = O\left(N^{q/2}\right)$ for some $q > 2$.

(ii) $E\left(u_{it} \mid x_{js}, \lambda_j, \gamma_s\right) = 0$ holds for all (i, j, t, s).

Assumption A1.3. *As (N, T) go to infinity simultaneously, then $N/T^2 \to 0$ and $T/N^2 \to 0$.*

Assumption A1.4. *There exist constants $m > 0$ small enough and $M > 0$ large enough such that:*
(i) $m < s_{min}\left(\frac{1}{NT^2}\sum_{i,j=1}^N \sum_{t,s=1}^T E\mathcal{G}_{it}\mathcal{G}_{js}^{*\prime}\right) \le s_{max}\left(\frac{1}{NT^2}\sum_{i,j=1}^N \sum_{t,s=1}^T E\mathcal{G}_{it}^{\lambda}\mathcal{G}_{js}^{\lambda\,*\prime}\right) < M$ holds with $\mathcal{G}_{it} \equiv x_{it}^* x_{it}^{*\prime}\lambda_i, x_{it}^* x_{it}^{*\prime}\gamma_t$, or $x_{it}^* u_{it}$ and $\mathcal{G}_{js}$ is similarly defined.
(ii) $m < \min_{1\le t\le T} s_{min}\left(\frac{1}{N}\sum_{i,j=1}^N E\mathbb{G}_{it}\mathbb{G}_{jt}'\right) \le \max_{1\le t\le T} s_{max}\left(\frac{1}{N}\sum_{i,j=1}^N E\mathbb{G}_{it}\mathbb{G}_{jt}'\right) < M$, where $\mathbb{G}_{it} \equiv \left(x_{it}^* x_{it}^{*\prime} - \Sigma_x\right)\lambda_i$ or $x_{it}^* u_{it}$, and $\mathbb{G}_{jt}$ is similarly defined.
(iii) $m < \min_{1\le i\le N} s_{min}\left(\frac{1}{T}\sum_{t,s=1}^T E\mathbb{G}_{it}^*\mathbb{G}_{is}^{*\prime}\right) \le \max_{1\le i\le N} s_{max}\left(\frac{1}{T}\sum_{t,s=1}^T E\mathbb{G}_{it}^*\mathbb{G}_{is}^{*\prime}\right) < M$, where $\mathbb{G}_{it}^* \equiv \left(x_{it}^* x_{it}^{*\prime} - \Sigma_x\right)\gamma_t$ or $x_{it}^* u_{it}$, $\mathbb{G}_{is}^*$ is similarly defined.

Assumption A1.5. *Suppose following CLT-type results hold,*

(i) $N^{-1/2}T^{-1}\sum_{i=1}^N \sum_{t=1}^T x_{it}^* x_{it}^{*\prime}\lambda_i \xrightarrow{d} \mathcal{N}(0, V_\lambda)$, where $V_\lambda = p\lim_{N,T\to\infty} N^{-1}T^{-2}\sum_{i,j=1}^N \sum_{t,s=1}^T \mathcal{G}_{it}^{\lambda}\mathcal{G}_{js}^{\lambda\,\prime}$ with $\mathcal{G}_{it}^{\lambda} = x_{it}^* x_{it}^{*\prime}\lambda_i$, and $\mathcal{G}_{jt}^{\lambda}$ is similarly defined.

(ii) $N^{-1/2}\sum_{i=1}^N \left(x_{it}^* x_{it}^{*\prime} - \Sigma_x\right)\lambda_i \xrightarrow{d} \mathcal{N}(0, \Sigma_{t,x\lambda})$, where $\Sigma_{t,x\lambda} = p\lim_{N\to\infty} N^{-1}\sum_{i,j=1}^N \mathbb{G}_{it}^{\lambda}\mathbb{G}_{jt}^{\lambda\,\prime}$ for all t with $\mathbb{G}_{it}^{\lambda} = \left(x_{it}^* x_{it}^{*\prime} - \Sigma_x\right)\lambda_i$ and $\mathbb{G}_{jt}^{\lambda}$ is similarly defined.

(iii) $N^{-1/2}\sum_{i=1}^N x_{it}^* u_{it} \xrightarrow{d} \mathcal{N}\left(0, \breve{\Sigma}_{t,xu}\right)$, where $\breve{\Sigma}_{t,xu} = p\lim_{N\to\infty} N^{-1}\sum_{i,j=1}^N \mathbb{G}_{it}^u \mathcal{G}_{jt}^{u\,\prime}$ for all t with $\mathbb{G}_{it}^u = x_{it}^* u_{it}$ and $\mathbb{G}_{jt}^u$ is similarly defined.

Assumption A1.6. *(i) Let η_i be random variables that have mean 0 and finite variance, and uncorrelated with u_{it} for any $1 \le i \le N$. (ii) Let ω_t be random variables that have mean 0 and finite variance, and uncorrelated with u_{it} for any $1 \le t \le T$.*

Assumption A1.1 requires that the sequences $\{v_{it}\}$ are stationary and strong mixing over t and also imposes the mild assumptions for the random coefficients, which is standard in related studies. For example, Assumption A1.1(i) incorporates Assumption A.4 in Feng et al. (2017) and Assumption A4 in Chen et al. (2012) and extents them to our setups; Assumption A1.1(iii)-(v) extents Assumption 2.2 to 2.8 Hsiao (1974) to our setups. Under Assumptions A1.1(iii)(a) and (v)(a), heterogeneous and time-varying slopes $\{\lambda_i\}_{i=1}^N$ and $\{\gamma_t\}_{t=1}^T$ can correlated to x_{it} in higher moments. Assumptions A1.1(iii)(b) and (v)(b) allow for the correlations among the sequences $\{\lambda_i\}_{i=1}^N$ and $\{\gamma_t\}_{t=1}^T$ correspondingly. Assumption A1.2(i) imposes similar mixing conditions on $\{u_{it}\}$ as Assumption A1.1(i). The strict exogeneity condition in Assumption A1.2(ii) simplifies the the asymptotic analysis. It allows for conditional heteroskedasticity, skewness, or kurtosis of unknown form in u_{it}. Assumption A1.3 indicates the rate conditions, which accommodates the case that N and T are comparable. Assumption A1.4 assumes that the eigenvalues of referred matrices are bounded from below and above, which is a technical assumption that is standard in related studies, see e.g., Assumption 2(iii) in Zhu (2017). Assumption A1.5 is frequently used in related studies and can be replaced by imposing some primitive assumptions; see e.g., Kuersteiner and Prucha (2013), Castagnetti et al. (2015), and Hidalgo and Schafgans (2017). Assumption A1.6 imposes very mild restrictions on the fixed effects..

Remark 1.1. Assumption A1.1 (ii) assumes the homoskedasticity of x_{it}. This condition is just imposed to simplify the exposition of the results. From the perspective of theoretical justifications, all results in Section 1.3 and Section 1.4 still pertain at the cost of more lengthy proofs and does not change substantially by defining corresponding notations appropriately. We present the discussions on the relaxation of such conditions in the Appendix A such that $Ex_{it} = \mu_i$, $Var\left(x_{it}\right) = \Sigma_{i,x}$ and $Cov\left(x_{it}, x_{is}\right) = \Gamma_{i,t-s}$. For details, we refer readers to the Appendix A.

Remark 1.2. The strict exogeneity condition in Assumption A1.2 (ii) can be replaced by the sequential exogeneity condition such that $E\left(u_{it} \mid x_{it}, \lambda_j, \gamma_s\right) = 0$ for all (i, j, t, s) as Assumption A3(ii) in Lu and Su (2021). In the presence of sequential exogenous but not strictly exogenous regressors, $E\left(x_{it}^* u_{js}\right) \neq 0$ in general for $i \neq j$ or $t \neq s$, which can lead to the asymptotic bias. To handle the bias caused by the sequential exogeneity condition on x_{it}, we can employ the half-panel jackkinfe estimation proposed by Dhaene and Jochmans (2015), see e.g., Chudik et al. (2018) and Lu and Su (2021).

1.3.2 Main results on β

In this subsection, we study the asympotic properties of $\widehat{\beta}$. To present the main results, define

$$
\begin{aligned}
V_\lambda &\equiv \lim_{(N,T)\to\infty} \frac{1}{NT^2} \sum_{i,j} \sum_{t,s} E\left(x_{it}^* x_{it}^{*\prime} \lambda_i \lambda_j^\prime x_{js}^* x_{js}^{*\prime}\right), \\
V_\gamma &\equiv \lim_{(N,T)\to\infty} \frac{1}{N^2T} \sum_{i,j} \sum_{t,s} E\left(x_{it}^* x_{it}^{*\prime} \gamma_t \gamma_s^\prime x_{js}^* x_{js}^{*\prime}\right), \\
V_u &\equiv \lim_{(N,T)\to\infty} \frac{1}{NT} \sum_{i,j} \sum_{t,s} E\left(x_{it}^* u_{it} u_{is} x_{jt}^{*\prime}\right),
\end{aligned}
$$

The key result is summarized in the following theorem.

Theorem 1.3.1. *Suppose Assumptions A1.1-A1.5. Suppose that V_λ, V_γ, and V_u exist and are nonsingular. Then as $(N,T)\to\infty$,*

(i) $\|\widehat{\beta}-\beta\| = O_p\left(\delta_{NT}^{-1}\right)$;

(ii) $\mathcal{V}_\beta^{-1/2}\left(\widehat{\beta}-\beta\right) \xrightarrow{d} \mathcal{N}(0_K, I_K)$,

where $\mathcal{V}_\beta = \frac{1}{N}\Sigma_x^{-1} V_\lambda \Sigma_x^{-1} + \frac{1}{T}\Sigma_x^{-1} V_\gamma \Sigma_x^{-1} + \frac{1}{NT}\Sigma_x^{-1} V_u \Sigma_x^{-1}$.

Theorem 1.3.1(i) reports the convergence rate of $\widehat{\beta}$. In comparison with the usual $\sqrt{NT}$-rate of convergence for the slope estimators in homogenous panels, our estimator $\widehat{\beta}$ converges to β at the slow rate $\sqrt{N} \wedge \sqrt{T}$ when the slope coefficient has heterogeneity along both the time and indiviual dimensions. Theorem 1.3.1(ii) estabilishes the asymptotic normality of $\widehat{\beta}$. The asymptotic variance $\mathcal{V}_\beta$, contains three components, that can be attributed to the individul heterogeneity in the slope coefficient, the time heterogeneity in the slope coefficient, and the idiosyncractic error term in the model, respectively. We both the individual and time heterogeneities are absent from the slope coefficient, the first two components are vanishing so that the convergence rate of $\widehat{\beta}$ becomes the standard $(NT)^{-1/2}$-rate for two-way FE estimators of the slope cofficients in homogeneous panels.

Note that x^*_{it}, λ_i and γ_t are not observed in the reality, to implement the inference on β, we replace x^*_{it}, λ_i and γ_t with their consistent estimators. The corollary below shows the result:

Corollary 1.3.2. *Under Assumptions A1.1-A1.4, let $\widehat{\lambda}_i$ be any consistent estimator of λ_i for each i, $\widehat{\gamma}_t$ be any consistent estimator of γ_t for each t, and $\widehat{u}_{it}$ be any consistent estimator of u_{it}, as (N, T) go to infinity jointly,*

$$\widehat{\mathcal{V}}_\beta^{-1/2}\left(\widehat{\beta}-\beta\right) \xrightarrow{d} \mathcal{N}(0_K, I_K),$$

where $\widehat{\mathcal{V}}_\beta = \frac{1}{N}\widehat{Q}_{\ddot{x}}^{-1}\widehat{V}_\lambda\widehat{Q}_{\ddot{x}}^{-1} + \frac{1}{T}\widehat{Q}_{\ddot{x}}^{-1}\widehat{V}_\gamma\widehat{Q}_{\ddot{x}}^{-1} + \frac{1}{NT}\widehat{Q}_{\ddot{x}}^{-1}\widehat{V}_u\widehat{Q}_{\ddot{x}}^{-1}$, *and* $\widehat{Q}_{\ddot{x}} = \frac{1}{NT}\sum_i\sum_t \ddot{x}_{it}\ddot{x}_{it}'$,

$$\widehat{V}_\lambda = \frac{1}{NT^2}\sum_i\sum_j\sum_t\sum_{k=-p}^{p}\left(1-\frac{|k|}{|p|+1}\right)\ddot{x}_{it}\ddot{x}_{it}'\widehat{\lambda}_i\widehat{\lambda}_j'\ddot{x}_{jt+k}\ddot{x}_{jt+k}',$$

$$\widehat{V}_\gamma = \frac{1}{NT^2}\sum_i\sum_j\sum_t\sum_{k=-p}^{p}\left(1-\frac{|k|}{|p|+1}\right)\ddot{x}_{it}\ddot{x}_{it}\widehat{\gamma}_t\widehat{\gamma}_{t+k}'\ddot{x}_{jt}\ddot{x}_{jt+k}',$$

$$\widehat{V}_u = \frac{1}{NT^2}\sum_i\sum_j\sum_t\sum_{k=-p}^{p}\left(1-\frac{|k|}{|p|+1}\right)\ddot{x}_{it}\widehat{u}_{it}\widehat{u}_{it+k}\ddot{x}_{jt+k}'.$$

In order to handle the serial correlation, we use HAC estimators with the Bartlett kernel in Corollary 1.3.2 to implement $\mathcal{V}_\beta$, and p above is the truncation parameter that is user-specified. The choice of kernel can be different in practice. Note that we account for weak cross-sectional dependence under Assumption A1.8 in the current paper. However, if the strong cross-sectional dependence is considered, one can also incorporate the idea of spatial HAC estimator in Kelejian and Prucha (2007) to modify $\widehat{V}_\lambda$, $\widehat{V}_\gamma$ and $\widehat{V}_u$ given above.

1.3.3 Main results on λ_i and γ_t

In order to obtain unbiased and $\sqrt{T}$-consistent estimator of λ_i for each given i, we proposed following feasible bias-corrected estimates for $\{\lambda_i\}_{i=1}^N$,

$$\widehat{\lambda}_i = \frac{N}{N-1}\widehat{\widehat{\lambda}}_i, \tag{1.3.1}$$

where $\widehat{\widehat{\lambda}}_i$ is defined in (1.2 .5). The following theorem gives the asymptotic distribution of $\widehat{\lambda}_i$ in (1.3 .1).

Theorem 1.3.3. *Under Assumptions A1.1-A1.5, as (N,T) go to infinity jointly, for each i,*

$$\mathcal{V}_{\lambda i}^{-1/2}\left(\widehat{\lambda}_i - \lambda_i\right) \xrightarrow{d} \mathcal{N}(0_K, I_K),$$

where $\mathcal{V}_{\lambda i} = \Sigma_x^{-1}\left(V_\lambda + \Omega_{\lambda i}\right)\Sigma_x^{-1}$, V_λ is the element of covariance matrix of $\widehat{\beta}$ defined in Theorem 1.3.1 above, and

$$\Omega_{\lambda i} = \frac{1}{T}\Sigma_{i,x\gamma} + \frac{1}{T}\tilde{\Sigma}_{i,xu},$$

where the matrices Σ_x^{-1}, $\Sigma_{i,x\gamma} = \lim_{T\to\infty}\frac{1}{T}\sum_{t,s=1}^{T} E\left[\left(x_{it}^{*}x_{it}^{*\prime} - \Sigma_x\right)\gamma_t\gamma_s'\left(x_{is}^{*}x_{is}^{*\prime} - \Sigma_x\right)\right]$, and $\tilde{\Sigma}_{i,xu} = \lim_{T\to\infty}\frac{1}{T}\sum_{t,s=1}^{T} E\left(x_{it}^{*}u_{it}u_{is}x_{is}^{*\prime}\right)$ exist and are positive definite matrix.

Remark 1.3. It is worthwhile to observe that for each i, the structure of $\mathcal{V}_{\lambda i}$ involves three components, namely, V_λ, $\Sigma_{i,x\gamma}$ and $\tilde{\Sigma}_{i,xu}$ in the above formula. In particular, the existence of $\Sigma_{i,x\gamma}$ in the above formula exhibit the effects of time-varying slopes, and the existence of V_λ in the above formula exhibit the effects of heterogeneous slopes $\{\lambda_j\}_{j\neq i}$ on the efficiency of the estimator $\widehat{\lambda}_i$ for each i.

Similarly, in order to obtain unbiased and $\sqrt{N}$-consistent estimator of γ_t for each given t, we proposed following bias-corrected estimate $\widehat{\gamma}_t$ for each t,

$$\widehat{\gamma}_t = \frac{T}{T-1}\widehat{\widehat{\gamma}}_t, \tag{1.3 .2}$$

where $\widehat{\widehat{\gamma}}_t$ is defined in (1.2 .7). The following theorem gives the asymptotic distribution of $\widehat{\gamma}_t$:

Theorem 1.3.4. *Under Assumptions A1.1-A1.5, as (N,T) go to infinity jointly, for each t,*

$$\mathcal{V}_{\gamma t}^{-1/2}\left(\widehat{\gamma}_t - \gamma_t\right) \xrightarrow{d} \mathcal{N}(0_K, I_K),$$

where $\mathcal{V}_{\gamma t} = \Sigma_x^{-1}\left(V_\gamma + \Omega_{\gamma t}\right)\Sigma_x^{-1}$, V_γ is the element of covariance matrix of $\widehat{\beta}$ defined in Theorem

1.3.1 above, and

$$\Omega_{\gamma t} = \frac{1}{N}\Sigma_{t,x\lambda} + \frac{1}{N}\breve{\Sigma}_{t,xu},$$

where the matrices Σ_x^{-1}, $\Sigma_{t,x\lambda} = \lim\limits_{N\to\infty} \frac{1}{N}\sum_{i,j} E\left[\left(x_{it}^* x_{it}^{*\prime} - \Sigma_x\right)\lambda_i\lambda_j'\left(x_{jt}^* x_{jt}^{*\prime} - \Sigma_x\right)\right]$, *and* $\breve{\Sigma}_{t,xu} = \lim\limits_{N\to\infty} \frac{1}{N}\sum_{i,j} E\left(x_{it}^* u_{it} u_{jt} x_{jt}^{*\prime}\right)$ *exist and are positive definite matrix.*

Remark 1.4. Similar to Remark 1.3, the presence of the term $\Sigma_{t,x\lambda}$ in the structure of $\mathcal{V}_{\gamma t}$ demonstrates the effects of heterogeneous slopes, and the existence of V_γ reveals the effects of time-varying slopes $\{\gamma_s\}_{s\neq t}$ on the efficiency of the estimator $\widehat{\gamma}_t$.

Note that x_{it}^*, u_{it}, λ_i and γ_t in Theorem 1.3.3 and Theorem 1.3.4 are not available in practice. To implement the inference on $\{\lambda_i\}_{i=1}^N$ and $\{\gamma_t\}_{t=1}^T$, we have to replace x_{it}^*, u_{it}, λ_i and γ_t with their consistent estimates, which leads us to the following two corollaries,

Corollary 1.3.5. *Under Assumptions A1.1-A1.5, as (N,T) go to* (∞,∞) *jointly, for each* i,

$$\widehat{\mathcal{V}}_{\lambda i}^{-1/2}\left(\widehat{\lambda}_i - \lambda_i\right) \xrightarrow{d} \mathcal{N}(0_K, I_K),$$

where $\widehat{\mathcal{V}}_{\lambda i} = \widehat{Q}_{\tilde{x}}^{-1}\left(\widehat{V}_\lambda + \widehat{\Omega}_{\lambda i}\right)\widehat{Q}_{\tilde{x}}^{-1}$, $\widehat{V}_\lambda$ *is an consistent estimator of covariance matrix of* V_λ *defined in Corollary 1.3.2,* $\widehat{Q}_{\tilde{x}} = \frac{1}{T}\sum_t \tilde{x}_{it}\tilde{x}_{it}'$, *and*

$$\widehat{\Omega}_{\lambda i} = \frac{1}{T}\widehat{\Sigma}_{i,x\gamma} + \frac{1}{T}\widehat{\tilde{\Sigma}}_{i,xu},$$

where $\widehat{\Sigma}_{i,x\gamma} = \frac{1}{T}\sum_{t=1}^T\sum_{k=-p}^p\left(1 - \frac{|k|}{|p|+1}\right)\left(\tilde{x}_{it}\tilde{x}_{it}' - \widehat{\tilde{\Sigma}}_X\right)\widehat{\gamma}_t\widehat{\gamma}_{t+k}'\left(\tilde{x}_{it+k}\tilde{x}_{it+k}' - \widehat{\tilde{\Sigma}}_X\right)'$, $\widehat{\gamma}_t$ *is any consistent estimator of* γ_t, *and* $\widehat{\tilde{\Sigma}}_{i,xu} = \frac{1}{T}\sum_{t=1}^T\sum_{k=-p}^p\left(1 - \frac{|k|}{|p|+1}\right)\tilde{x}_{it}\widehat{\tilde{u}}_{it}\widehat{\tilde{u}}_{it+k}\tilde{x}_{it+k}'$, *where* $\widehat{\tilde{\Sigma}}_X = \frac{1}{T}\sum_t \tilde{x}_{it}\tilde{x}_{it}'$, $\widehat{\tilde{u}}_{it} = \widehat{u}_{it} - \frac{1}{T}\sum_{s=1}^T \widehat{u}_{is}$, $\widehat{u}_{it}$ *is any consistent estimator of* u_{it}.

Remark 1.5. In order to allow for the serial correlations, we use HAC estimator of $\tilde{\Sigma}_{i,xu}$ with the Bartlett kernel in Corollary 1.3.5 to implement $\tilde{\Sigma}_{i,xu}$, and p is the user-specified truncation parameter.

Corollary 1.3.6. *Under Assumptions A1.1-A1.5, as (N,T) go to* (∞, ∞) *jointly, for each* t*,*

$$\widehat{\mathcal{V}}_{\gamma t}^{-1/2}\left(\widehat{\gamma}_t - \gamma_t\right) \xrightarrow{d} \mathcal{N}(0_K, I_K),$$

where $\widehat{\mathcal{V}}_{\gamma t} = \widehat{Q}_{\breve{x}}^{-1}\left(\widehat{V}_\gamma + \widehat{\Omega}_{\gamma t}\right)\widehat{Q}_{\breve{x}}^{-1}$*,* $\widehat{V}_\gamma$ *is an consistent estimator of covariance matrix of* V_γ *defined in Corollary 1.3.2,* $\widehat{Q}_{\breve{x}} = \frac{1}{N}\sum_i \breve{x}_{it}\breve{x}_{it}'$*, and*

$$\widehat{\Omega}_{\gamma t} = \frac{1}{N}\widehat{\Sigma}_{t,x\lambda} + \frac{1}{N}\widehat{\breve{\Sigma}}_{t,xu},$$

where $\widehat{\Sigma}_{t,x\lambda} = \frac{1}{N}\sum_{i,j}\left(\breve{x}_{it}\breve{x}_{it}' - \widehat{\breve{\Sigma}}_X\right)\widehat{\lambda}_i\widehat{\lambda}_j'\left(\breve{x}_{jt}\breve{x}_{jt}' - \widehat{\breve{\Sigma}}_X\right)'$*,* $\widehat{\lambda}_i$ *is any consistent estimator of* λ_i*, and* $\widehat{\breve{\Sigma}}_{t,xu} = \frac{1}{N}\sum_{i,j}\breve{x}_{it}\widehat{\breve{u}}_{it}\widehat{\breve{u}}_{jt}\breve{x}_{jt}'$*, where* $\widehat{\breve{u}}_{it} = \widehat{u}_{it} - \frac{1}{N}\sum_{j=1}^N \widehat{u}_{jt}$*,* $\widehat{\breve{\Sigma}}_X = \frac{1}{N}\sum_i \breve{x}_{it}\breve{x}_{it}'$*, and* $\widehat{u}_{it}$ *is any consistent estimator of* u_{it} *for each* t*.*

Remark 1.6. As point out earlier, if the strong cross-sectional dependence is considered, one can also incorporate the idea of spatial HAC estimator in Kelejian and Prucha (2007) to modify $\widehat{\Sigma}_{t,x\lambda}$, $\widehat{V}_\gamma$ and $\widehat{\breve{\Sigma}}_{t,xu}$ given above.

1.3.4 The two-way heterogeneous bias test

According to the analyses in the Appendix A, we can see that the consistent and unbiased estimates of β, $\{\lambda_i\}_{i=1}^N$, and $\{\gamma_t\}_{t=1}^T$ crucially depend on Assumption A1.1(iii)(a), (iv)(a) and Assumption A1.2(ii). As pointed out in Remark 1.2, if the strict exogeneity condition in Assumption A1.2(ii) is relaxed, the asymptotic bias can be removed by employing the half-panel jackknife method.

As for Assumption A1.1(iii)(a) and (iv)(a), they imply following two conditions will hold simultaneously,

$$\sum_i E\left(\breve{x}_{it}^*\tilde{x}_{it}^{*\prime}\lambda_i\right) = 0 \quad \text{holds for each } t. \tag{1.3 .3}$$

and

$$\sum_t E\left(\tilde{x}_{it}^{*}\breve{x}_{it}^{*\prime}\gamma_t\right) = 0 \quad \text{holds for each } i. \tag{I.3 .4}$$

Indeed, the validity of (I.3 .4) and (I.3 .3) indicate that the estimate of β based on (I.2 .2), the estimates of $\{\lambda_i\}_{i=1}^{N}$ based on (I.2 .4) and the estimate of $\{\gamma_t\}_{t=1}^{T}$ based on (I.2 .6) still can be desirable even if the slope heterogeneity and time-varyingness exist. Naturally, it is of practical and theoretical interests to check the validity of (I.3 .4) and (I.3 .3). So the rest of this subsection is devoted to test whether (I.3 .4) and (I.3 .3) holds simultaneously or not.

To test the null of hypothesis that (I.3 .3) and (I.3 .4) hold simultaneously, we propose the following test statistic,

$$\mathcal{J}_{TWHB} = \widehat{\mathcal{V}}_{\upsilon}^{-1/2}\frac{1}{\sqrt{T}}\sum_t \left(\frac{1}{N}\widehat{\upsilon}_t^{\prime}\widehat{\upsilon}_t - \widehat{\mathcal{B}}_{\upsilon}\right), \tag{I.3 .5}$$

where $\widehat{\upsilon}_t = \sum_i \breve{x}_{it}\tilde{x}_{it}^{\prime}\widehat{\lambda}_i$, $\widehat{\mathcal{B}}_{\upsilon} = \frac{1}{T}\sum_t \widehat{\mathcal{Z}}_{t,\upsilon}$, $\widehat{\mathcal{V}}_{\upsilon} = \frac{1}{T}\sum_{t,s}\widehat{\zeta}_{t,\upsilon}^{\prime}\widehat{\zeta}_{s,\upsilon}$, and we define $\widehat{\mathcal{Z}}_{t,\upsilon} \equiv \frac{1}{N}\sum_{i,j}\widehat{\lambda}_i^{\prime}\left(\breve{x}_{it}\breve{x}_{it}^{\prime} - \widehat{\breve{\Sigma}}_X\right)^{\prime}\left(\breve{x}_{jt}\breve{x}_{jt}^{\prime} - \widehat{\breve{\Sigma}}_X\right)\widehat{\lambda}_j$, $\widehat{\zeta}_{t,\upsilon} \equiv \widehat{\mathcal{Z}}_{t,\upsilon} - \widehat{\mathcal{B}}_{\upsilon}$.

Recently, Campello et al. (2019) propose the one-way heterogeneity bias test when the slope heterogeneity exits in panel data models. As a comparison, our two-way heterogeneity bias test aims to check the bias possibly induced by the existence of slope heterogeneity or time-varyingness and thus generalizes their test.

The motivation behind $\mathcal{J}_{TWHB}$ is simple and intuitive. Denote $\upsilon_t \equiv \sum_i \breve{x}_{it}^{*}\tilde{x}_{it}^{*\prime}\lambda_i$, and $\upsilon_i \equiv \sum_t \tilde{x}_{it}^{*}\breve{x}_{it}^{*\prime}\gamma_t$. If x_{it}^{*}, λ_i and γ_t in (I.3 .3) and (I.3 .4) are available, ones can show that $\mathcal{J}_{TWHB}^{\star} \equiv N^{-1}T^{-1/2}\sum_t \upsilon_t^{\prime}\upsilon_t + N^{-1/2}T^{-1}\sum_i \upsilon_i^{\prime}\upsilon_i \equiv \mathcal{J}_{TWHB,\lambda}^{\star} + \mathcal{J}_{TWHB,\gamma}^{\star}$ will follow the standard normal distribution after being centered and rescaled appropriately and thus can be used to test whether (I.3 .3) and (I.3 .4) holds simultaneously or not. Our starting point is to implement $\mathcal{J}_{TWHB,\lambda}^{\star} = N^{-1}T^{-1/2}\upsilon_t^{\prime}\upsilon_t$, and it is the common practice to replace υ_t by its plug-in estimator $\widehat{\upsilon}_t$ based on $\{\widehat{\lambda}_i\}_{i=1}^{N}$ and x_{it}. Meanwhile, the unbiasedness and consistency of $\widehat{\lambda}_i$ holds as long as (I.3 .3) and (I.3 .4) holds simultaneously. Therefore, it suffices to use the proposed test $\mathcal{J}_{TWHB}$ instead.

Before we present the asymptotic null distribution of $\mathcal{J}_{TWHB}$, we impose additional assumption as below.

Assumption A1.7. *$E\mathcal{Z}_{t,v}$ is a constant, where $\mathcal{Z}_{t,v} \equiv \frac{1}{N}\sum_i \sum_j \lambda_i' \left(x_{it}^* x_{it}^{*\prime} - \Sigma_x\right)\left(x_{jt}^* x_{it}^{*\prime} - \Sigma_x\right)\lambda_j$.*

Assumption A1.7 requires $E\mathcal{Z}_{t,v}$ does not vary with time t such that $\widehat{\mathcal{B}}_v$ can be used to estimate this population mean consistently. This condition simplifies our theoretical analysis. On the contrary, if $E\mathcal{Z}_{t,v}$ is allowed to vary with time t, ones can employ jackknife method by making full use of cross-section observations for each given t to estimate $E\mathcal{Z}_{t,v}$ consistently for each t.

The following theorem states the asymptotic null distribution of $\mathcal{J}_{TWHB}$.

Theorem 1.3.7. *Under Assumptions A1.1-A1.5, in addition, A1.1(iii)(a) and (iv)(a) are replaced by* (1.3 .4) *and* (1.3 .3) *respectively, as (N,T) go to infinity jointly, $N/T^{3/2} \to 0$, then the following hodls,*

$$\mathcal{J}_{TWHB} \xrightarrow{d} \mathcal{N}(0,1)$$

In practice, ones can implement $\mathcal{J}_{TWHB}$ to make sure the desirability of $\widehat{\beta}$, $\{\widehat{\lambda}_i\}_{i=1}^N$ and $\{\widehat{\gamma}_t\}_{t=1}^T$ based on our estimation strategy. For the asymptotic local power properties of $\mathcal{J}_{TWHB}$, we can follow Su and Chen (2013) and Campello et al. (2019) to show that $\mathcal{J}_{TWHB}$ has the non-trivial power against the local alternatives as long as (1.3 .4) or (1.3 .3) is violated. Besides, as pointed out above, $\mathcal{J}_{TWHB}$ stems from $\mathcal{J}^{\star}_{TWHB,\lambda}$, we can similarly devise a test statistic related to $\mathcal{J}^{\star}_{TWHB,\gamma}$, and the result and corresponding theoretical justifications are very similar to those for $\mathcal{J}_{TWHB}$, which is also convenient to implement in practice.

1.3.5 Uniform consistency on estimates of time-varying and heterogeneous slopes

In this subsection, we establish the uniform consistency to quantify the upper bounds on the maximum deviation of the estimates from corresponding true parameters by imposing an additional assumption.

Specifically, we extent Assumption 1(i)-(ii) in Zhu (2017) to strengthen Assumptions A1.1 and A1.2 to specify the behaviors of the entries of $\{\lambda_i, \gamma_t, x_{it}, u_{it}\}$ for both individual and time dimensions further.

Definition 1. *A random variable Z is said to have an exponential-type tail with parameter (b, ν) if $\forall z > 0, P(|Z| > z) \leq \exp[1 - (z/b)^{\nu}]$.*

Assumption A1.8. *Assume that the following hold:*

(i) There exist constants $b_*, \nu_* > 0$ such that $\forall(i,t) \in \{1,\ldots,n\} \times \{1,\ldots,T\}$, each entry of λ_i, γ_t, x_{it} and $u_{i,t}$ has an exponential-type tail with parameter (b_*, ν_*);

(ii) There exist constants $c_*, \nu_{**} > 0$ such that $\alpha(\tau) \leq c_* \exp(-\tau^{\nu_{**}}) \,\forall\tau \geq 1$. For the same (c_*, ν_{**}), assume that $\max_i \sum_{j\in\mathcal{S}_N} \alpha_{ij}(0) \leq c_* \exp(-Card(\mathcal{S}_N)^{\nu_{**}})$ and $\max_i \sum_{j\in\widetilde{\mathcal{S}}_N} \alpha_{u,ij}(0) \leq c_* \exp\left(-Card(\widetilde{\mathcal{S}_N})^{\nu_{**}}\right)$, where $Card(\mathcal{S}_N)$ and $Card(\widetilde{\mathcal{S}_N})$ go to infinity as N goes to infinity, $\mathcal{S}_N$ and $\widetilde{\mathcal{S}}_N$ are subsets of $\{1,\ldots,N\}$, and $Card(A)$ stands for the cardinality of a set A; $\alpha(.)$, $\alpha_{ij}(.)$, and $\alpha_{u,ij}(.)$ are defined in Assumptions A1.1 and A1.2.

Compared with Assumption 1(ii) in Zhu (2017), we impose the additional restriction that the cross-sectional dependence for v_{it} and u_{it} in Assumptions A1.1 and A1.2 decays at the exponential rate in Assumption A1.8(ii) above, and this condition will guarantee the use of Theorem 1 in Merlevède et al. (2011) in our proofs.

Proposition 1.3.8. *Under Assumptions A1.1-A1.4, and A1.8*

(1) $\|\widehat{\lambda} - \lambda\|_\infty = O_p\left(\sqrt{\frac{lnN}{T}}\right)$, *where* $\lambda \equiv (\lambda_1', .., \lambda_N')' \in \mathbb{R}^{NK}$, $\widehat{\lambda} \equiv \left(\widehat{\lambda}_1', .., \widehat{\lambda}_N'\right)' \in \mathbb{R}^{NK}$.

(2) $\|\widehat{\gamma} - \gamma\|_\infty = O_p\left(\sqrt{\frac{lnT}{N}}\right)$, *where* $\gamma \equiv (\gamma_1', .., \gamma_T')' \in \mathbb{R}^{TK}$, $\widehat{\gamma} \equiv (\widehat{\gamma}_1', .., \widehat{\gamma}_T')' \in \mathbb{R}^{TK}$.

This result says that $\{\widehat{\lambda}_i\}_{i=1}^N$ and $\{\widehat{\gamma}_t\}_{t=1}^T$ defined in (1.3 .1) and (1.3 .2) correspondingly are uniformly consistent estimators of $\{\lambda_i\}_{i=1}^N$ and $\{\gamma_t\}_{t=1}^T$ correspondingly under mild conditions. In particular, the logarithm terms are present in above results because of the high dimensionality of λ and γ. The extra logarithm term is the price often paid in high-dimensional studies; see e.g., Zhu (2017) and Wang et al. (2018).

1.4 The specification tests

In this section, we propose two specifications tests to examine the existence of $\{\lambda_i\}_{i=1}^N$ and $\{\gamma_t\}_{t=1}^T$ correspondingly. We first study the asymptotic null distributions and asymptotic local power properties of the specification tests. Then, we give the implementations of our specification tests in practice. And finally, we propose the bootstrap version of our specification tests.

1.4.1 Max-type tests for slope heterogeneity and time-varyingness

It is of practical interests in empirical studies to test whether slope heterogeneity or time-vartingnes exist or not. To this end, we consider to test one of following two null hypotheses:

H_{0a}: $\lambda_i = 0$ for all i;

H_{0b}: $\gamma_t = 0$ for all t.

The corresponding alternative hypotheses are the negations of H_{0a} and H_{0b} shown above,

H_{1a}: $\lambda_i \neq 0$ for some i;

H_{1b}: $\gamma_t \neq 0$ for some t.

From above, we can see that under H_{0a}, there is no slope heterogeneity, and under H_{0b}, the slope time-varyingess is excluded. In contrast, under the alternatives, slope may exhibit heterogeneity, or time-varyingess, or both of them.

Now, to test H_{0a} above, we propose to use the max-type test statistic $\mathcal{T}_\lambda$ defined below:

$$\mathcal{T}_\lambda = \max_{1\leq i\leq N} \widehat{\lambda}_i' \widehat{\mathcal{V}}_{\lambda i}^{-1} \widehat{\lambda}_i. \quad (1.4.1)$$

Similarly, to test H_{0b} above, we propose to use the max-type test statistic $\mathcal{T}_\gamma$ defined below:

$$\mathcal{T}_\gamma = \max_{1\leq t\leq T} \widehat{\gamma}_t' \widehat{\mathcal{V}}_{\gamma}^{-1} \widehat{\gamma}_t. \quad (1.4.2)$$

To study the asymptotic local power of the tests proposed above, we specify the following sequence of Pitman local alternatives:

$$H_{1a,NT}: \lambda_i = \sqrt{\frac{\ln N}{T}} \cdot \widetilde{\lambda}_i \quad \text{for all } i \text{, and} \quad H_{1b,NT}: \gamma_t = \sqrt{\frac{\ln T}{N}} \cdot \widetilde{\gamma}_t \quad \text{for all } t,$$

where $\widetilde{\lambda}_i$ and $\widetilde{\gamma}_t$ satisfy the corresponding conditions in Assumption A1.1.

The motivation behind our test statistics $\mathcal{T}_\lambda$ and $\mathcal{T}_\gamma$ is simple and intuitive. In a nutshell, replying on the results in section 1.3, for instance, under H_{0a}, we have $\widehat{\mathcal{V}}_{\lambda i}^{-1/2}\widehat{\lambda}_i \xrightarrow{d} \mathcal{N}(0_K, I_K)$, then it follows that $\widehat{\lambda}_i \widehat{\mathcal{V}}_{\lambda i}^{-1}\widehat{\lambda}_i \xrightarrow{d} \chi_K^2$ for each i. We will show that $\mathcal{T}_\lambda$ follow standard the Gumbel distribution after being rescaled appropriately under the null. In contrast, under $H_{1a,NT}$, $\mathcal{T}_\lambda$ will deviate from the null distribution quickly and substantially.

Denote $\Pi_{NT} = \ln \delta_{NT}$. To study the asymptotic properties of $\mathcal{T}_\lambda$ and $\mathcal{T}_\gamma$, we further impose following two assumptions.

Assumption A1.9. *Let κ be positive but small enough,*

(i) $\gamma_t = \Pi_{NT}^{-\kappa} \cdot \widetilde{\gamma}_t^\star$ *for* $1 \le t \le T$*, where* $\widetilde{\gamma}_t^\star$ *satisfies the the corresponding conditions in Assumption A1.1.*

(ii) $\lambda_i = \Pi_{NT}^{-\kappa} \cdot \widetilde{\lambda}_i^\star$ *for* $1 \le i \le N$*, where* $\widetilde{\lambda}_i^\star$ *satisfies the the corresponding conditions in Assumption A1.1.*

Note that κ in Assumption A1.9(i)-(ii) can be different. Since Π_{NT} is the logarithmic function of N or T, it is not stringent for the real data in empirical studies. Actually, Assumptions A1.9 imposed above are driven by practical considerations. For such considerations, we will discuss in the next subsection after we give the implementations of $\mathcal{T}_\lambda$ and $\mathcal{T}_\gamma$. Meanwhile, Assumption A1.9 also has some theoretical implications as follows: unlike studies on the CRC panel data models or time-varying panel data models that exclude the existence of the slope time-varyingness or heterogeneity by the model setup, $\mathcal{T}_\lambda$ can allow for the the existence of $\{\gamma_t\}_{t=1}^T$ as long as the magnitude of $\{\gamma_t\}_{t=1}^T$ is mild; similar implications also apply to $\mathcal{T}_\gamma$. We can ease the exposition of proofs for Theorem

1.4.1 and Theorem 1.4.2 under Assumption A1.9, however, we address that Assumption A1.9 can dropped at the cost of lengthy arguments in the view of theoretical derivations.

The following two theorems state the asymptotic behaviors of $\mathcal{T}_\lambda$ defined in (1.4 .1), and $\mathcal{T}_\gamma$ defined in (1.4 .2) under the null and alternatives formally.

Theorem 1.4.1. *Under Assumptions A1.1-A1.5,A1.8, and A1.9(i), as (N,T) go to* (∞, ∞) *jointly, under* H_{0a}*, it holds that*

$$P\left(A_N \mathcal{T}_\lambda \leq x + B_N\right) = e^{-e^{-x}}, \tag{1.4 .3}$$

where $A_N = \frac{1}{2}$*, and* $B_N = \ln N + (\frac{K}{2} - 1)\ln\ln N - \ln\Gamma(\frac{K}{2})$*,* $\Gamma(.)$ *here is the Gamma function. Under the local alternative* $H_{1a,NT}$*, if* $\frac{T}{\ln N}\left\|\lambda_i\right\|_2^2 \to \infty$ *holds for at least one* i*, it holds that*

$$P\left(\mathcal{T}_\lambda > c_{\alpha,N}\right) = 1, \tag{1.4 .4}$$

where $c_{\alpha,N} = 2B_N - \ln|\ln(1-\alpha)|^2$*,* α *is the significant level selected by user.*

Theorem 1.4.2. *Under Assumptions A1.1-A1.5,A1.8, A1.9(ii), as (N,T) go to* (∞, ∞) *jointly, under* H_{0b}*, it holds that*

$$P\left(A_T \mathcal{T}_\gamma \leq x + B_T\right) = e^{-e^{-x}}, \tag{1.4 .5}$$

where $A_T = \frac{1}{2}$*, and* $B_T = \ln T + (\frac{K}{2} - 1)\ln\ln T - \ln\Gamma(\frac{K}{2})$*,* $\Gamma(.)$ *here is the Gamma function. Under the local alternative* $H_{1b,NT}$*, if* $\frac{N}{\ln T}\left\|\gamma_t\right\|_2^2 \to \infty$ *holds for at least one* t*, it holds that*

$$P\left(\mathcal{T}_\gamma > c_{\alpha,T}\right) = 1, \tag{1.4 .6}$$

where $c_{\alpha,T} = 2B_T - \ln|\ln(1-\alpha)|^2$*,* α *is the significant level selected by user.*

Note that (1.4 .3) in Theorem 1.4.1 and (1.4 .5) in Theorem 1.4.2 say that under the null, our specification will converge to the Gumbel distribution after being centered and rescaled appropriately.

As long as the consistency of our specification tests are concerned, (1.4 .4) in Theorem 1.4.1 reveals that $\mathcal{T}_\lambda$ has nontrivial power against the local alternatives shrinking to the null at rate $O\left(\sqrt{\ln N/T}\right)$, and (1.4 .6) in Theorem 1.4.2 has the similar implication.

1.4.2 Implementations of the specification tests

In this section, we offer two algorithms to implement $\mathcal{T}_\lambda$ and $\mathcal{T}_\gamma$ respectively in practice.

Algorithm 1.1. We implement $\mathcal{T}_\lambda$ by following steps,

(1) Regress $\ddot{y}_{it}$ on $\ddot{x}_{it}$ to obtain $\widehat{\beta}$ and $\widehat{\ddot{y}}_{it} = \ddot{y}_{it} - \ddot{x}_{it}\widehat{\beta}$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$.

(2) Regress $\widehat{\ddot{y}}_{it}$ on $\breve{x}_{it}$ to obtain $\widehat{\gamma}_t$ for $t = 1, \ldots, T$. Further, let $\widehat{\ddot{y}}^{*}_{it} = \widehat{\ddot{y}}_{it} - \breve{x}'_{it}\widehat{\gamma}_t$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$.

(3) Regress $\widehat{\ddot{y}}^{*}_{it}$ on $\tilde{x}_{it}$ to obtain $\widehat{\lambda}_i$, and $\widehat{\ddot{u}}_{it} = \widehat{\ddot{y}}^{*}_{it} - \tilde{x}'_{it}\widehat{\lambda}_i$ for $i = 1, \ldots, N$.

(4) According to Corollary 1.3.5, for each i, calculate $\widehat{\mathcal{V}}_{\lambda i}$ based on $\tilde{x}_{it}$, $\widehat{\gamma}_t$ for $t = 1, \ldots, T$ and $\widehat{\ddot{u}}_{it}$ for $i = 1, \ldots, N$, $t = 1, \ldots, T$ obtained in Step (3) above.

(5) Calculate $\mathcal{T}_\lambda$ based on $\widehat{\lambda}_i$ in step (3) and $\widehat{\mathcal{V}}_{\lambda i}$ in step (4) for $i = 1, \ldots, N$

Particularly, in Step (2) in Algorithm 1.1, we actually calculate $\widehat{\ddot{y}}^{*}_{it} = \ddot{y}_{it} - \ddot{x}'_{it}\widehat{\beta} - \breve{x}'_{it}\widehat{\gamma}_t$, in doing so we get the consistent estimator $\widehat{\ddot{u}}_{it}$ that allows for the existence of $\{\gamma_t\}$ under H_{0a}.

Remark 1.7. (The choice of lags p) In Algorithm 1.1, we employ HAC estimator of $\widehat{\mathcal{V}}_{\lambda i}$ for $i = 1, \ldots, N$. In practice, the selection of lags p will affect the actual size of our test $\mathcal{T}_\lambda$. There are some rule-of-thumbs to determine p as follows: $p = \lceil 0.75T^{1/3} \rceil$, $p = \lceil T^{1/4} \rceil$, and $p = \lceil 4(T/100)^{2/9} \rceil$, where $\lceil C \rceil$ represent the integer that is nearest to C. In our simulations, $p = [4(T/100)^{2/9}]$ outperforms other two candidates, which implies suitable and longer lags will improve finite-sample performances of the specification tests.

Algorithm 1.2. We implement $\mathcal{T}_\gamma$ by following steps,

(1) Regress $\ddot{y}_{it}$ on $\ddot{x}_{it}$ to obtain $\widehat{\beta}$ and $\widehat{\ddot{y}}_{it} = \ddot{y}_{it} - \ddot{x}_{it}\widehat{\beta}$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$.

(2) Regress $\widehat{\ddot{y}}_{it}$ on $\tilde{x}_{it}$ to obtain $\widehat{\lambda}_i$ for $i = 1, \ldots, N$. Further, let $\widehat{\ddot{y}}^{\star}_{it} = \widehat{\ddot{y}}_{it} - \tilde{x}_{it}\widehat{\lambda}_i$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$.

(3) Regress $\widehat{\ddot{y}}^{\star}_{it}$ on $\breve{x}_{it}$ to obtain $\widehat{\gamma}_t$ and and $\widehat{\ddot{u}}_{it} = \widehat{\ddot{y}}^{\star}_{it} - \breve{x}'_{it}\widehat{\gamma}_t$ for $t = 1, \ldots, T$.

(4) According to Corollary 1.3.6, calculate $\widehat{\mathcal{V}}_{\gamma t}$ for each t based on $\breve{x}_{it}$, $\widehat{\lambda}_i$ for $i = 1, \ldots, N$ and $\widehat{\ddot{u}}_{it}$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$ obtained in Step (2) above.

(5) Calculate $\mathcal{T}_{\gamma}$ based on $\widehat{\gamma}_t$ in step 3 and $\widehat{\mathcal{V}}_{\gamma t}$ in step 4 for $t = 1, \ldots, T$.

Similarly, in Step (2) in Algorithm 1.1, we actually calculate $\widehat{\ddot{u}}_{it} = \ddot{y}_{it} - \ddot{x}'_{it}\widehat{\beta} - \tilde{x}'_{it}\widehat{\lambda}_i$, in doing so we get the consistent estimator $\widehat{\ddot{u}}_{it}$ that allows for for the existence of λ_i for $i = 1, \ldots, N$ under H_{0b}.

Remark 1.8. From Algorithm 1.1, we can see that the effects of $\{\gamma_t\}_{t=1}^{T}$ are considered and removed by its consistent estimator $\{\widehat{\gamma}_t\}_{t=1}^{T}$ when we implement $\mathcal{T}_{\lambda}$ in practice even if we have no prior knowledge about the existence of $\{\gamma_t\}_{t=1}^{T}$. It is reasonable to do so, because practitioners may worry about the adverse effects of $\{\gamma_t\}_{t=1}^{T}$ on the finite-sample properties of $\mathcal{T}_{\lambda}$. Indeed, if practitioners do iterated estimations for β , λ_i and γ_t, according to our proofs for Theorem 1.4.1, the term $\zeta^{\lambda}_{iT} + \xi^{\lambda}_{NT}$ can turn out to be $\zeta^{\lambda}_{iT} + \xi^{\lambda}_{NT} = N^{-1/2}T^{-1}\sum_{j\neq i}\sum_t x^*_{jt}x^{*\prime}_{jt}\left(\lambda_j - \widehat{\lambda}_j\right) + T^{-1/2}\sum_t\left(x^*_{it}x^{*\prime}_{it} - \Sigma_x\right)\left(\gamma_t - \widehat{\gamma}_t\right)$. We can show that this term is of order $O_p\left(N^{-1/2}T^{-1/2}\right) = o_p(1)$ by using Proposition 1.3.8(ii), which is negligible for in the derivations of the null distribution of $\mathcal{T}_{\lambda}$. Because of this fact, we impose a weaker condition in Assumption A1.9(i), because the terms bounded by $O_p\left(N^{-1/2}T^{-1/2}\right)$ will be dominated by the terms bounded by $O_p\left(\Pi_{NT}\right)$. We impose a weaker condition in Assumption A1.9(ii) by the similar considerations.

1.4.3 A bootstrap version of the specification tests

Consistent with comments in Castagnetti et al. (2015), our specification test statistics $\mathcal{T}_{\lambda}$ and $\mathcal{T}_{\gamma}$ also suffer from the slow convergence to the Gumbel distribution. The asymptotic distribution approximates the finite sample distribution poorly. As a consequence, we expect the use of asymptotic critical values will deteriorate the power and size of our specification tests. To rectify this issue, we propose to use following two bootstrap schemes to improve the finite sample performance of our specification tests. The first one is so-called block wild bootstrap (BWB) scheme and the second one is the wild bootstrap (WB) scheme. Particularly, the BWB is an extension of the WB that handles the serial correlations among time-series observations.

Remark 1.9. Note that we use time series data for each individual i to construct $\mathcal{T}_\lambda$. Because Assumption A1.8 implies the weak dependence, it can be sufficient for practitioners to use the WB scheme only to obtain the bootstrap p-values for both $\mathcal{T}_\lambda$ and $\mathcal{T}_\gamma$, the WB implementations works for both $\mathcal{T}_\lambda$ and $\mathcal{T}_\gamma$ from our simulation results in Section 1.5 even when the dependence among time-series observations are moderate. However, for robustness checks, practitioners can further apply the BWB scheme to $\mathcal{T}_\lambda$ to handle the unknown but possibly strong time-series persistence.

First, we present the algorithm for the implementation of $\mathcal{T}_\lambda$ under the BWB scheme.

Algorithm BWB: Block Wild Bootstrap Scheme for $\mathcal{T}_\lambda$

1. Execute Algorithm 1.1 to the original test statistics $\mathcal{T}_\lambda$.

2. Set a block size b_T, such that $1 \leq b_T < T$. Denote the blocks by $B_s = \{(s-1)b_T + 1, \ldots, sb_T\}$, $s = 1, \ldots, L_T$, $L_T = \lceil T/b_T \rceil$, the number of blocks, is constructed to be an integer for the convenience of presentation.

3. For each $i = 1, 2, \ldots, N$, take i.i.d random draws $\{\xi_{is}\}_{s=1}^{L_T}$, independent of the data, from a common distribution W, where $\mathbb{E}(W) = 0$, $\mathbb{E}(W^2) = 1$. Define the auxiliary variables $w_{it} = \xi_{is}$, if $t \in B_s$, for $t = 1, \ldots, T$, where B_s is constructed in Step 2.

4. For ecah i, obtain the bootstrap error $\widehat{\ddot{u}}_{it}^* = \widehat{\ddot{u}}_{it} w_{it}$ with w_{it} defined in Step 3 for $t = 1, 2, \ldots, T$,. Then, we construct the bootstrap sample $\widehat{\ddot{y}}_{it}^* = \tilde{x}_{it}'\lambda_i + \widehat{\ddot{u}}_{it}^*$ for $i = 1, \ldots, N$ and $t = 1, 2, \ldots, T$ as if the null $\lambda_i = 0$ for all i holds.

5. Given the bootstrap sample $\{\widehat{\ddot{y}}_{it}^*, x_{it}\}$, re-execute Step (3)-(5) in Algorithm 1.1 above to obtain test statistic $\mathcal{T}_\lambda^*$ in bootstrap world.

6. Repeat steps 2 to 5 for B times and index the bootstrap test statistics $\left\{\mathcal{T}_{\lambda,b}^{*}\right\}_{b=1}^{B}$. Then calculate the bootstrap $p-$value by $p^{*}=B^{-1}\sum_{b=1}^{B}\mathbb{I}\left\{\mathcal{T}_{\lambda,b}^{*}>\mathcal{T}_{\lambda}\right\}$.

Algorithm BWB says that we generate the bootstrap samples for each individual i under the BWB scheme as follows: (1) divide T consecutive observations into L_T blocks with b_T consecutive observations in each block. Assume $L_T b_T = T$ implicitly for simplicity; (2) generate L_T external and auxiliary random variables $\xi_{il} \overset{i.i.d}{\sim} (0,1)$ with $l = 1, \ldots, L_T$, all residuals in the block l then are multiplied by the external random variable ξ_{il} ; (3) obtain the bootstrap samples as if $\lambda_i = 0$ holds for all i.

Remark 1.10. Let the block size $b_T = 1$ in Step 2 of Algorithm BWB, we actually implement the wild bootstrap version of $\mathcal{T}_\lambda$.

We then provide the algorithm for the implementation of $\mathcal{T}_\gamma$ under the WB scheme.

Algorithm WB: Wild Bootstrap Scheme for $\mathcal{T}_\gamma$

1. Execute Algorithm 1.2 stated above to obtain the test statistics $\mathcal{T}_\gamma$.

2. For each i, take i.i.d random draws $\{\xi_{it}\}_{t=1}^{T}$ independent of the data, from a common distribution W, where $\mathbb{E}(W) = 0$, $\mathbb{E}\left(W^2\right) = 1$.

3. For each $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$, obtain the bootstrap error $\widehat{\breve{u}}_{it}^{*} = \widehat{\breve{u}}_{it}\xi_{it}$ with ξ_{it} defined in Step 2. Then, we construct the bootstrap sample $\widehat{\breve{y}}_{it}^{*} = \breve{x}_{it}'\gamma_t + \widehat{\breve{u}}_{it}^{*}$ for $t = 1, 2, \ldots, T$ and $i = 1, \ldots, N$ as if the null $\gamma_t = 0$ for $t = 1, \ldots, T$ holds.

4. Given the bootstrap sample $\{\widehat{\breve{y}}_{it}^{*}, x_{it}\}$, re-execute Step (3)-(5) in Algorithm 1.2 presented above to obtain test statistic $\mathcal{T}_\gamma^{*}$ in bootstrap world.

5. Repeat steps 2 to 5 for B times and index the bootstrap test statistics $\left\{\mathcal{T}_{\gamma,b}^{*}\right\}_{b=1}^{B}$. Then calculate the bootstrap $p-$value by $p^{*}=B^{-1}\sum_{b=1}^{B}\mathbb{I}\left\{\mathcal{T}_{\gamma,b}^{*}>\mathcal{T}_{\gamma}\right\}$.

By comparing the details of Algorithm BWB and Algorithm WB, we can see that the BWB scheme is a generalization of the WB scheme in the sense that the BWB scheme uses the block structure to capture the serial correlations among time-series observations .

Remark 1.11. For external random variables used in both Algorithm BWB and Algorithm WB, we recommend to use so-called Rademacher sequences, namely,

$$\xi_{it} = \begin{cases} 1 & \text{with probability } p = \frac{1}{2}, \\ -1 & \text{with probability } 1-p, \end{cases}$$

with the properties $E\xi_{it} = 0$, $E\xi_{it}^2 = 1$, $E\xi_{it}^3 = 0$, and $E\xi_{it}^4 = 1$.

The theorem below justifies the asymptotic validity of the bootstrap version of our specification tests:

Theorem 1.4.3. ***(a)*** *Suppose Assumptions A1.1-A1.5, A1.8 and A1.9 hold.*

In addition, under Block Wild Bootstrap scheme, assume that $1/b_T + b_T/T^{1/3} \to 0$*, then, as* $(N,T) \to \infty$*,*

$$\sup_{x\in\mathbb{R}} \left| P^* \left(A_N \mathcal{T}_\lambda^* \leq x + B_N \right) - P \left(A_N \mathcal{T}_\lambda \leq x + B_N \right) \right| \xrightarrow{P^*} 0,$$

as $N, T \to \infty$*, where* P^* *denotes probability measure induced by the wild bootstrap conditional on the observed sample* $\mathcal{W}_{NT} \equiv \{(x_{it}, y_{it}), i = 1, \ldots, N, t = 1, \ldots, T\}$.

(b) *Similarly, under Assumptions A1.1-A1.5, A1.8 and A1.9,*

$$\sup_{x\in\mathbb{R}} \left| P^* \left(A_N \mathcal{T}_\gamma^* \leq x + B_N \right) - P \left(A_N \mathcal{T}_\gamma \leq x + B_N \right) \right| \xrightarrow{P^*} 0,$$

as $N, T \to \infty$*, where* P^* *denotes probability measure induced by the wild bootstrap conditional on the observed sample* $\mathcal{W}_{NT} \equiv \{(x_{it}, y_{it}), i = 1, \ldots, N, t = 1, \ldots, T\}$.

Remark 1.12. For strong serial and cross-sectional dependence, we conjecture that the Moving Block Bootstrap (MBB) scheme may work. Gonçalves (2011) show the validity of the MBB scheme for a homogeneous panel data model with individual fixed effects that allows for both serial and cross-sectional dependence. We leave this extension as future research.

1.5 Monte Carlo Simulations

In this section we evaluate the finite sample performance of the specification tests by means of a set of Monte Carlo experiments.

1.5.1 Wild Bootstrap Scheme

We implement the bootstrap version $\mathcal{T}_\lambda$ and $\mathcal{T}_\gamma$ under the Wild Bootstrap scheme in this subsection to study the size and power of the bootstrap version of our specification test statistics in practice. In particular, for the bootstrap version $\mathcal{T}_\lambda$, we just set block size $b_T = 1$ such that the Block Wild Bootstrap scheme degenerates to Wild Bootstrap scheme.

We consider following data generating processes (DGPs),

DGP 1: $\quad y_{it} = (\beta + \gamma_t)\, x_{it} + \eta_i + \omega_t + u_{it},$

DGP 2: $\quad y_{it} = (\beta + \lambda_i)\, x_{it} + \eta_i + \omega_t + u_{it},$

DGP 3: $\quad y_{it} = (\beta + \lambda_i + \gamma_t)\, x_{it} + \eta_i + \omega_t + u_{it}.$

For all DGPs, $\{\eta_i\}_{i=1}^N$ is drawn independently from $\mathcal{N}\left(0, \sigma_\eta^2\right)$, $\{\omega_t\}_{t=1}^T$ is drawn independently from $\mathcal{N}\left(0, \sigma_\omega^2\right)$, x_{it} and u_{it} follow the AR(1) process, namely

$$
\begin{aligned}
x_{it} &= \rho_x' x_{it-1} + \pi\left(\eta_i + \omega_t\right) + v_{it}, \\
u_{it} &= \rho_u u_{it-1} + \epsilon_{it},
\end{aligned}
$$

where v_{it} is drawn independently from $\mathcal{N}\left(0, \sigma_v^2\right)$, ϵ_{it} is drawn independently from $\mathcal{N}\left(0, \sigma_\epsilon^2\right)$ for each i. For *DGP 1*, $\{\gamma_t\}_{t=1}^T$ is drawn independently from $\mathcal{N}\left(0, \sigma_\gamma^2\right)$. For *DGP 2*, $\{\lambda_i\}_{i=1}^N$ is drawn independently from $\mathcal{N}\left(0, \sigma_\lambda^2\right)$.

When we study the size and power of $\mathcal{T}_\lambda$, we generate data from *DGP 1* and *DGP 3* under $H_{1a,NT}$ respectively, for $\{\gamma_t\}_{t=1}^T$, $\gamma_t = N^{-1/2} \cdot \gamma_t^*$, where γ_t^* is drawn independently from $\mathcal{N}\left(0, \sigma_\gamma^2\right)$ for $t = 1, \ldots, T$.

When we study the size and power of $\mathcal{T}_\gamma$, we generate data from *DGP 2* and *DGP 3* under $H_{1b,NT}$

respectively, for $\{\lambda_i\}_{i=1}^N$, $\lambda_i = T^{-1/2} \cdot \lambda_i^*$, where λ_i^* is drawn independently from $\mathcal{N}(0, \sigma_\lambda^2)$ for $i = 1, \ldots, N$.

Besides, for the power study under *DGP 3*, we consider the local alternatives as follows,

$H_{1a,NT}$: $\widetilde{\lambda}_i = C \cdot \sqrt{\frac{\ln N}{T}} \cdot \lambda_i$ such that λ_i are i.i.d sequences with $\lambda_i = \begin{cases} 1 & \text{w.p. } 1/2 \\ -1 & \text{w.p. } 1/2 \end{cases}$, and $C \in \{1, 2, 3\}$.

$H_{1b,NT}$: $\widetilde{\gamma}_t = C \cdot \sqrt{\frac{\ln T}{N}} \cdot \gamma_t$ such that γ_t are i.i.d sequences with $\gamma_t = \begin{cases} 1 & \text{w.p. } 1/2 \\ -1 & \text{w.p. } 1/2 \end{cases}$, and $C \in \{0.5, 1, 1.5\}$.

Now, we set fixed values for the user-specified parameters as follows: $\beta = 0.5$, $\rho_x = \rho_u = 0.5$, $\pi = 1$, $\sigma_v^2 = \sigma_\epsilon^2 = 1$, $\sigma_\lambda^2 = \{0.3, 0.5, 1\}$, $\sigma_\gamma^2 = \{0.3, 0.5, 1\}$, $(\sigma_\eta^2, \sigma_\omega^2) = (0.1, 0.1)$. For choices of lags p for the HAC estimators, we recommend to use $p = \lceil 4(T/100)^{2/9} \rceil$. We define variance ratio as $vr \equiv \sigma_\gamma^2/\sigma_\epsilon^2$ for *DGP 1* and $vr \equiv \sigma_\lambda^2/\sigma_\epsilon^2$ for *DGP 2* , which represent the signal-noise ratio.

In this subsection, we calculate $\mathcal{T}_\lambda$ by Algorithm 1.1 and the wild bootstrap version of $\mathcal{T}_\lambda$ by Algorithm BWB with the block size $b_T = 1$; we also obtain $\mathcal{T}_\gamma$ by Algorithm 1.2 and the bootstrap version of $\mathcal{T}_\gamma$ by Algorithm WB. For each scenario, we conduct 500 replications with $B = 250$ bootstrap resamples in each replication. We consider the 5% nominal level in all cases.

1.5.2 Block Wild Bootstrap Scheme

In this subsection, we implement the block wild bootstrap version $\mathcal{T}_\lambda$ to study the effects of the degree of serial correlations among time-series observations on the size of the block wild bootstrap version of $\mathcal{T}_\lambda$ under the null H_{0a}. We still assume $\{x_{it}, u_{it}\}$ to follow the AR(1) process,

$$x_{it} = \rho_x x_{it-1} + \pi\left(\eta_i + \omega_t\right) + v_{it}$$
$$u_{it} = \rho_u u_{it-1} + \epsilon_{it},$$

where π is the tuning parameter that controls the degree of correlations with fixed effects, and we set $\pi = 1$ as the default value for all cases. For each i, $\{v_{it}\}_{t=1}^{T}$ is drawn independently from $N(0,1)$; for each i, $\{\epsilon_{it}\}_{t=1}^{T}$ is drawn independently from $N(0,1)$. Further, we set $\rho_x \in \{0.3, 0.6, 0.9\}$ and $\rho_u \in \{0.3, 0.6, 0.9\}$ to study the effects of the degree of serial correlations among observations on the size of the block wild bootstrap version of our specification test statistic $\mathcal{T}_\lambda$ under H_{0a}.

We generate data by *DGP 4* specified below,

DGP 4: $y_{it} = (\beta + \gamma_t) x_{it} + \eta_i + \omega_t + u_{it}$ with $\gamma_t = N^{-1/2} \cdot \gamma_t^*$ for $t = 1, \ldots, T$, where γ_t^* is drawn independently from $\mathcal{N}\left(0, \sigma_\gamma^2\right)$ with $\sigma_\gamma^2 = 0.3$.

We set $\beta = 0.5$ as fixed in all simulations, meanwhile, for fixed effects, $\{\eta_i\}_{i=1}^{N}$ is drawn independently from $\mathcal{N}\left(0, \sigma_\eta^2\right)$ with $\sigma_\eta^2 = 0.1$ and $\{\omega_t\}_{t=1}^{T}$ is drawn independently from $\mathcal{N}\left(0, \sigma_\omega^2\right)$ with $\sigma_\omega^2 = 0.1$.

We calculate $\mathcal{T}_\lambda$ by Algorithm 1.1 and the wild block bootstrap version of $\mathcal{T}_\lambda$ by Algorithm BWB with different block size b_T shown in the related Tables. For each scenario, we conduct 1000 replications with $B = 250$ bootstrap resamples in each replication. We consider the 5% nominal level in all cases.

1.5.3 Results

Table 1.1 displays empirical rejection rates of $\mathcal{T}_\lambda$ and $\mathcal{T}_\gamma$ under Wild Bootstrap scheme for each combination of N and T with different variance ratios. The left panel of Table 1.1 shows that both $\mathcal{T}_\lambda$ and $\mathcal{T}_\gamma$ are too conservative in most cases regardless of values of the variance ratio based on the asymptotic critical values. These results are due to the poor approximation of asymptotic distributions to the finite-sample distribution and consistent with our anticipations. The empirical size of the wild bootstrap version of $\mathcal{T}_\lambda$ and $\mathcal{T}_\gamma$ in the right panel of Table 1.1 are very close to the 5% nominal level of significance. It is noteworthy to see that in several cases, $\mathcal{T}_\lambda$ tends to be over-sized slightly. We conjecture that the slight over-rejections originate from the moderate serial correlations in $\{x_{it}, u_{it}\}$ because we set $\rho_x = \rho_u = 0.5$.

Table 1.2 and 1.3 exhibit the power of $\mathcal{T}_\lambda$ and $\mathcal{T}_\gamma$ for each combination of N and T with different

Table 1.1: Finite sample rejection frequency under the null hypotheses

size of $\mathcal{T}_\lambda$ under H_{0a}								
DGP	N	T	Based on Asy. CV			Based on Wild Bootstrap		
			vr=0.3	**vr=0.5**	**vr=1**	**vr=0.3**	**vr=0.5**	**vr=1**
1	30	30	0.010	0.016	0.036	0.028	0.022	0.028
		50	0.030	0.016	0.034	0.050	0.046	0.044
		100	0.042	0.034	0.060	0.056	0.044	0.052
1	50	30	0.002	0.000	0.004	0.020	0.024	0.022
		50	0.008	0.014	0.014	0.064	0.066	0.044
		100	0.016	0.022	0.022	0.054	0.052	0.062
1	100	30	0.000	0.000	0.000	0.032	0.028	0.024
		50	0.000	0.000	0.002	0.070	0.060	0.066
		100	0.002	0.004	0.004	0.064	0.070	0.050
size of $\mathcal{T}_\gamma$ under H_{0b}								
DGP	T	N	Based on Asy. CV			Based on Wild Bootstrap		
			vr=0.3	**vr=0.5**	**vr=1**	**vr=0.3**	**vr=0.5**	**vr=1**
2	30	30	0.056	0.066	0.068	0.040	0.044	0.050
		50	0.062	0.066	0.066	0.050	0.056	0.044
		100	0.068	0.076	0.070	0.042	0.050	0.050
2	50	30	0.020	0.024	0.030	0.040	0.042	0.046
		50	0.032	0.042	0.032	0.060	0.058	0.046
		100	0.046	0.040	0.050	0.050	0.052	0.048
2	100	30	0.016	0.006	0.010	0.054	0.054	0.054
		50	0.010	0.014	0.028	0.050	0.058	0.056
		100	0.022	0.022	0.022	0.054	0.062	0.050

variance ratios. The left panels of Table 1.2 and 1.3 indicate that $\mathcal{T}_\lambda$ and $\mathcal{T}_\gamma$ have very low power against the alternatives in most cases regardless of values of the variance ratio based on asymptotic critical values. In contrast,the right panels of Table 1.2 and 1.3 disclose that the wild bootstrap version of $\mathcal{T}_\lambda$ and $\mathcal{T}_\gamma$ have decent power against the alternatives.

In Table 1.4, we present the empirical rejection rates with block size $b_T = 1$ and $b_T = \sqrt{T}/2$ under various combinations of the degree of serial correlations among $\{x_{it}\}$ and $\{u_{it}\}$ under the null. When $(\rho_x, \rho_u) = (0.3, 0.3)$, $(\rho_x, \rho_u) = (0.6, 0.3)$ and $(\rho_x, \rho_u) = (0.3, 0.6)$, Table 1.4 reveals that Wild Bootstrap scheme and Block Wild Bootstrap scheme have almost same performances on the empirical size. If the serial correlation is very strong, the Block Wild Bootstrap scheme outperforms Wild Bootstrap scheme significantly and has the proper size close to the nominal level.

Table 1.2: Finite sample rejection frequency under the local alternatives

power of $\mathcal{T}_\lambda$ under $H_{1a,NT}$								
			Based on Asy. CV			Based on Wild Bootstrap		
DGP	N	T	C=1	C=2	C=3	C=1	C=2	C=3
				vr=0.3			**vr=0.3**	
3	30	30	0.011	0.020	0.073	0.066	0.176	0.356
		50	0.066	0.265	0.656	0.104	0.406	0.700
		100	0.332	0.921	0.999	0.292	0.866	1.000
3	50	30	0.000	0.003	0.009	0.116	0.362	0.624
		50	0.013	0.090	0.382	0.166	0.644	0.934
		100	0.242	0.931	1.000	0.456	0.996	1.000
3	100	30	0.000	0.000	0.000	0.198	0.626	0.922
		50	0.004	0.010	0.072	0.358	0.916	1.000
		100	0.175	0.911	1.000	0.698	1.000	1.000
				vr=0.5			**vr=0.5**	
3	30	30	0.015	0.032	0.070	0.078	0.184	0.306
		50	0.057	0.276	0.670	0.090	0.346	0.696
		100	0.296	0.902	1.000	0.306	0.894	0.994
3	50	30	0.002	0.005	0.006	0.096	0.314	0.570
		50	0.015	0.119	0.359	0.176	0.576	0.912
		100	0.252	0.933	1.000	0.490	0.988	1.000
3	100	30	0.000	0.000	0.000	0.196	0.564	0.920
		50	0.002	0.008	0.067	0.342	0.890	0.996
		100	0.170	0.924	1.000	0.748	0.728	1.000
				vr=1			**vr=1**	
3	30	30	0.031	0.062	0.097	0.056	0.152	0.280
		50	0.083	0.355	0.713	0.076	0.350	0.616
		100	0.336	0.941	1.000	0.246	0.846	0.998
3	50	30	0.004	0.012	0.002	0.074	0.284	0.542
		50	0.027	0.027	0.390	0.158	0.602	0.892
		100	0.277	0.935	1.000	0.464	0.974	1.000
3	100	30	0.000	0.002	0.000	0.184	0.544	0.864
		50	0.006	0.019	0.067	0.320	0.870	1.000
		100	0.197	0.932	1.000	0.722	1.000	1.000

Table 1.3: Finite sample rejection frequency under the local alternatives

power of $\mathcal{T}_\gamma$ under $H_{1b,NT}$								
			Based on Asy. CV			Based on Wild Bootstrap		
DGP	T	N	C=0.5	C=1	C=1.5	C=0.5	C=1	C=1.5
				vr=0.3			**vr=0.3**	
3	30	30	0.157	0.658	0.961	0.144	0.528	0.870
		50	0.258	0.891	0.999	0.244	0.810	0.998
		100	0.387	0.973	1.000	0.304	0.938	1.000
3	50	30	0.086	0.478	1.000	0.194	0.760	0.960
		50	0.180	0.866	1.000	0.322	0.962	1.000
		100	0.354	0.981	1.000	0.394	0.986	1.000
3	100	30	0.041	0.261	0.756	0.272	0.916	1.000
		50	0.130	0.805	1.000	0.514	0.994	1.000
		100	0.302	1.000	1.000	0.542	0.998	1.000
				vr=0.5			**vr=0.5**	
3	30	30	0.208	0.672	0.971	0.144	0.526	0.862
		50	0.263	0.891	1.000	0.224	0.836	0.994
		100	0.364	0.979	1.000	0.272	0.928	1.000
3	50	30	0.009	0.498	0.918	0.190	0.720	0.950
		50	0.179	0.875	1.000	0.318	0.940	1.000
		100	0.360	0.990	1.000	0.370	0.982	1.000
3	100	30	0.029	0.273	0.763	0.326	0.944	1.000
		50	0.140	0.778	1.000	0.480	0.994	1.000
		100	0.294	0.988	1.000	0.542	1.000	1.000
				vr=1			**vr=1**	
3	30	30	0.176	0.637	0.965	0.134	0.496	0.870
		50	0.278	0.888	1.000	0.204	0.824	0.990
		100	0.369	0.977	1.000	0.318	0.948	1.000
3	50	30	0.113	0.522	0.904	0.196	0.706	0.948
		50	0.222	0.869	0.999	0.288	0.924	1.000
		100	0.349	0.977	1.000	0.380	0.988	1.000
3	100	30	0.040	0.261	0.765	0.290	0.904	1.000
		50	0.139	0.804	0.999	0.512	0.990	1.000
		100	0.287	0.990	1.000	0.530	1.000	1.000

Table 1.4: Finite sample rejection frequency under the null hypothesis

Size of $\mathcal{T}_\lambda$ under H_{0a}				Block Wild Bootstrap				
						(ρ_x, ρ_u)		
DGP	Block Size	N	T	(0.3,0.3)	(0.3,0.6)	(0.3,0.9)	(0.6,0.3)	(0.6,0.6)
	$b_T = 1$	30	30	0.056	0.075	0.115	0.069	0.129
4	$b_T = \sqrt{T}/2$	30	30	0.065	0.050	0.057	0.078	0.086
	$b_T = 1$	100	100	0.061	0.073	0.086	0.060	0.149
	$b_T = \sqrt{T}/2$	100	100	0.051	0.064	0.041	0.064	0.066

1.6 Empirical Applications

In labor and demographic economics, an important study by Shimer (2001) discloses the negative effects of youth share of working-age population on the unemployment using U.S. state-level panel data. Skans (2005) applies the methodology of Shimer (2001) to study the effect of age distribution on unemployment rates and finds the reverse effects of a large share of workers aged 50-60 on the labor market in terms of higher unemployment and lower employment. Biagi and Lucifora (2005) use a similar empirical model to figure out the relationships among demographics, education changes and unemployment changes.

It is interesting to revisit the relation of unemployment rates and the youth shares of working-age population because many recent researches in labor and demographic economics concern the heterogeneous effects. As an illustrative example, following the empirical model in Shimer (2001), it is equivalent to apply the specification tests to check the slope heterogeneity and time-varyingness respectively.

The generalization of the empirical model in Shimer (2001) is given as below,

$$\log rate_{it} = \eta_i + \omega_t + \beta_{it} \log share_{it} + \epsilon_{it}, \tag{1.6 .1}$$

where the dependent variable $\log rate_{it}$ is the unemployment rate in country (or state) i and year t, and the explanatory variable $\log share_{it}$ is the youth shares of working-age population. Besides, η_i and ω_t are the time-specific and individual-specific fixed effects respectively. In particular, ω_t aims to capture any macroeconomic shocks while η_i accommodates considerable cross sectional variations in unemployment and demographics.

In order to apply our specification test statistics, we decompose $\beta_{it} = \beta + \lambda_i + \gamma_t$, thus, the corresponding null hypotheses are:

H_{0a}: $\lambda_i = 0$ for all i,

and

H_{0b}: $\gamma_t = 0$ for all t.

In our empirical study, we use panel dataset[1] consisting of six groups[2]. For each case, we collect annual data from 1991 to 2013 for unemployment rates and youth shares of working-age population.[3]Following Shimer (2001), we also calculate the youth share of working-age population as the ratio of population ages 16-24 over population ages 16-64. We report bootstrap p-values based on the bootstrap resamples $B = 1000$.

Table 1.5: Test statistics and p-values for Empirical Study

High Income Countries(N=52, T=23)		
	H_{0a}	H_{0b}
Test Statistics	8.470	9.787
Bootstrap p-value	0.002	0.230
Middle Income Countries(N=87, T=23)		
	H_{0a}	H_{0b}
Test Statistics	38.647	10.158
Bootstrap p-value	0.000	0.187
Upper Middle Income Countries (N=44, T=23)		
	H_{0a}	H_{0b}
Test Statistics	9.042	4.721
Bootstrap p-value	0.000	0.461
Lower Middle Income Countries(N=43, T=23)		
	H_{0a}	H_{0b}
Test Statistics	13.8979	5.812
Bootstrap p-value	0.000	0.849
Low Income Countries(N=32, T=23)		
	H_{0a}	H_{0b}
Test Statistics	9.2184	5.205
Bootstrap p-value	0.003	0.777
All Countries (N=171, T=23)		
	H_{0a}	H_{0b}
Test Statistics	38.217	15.606
Bootstrap p-value	0.000	0.073

Table 1.5 summarizes the testing results for our specification tests. We can reject the null hypotheses H_{0a} based on bootstrap p-values at significance level 5%. Meanwhile, we fail to reject H_{0b}

[1]The data is downloaded from World Development Indicators, https://databank.worldbank.org/data/source/world-development-indicators/preview/on,and Education Statistics, https://databank.worldbank.org/data/source/education-statistics-%5E-all-indicators/Type/TABLE/preview/on,provided by World Bank Open Data.

[2]low-income countries, lower-middle-income countries, upper-middle-income countries, middle income countries, high-income countries and all countries according to classifications in World Bank Open Data

[3]The countries that data are not available in each case are dropped to keep panel data balanced.

for all cases based on bootstrap p-values at significance level 5% . Therefore, it can be more reasonable to consider the random coefficients panel data models to accommodate the country-specific slope heterogeneity when we are of practical interest to have more insight on the relationship between unemployment rates and youth share of working-age population.

1.7 Conclusion

This paper introduces a two-way linear random coefficient panel data models with both time- and individual- specific fixed effects to capture slope heterogeneity and time-varyingness, in particular, we allow for the cross-sectional dependence in our model. The fixed effects included here can be arbitrarily correlated with explanatory variables. Therefore, our model extents the conventional random coefficients panel data models to accommodate both variable slopes in both time- and individual- dimensions and unobserved heterogeneity. This enlarges the potentials of applications of the random coefficients panel data models in practice.

To estimate parameters of interests, we follow the idea of the within-group fixed effects estimator for homogeneous panel data models with fixed effects. We then establish the limiting distributions of these estimates in the standard large N and large T framework so that practitioners can make the inference on parameters of interests in practice. Besides, to justify the desirability of such estimation strategy, we propose the two-way heterogeneity bias test.

Furthermore, we construct the max-type tests to examine the existence of slope heterogeneity and time-varyingness. Our specification tests suffer from the finite-sample size distortion associated when we use the asymptotic critical values in many cases. To rectify this issue, we employ two bootstrap schemes for our specification tests to correct the size of specification tests. Our Monte Carlo experiments confirm that the bootstrap implementations of our specification tests have reasonable size and decent power against the alternatives.

Finally, as an illustrative example, we apply our specification tests to reveal the relationship between the unemployment rates and youth shares of working-age population using panel data covering

171 countries at most and 23 years from 1991 to 2013, and testing results are in favor of the existence of country-specific slope heterogeneity.

We conclude the paper by pointing out some possible extensions for the current paper. First, we can replace time- and individual fixed effects with interactive fixed effects (IFEs) to accommodate the factor structure in our models. Second, following papers in high-dimensional statistics, we can impose the sparse structure on random coefficients and use some regularized estimators to revisit the two-way linear random coefficients panel data models in the current paper. Third, the group patterns and sparsity structure can co-exist among the random coefficients. The arguments are more lengthy for these extensions and we leave them for future research.

Chapter 2

Testing for Idiosyncratic Bubbles in the Presence of Nonstationary or Mildly Explosive Factors in Panel Data Models

2.1 Introduction

Since the pioneering works of Stock and Watson (2002a), Stock and Watson (2002b), Bai and Ng (2002), Bai (2003) and Bai (2009), panel factor models have become more and more popular in economics and finance. By employing a panel factor model, the variable of interest can be decomposed into a common component plus and an idiosyncratic error component. The former corresponds to the systematic risk, while the latter is linked to the specific risk for empirical researches in economics and finance. In the presence of speculative bubbles in the panel data, both regulators and financial analysts concern the sources of the bubbles' explosiveness in the panel of asset prices. That is, they have practical incentives to know whether the bubble risk arises from the individual-specific characteristics or the common component. Such concerns motivate our econometric analysis of the nature of the explosiveness of data through panel factor models in this paper. Ideally, we want to tell whether the explosive behavior is driven by the common component or by the idiosyctractic component, or both.

We provide a simple but effective procedure to test the bubbles' explosiveness in the idiosyncratic components by allowing for the presence of unit-root-type nonstationary factors or mildly explosive factors in large dimensional panel factor models. We derive the asymptotic null distribution and the asymptotic local power property of the proposed test statistic. We also provide a wild bootstrap scheme to improve our test's finite-sample performance. The asymptotic local power properties imply that the proposed test can detect effectively bubbles' explosiveness in the idiosyncratic components by exploiting the cross-sectional information in the panel data.

On the theoretical side, our study is extremely challenging due to the nonstationary behavior of the idiosyncractic component under the null and its mildly explosive behavior under the alternatives. Recently, Onatski and Wang (2020) have found highly persistent idiosyncratic components may lead to spurious factors in panel data models. These spurious factors are likely to be misidentified as common factors by existing principal-component-based (PC-based) method. Therefore, in order to disentangle the sources of the bubbles' explosiveness within the framework of panel factor models, it is of great importance to test the bubbles' explosiveness in idiosyncratic components as the first step. Besides, by virtue of spurious factors, the prevailing tests for the idiosyncratic bubbles can suffer from a lack of power against some alternatives when empirical researchers over-extract common factors in practice. In contrast, Monte Carlo simulations show that the wild bootstrap implementation of our proposed test has an appropriate size and non-trivial power in detecting idiosyncratic bubbles despite the issue caused by spurious factors.

As an empirical illustration, we apply the proposed test to investigate the explosiveness in the panel of cryptocurrencies' daily prices from November 15, 2017, to January 31, 2018. We find no evidence of idiosyncratic bubbles' explosiveness in the cryptocurrency prices during this period and discover that the bubbles' explosiveness in the first estimated factor by the classical time-series approach.

Our paper is closely related to three lines of studies on panel data models.

First, our paper contributes to the growing literature on the panel factor models with possibly explosive factors. Representative works include Horie and Yamamoto (2016), Chen et al. (2019), and Peng et al. (2020). In particular, Horie and Yamamoto (2016) and Chen et al. (2019) also discuss the detection of the bubbles' explosiveness within the framework of panel factor models. It is worth mentioning that Horie and Yamamoto (2016) consider tests for the bubbles' explosiveness in the common factors and idiosyncratic components separately, but their tests for the idiosyncratic components are conducted series-by-series and thus do not take advantage of the cross-section dimension of panel data effectively. In contrast, our test for the explosiveness of the idiosyncratic component is a panel approach and the asymptitic local power analysis of our test reveals that the cross-sectional information indeed enhances the power of our test against the alternatives. In addition, Horie and Yamamoto (2016) recommend determining the working number of factors by existing information criteria. But we find that the PC-based information criteron tends to underestimate the number of factors due to one or two overwhelming dominant factors in the real data. The under-extraction of true factors can result in the oversize of the idiosyncratic tests in Horie and Yamamoto (2016). In contrast, Chen et al. (2019) implement Phillips, Shi and Yu's (2015, PSY hereafter) procedure to test the presence of bubbles in the first estimated factor series by assuming that the idioyscrtic error term processes are stationary. In the case where the idiosyncratic components are mildly explosive but common factors follow unit root or stationary processes, their methodology can potentially attribute the source of bubbles' explosiveness to the common factors due to spurious factors caused by highly persistent series in the idiosyncratic components. In some sense, our work complements that of Chen et al. (2019). Only in the absence of mildly explosive and unit-root-type nonstationary behavior in the idiosyncratic components, can their test have the right interpretation.

Second, our paper is linked with the large literature on nonstationary panel data models. Bai (2004) shows that latent factors, factor loadings, and common components can be estimated consistently by the PC method if all factors follow unit-root processes in the presence of stationary idiosyncratic components. Bai and Ng (2004) propose a novel PANIC procedure to test the null of unit root separately in the common and idiosyncratic components for data of interest. Breitung and Das (2008) explore the theoretical properties and practical performances of the unit root test for panel factor models under various cases. Bai and Ng (2010) provide more alternative approaches to test the null of unit root for the nonstationary panel data models. In this paper, we allow for the bubbles' explosiveness in the common component and the idiosyncratic error component and thus extend the methodology in Bai (2004) and Bai and Ng (2004) to explosive panel data. We show that PC estimation can still yield consistent estimates of unobserved factors and factor loadings for the restricted model under the null. Our proposed test exploits the potentials of PANIC Pool Tests (PPT) in Bai and Ng (2010) to test the idiosyncratic bubbles' explosiveness.

Third, our paper is implicitly tied to panel data models with interactive fixed effects (IFEs); see, e.g., Bai (2009), Moon and Weidner (2015), and Miao et al. (2020). After the quasi-difference transformation of the generic model considered in the paper, the model turns out to be a dynamic panel data model with IFEs and heterogeneous explosive coefficients. Besides, our technical lemmas reveal that the PC estimation is still consistent for the first-differenced form of the restricted model under the null even if the unobserved common factors are mildly explosive but fairly weak. In contrast, Onatski (2009) studies the inconsistency of PC estimation for panel factor models when the unobserved common factors are stationary but very weak.

The remainder of the paper is structured as follows. In Section 2.2, we formally introduce our model, hypotheses, and estimation strategy. Section 2.3 reports the main theoretical results. Section 2.4 presents the information criterion for the determination of the working number of factors. In Section 2.5, we conduct Monte Carlo experiments to evaluate the finite sample performance of the proposed test. We apply the proposed test to study the cryptocurrency daily prices in Section 2.6. Section 2.7 concludes. Proofs of the main results in the paper are relegated to the Appendix B. Further technical details are also provided in the Appendix B.

NOTATION. Throughout the paper we adopt the following notation. For an rectangular real matrix $A \in \mathbb{R}^{m \times n}$, we denote its transpose as A', its trace as trA, its Frobenius norm as $\|A\|$ $\left(\equiv [tr\,(A'A)]^{1/2}\right)$, and its spectral norm as $\|A\|_{\text{sp}}$ $\left(\equiv \sqrt{\varphi_1\,(A'A)}\right)$, where $\equiv$ means "is defined as" and $\varphi_k(\cdot)$ denotes the k-th largest eigenvalue of a real symmetric matrix by counting eigenvalues of multiplicity multiple times. We also use $\varphi_{\min}$ and $\varphi_{\max}$ to stand for the minimum and maximum eigenvalues of a symmetric real matrix. Let diag$(a_1, \ldots, a_m)$ represent a $\mathbb{R}^{m\times m}$ diagonal matrix with entries $a_1, \ldots, a_m$ on its diagonal. We write $A \asymp B$ if there exist some finite positive constants c and C such that $c|A| \leq B \leq C|A|$. Besides, M stands for a generic large positive constant that may vary across lines. The operator $\xrightarrow{P}$ denotes convergence in probability, $\xrightarrow{d}$ convergence in distribution, and $plim$ probability limit. We use $\mathbb{I}(\cdot)$ to denote the usual indicator function. For a full rank $N \times R$ matrix $\boldsymbol{F}$ with $N > R$, we denote the corresponding orthogonal projection matrices as $P_{\boldsymbol{F}} = \boldsymbol{F}\left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1}\boldsymbol{F}$ and $M_{\boldsymbol{F}} = \boldsymbol{I}_N - P_{\boldsymbol{F}}$, where $\boldsymbol{I}_N$ denotes the $N \times N$ identity matrix. Besides, let all the time-series observations for each individual are available from period 1 to period T.

2.2 Basic Framework

2.2.1 The model

We consider the following panel factor model

$$X_{it} = \boldsymbol{\lambda}_i^{0\prime}\boldsymbol{F}_t^0 + e_{it}, \tag{2.2 .1}$$

where $i = 1, \ldots, N$, $t = 1, \ldots, T$, $\boldsymbol{F}_t$ and $\boldsymbol{\lambda}_i$ are $R_0 \times 1$ vectors of factors and factor loadings, respectively, and e_{it} is the idiosyncratic error term. (2.2 .1) also can be rewritten in matrix form

$$\boldsymbol{X} = \boldsymbol{\Lambda}^0\boldsymbol{F}^{0\prime} + \boldsymbol{e}, \tag{2.2 .2}$$

where $\boldsymbol{X} = (X_{it})$ denotes a $N \times T$ matrix. $\boldsymbol{F}^0 = (\boldsymbol{F}_1^0, \ldots, \boldsymbol{F}_T^0)'$ is a $T \times R_0$ matrix of unobserved factors, $\boldsymbol{\Lambda}^0 = (\boldsymbol{\lambda}_1^0, \ldots, \boldsymbol{\lambda}_N^0)'$ is an $N \times R_0$ matrix of factor loadings, and $\boldsymbol{e}$ represents an $N \times T$ matrix of idiosyncratic shocks.

We assume the unobserved common factors are generated as follows:

$$(1 - \boldsymbol{\rho}_0^0 L)\boldsymbol{F}_t^0 = \boldsymbol{C}(\boldsymbol{L})\boldsymbol{u}_t, \tag{2.2 .3}$$

where $\boldsymbol{\rho}_0^0$ =diag$(\rho_{01}^0, \ldots, \rho_{0R_0}^0)$, L is lag operator, $\boldsymbol{C}(L) = (C_1(L), \ldots, C_{R_0}(L))$, $C_r(L) = \sum_{j=0}^{\infty} C_{j,r} L^j$ with $C_{0,r} = 1$ for $r = 1, \ldots, R_0$, and $\boldsymbol{u}_t$ represents the error process in (2.2 .3). Throughout this paper, we consider two cases: (i) $\rho_{0,r}^0 > 1$ for $r = 1, \ldots, R_0$, and (ii) $\{\rho_{0,r}^0\}_{r=1}^{R_0}$ are all equal to 1.

The idiosyncratic components are generated by similar autoregressive processes,

$$(1 - \rho_i^0 L) e_{it} = D_i(L) \epsilon_{it}, \tag{2.2 .4}$$

where ρ_i^0 is the AR coefficient, $D_i(L) = \sum_{j=0}^{\infty} D_{ij} L^j$ with $D_{i0} = 1$ for $\forall i$, and ϵ_{it} represents the error process in (2.2 .4).

2.2.2 Hypotheses and estimation of the restricted model

The null hypothesis of interest is

$$\mathbb{H}_0 : \rho_i^0 = 1 \text{ for } \forall i = 1, \ldots, N. \tag{2.2 .5}$$

The alternative hypothesis is

$$\mathbb{H}_1 : \rho_i^0 > 1 \text{ for some } i. \tag{2.2 .6}$$

That is, we allow for explosive behavior under the alternative but not under the null. In the case of failure to reject the null, we can conclude the explosive behavior in the data is driven by that of the common factors. Under the alternatives, both the common factors and the idiosyncratic component can contribute to the explosiveness of $\{X_{it}\}$.

To construct a residual-based test statistic for $\mathbb{H}_0$, we propose to estimate the factors and factor loadings under the null and obtain the residuals based on (2.2 .1). For brevity, throughout this paper, let $C_r(L) = 1$ for $r = 1, \ldots, R_0$ and $D_i(L) = 1$ for $i = 1, \ldots, N$. Under the null, we can rewrite the restricted model as follows:

$$\begin{aligned} X_{it} &= \boldsymbol{\lambda}_i^{0\prime}\boldsymbol{F}_t^0 + e_{it}, \\ \boldsymbol{F}_t^0 &= \boldsymbol{\rho}_0^0\boldsymbol{F}_{t-1}^0 + \boldsymbol{u}_t, \\ e_{it} &= \rho_i^0 e_{it-1} + \epsilon_{it}. \end{aligned} \tag{2.2 .7}$$

Remark 2.1. As mentioned in Section 2.1, the model under study is closely linked with panel data models with IFEs. Note that

$$\begin{aligned} X_{it} &= \rho_i^0 X_{it-1} + \boldsymbol{\lambda}_i^{0\prime}\boldsymbol{F}_t^0 - \rho_i^0\boldsymbol{\lambda}_i^{0\prime}\boldsymbol{F}_{t-1}^0 + \epsilon_{it} \\ &= \rho_i^0 X_{it-1} + \boldsymbol{\delta}_i^{0\prime}\boldsymbol{G}_t^0 + \epsilon_{it}, \end{aligned} \tag{2.2 .8}$$

where $\boldsymbol{\delta}_i^0 = (\boldsymbol{\lambda}_i^{0\prime}, \rho_i^0\boldsymbol{\lambda}_i^{0\prime})'$ and $\boldsymbol{G}_t^0 = (\boldsymbol{F}_t^0, \boldsymbol{F}_{t-1}^{0\prime})'$. When $\rho_i = \rho < 1$ for all i and all factors in $\boldsymbol{F}_t^0$ are stationary, the above model specifies a linear dynamic panel data model with IFEs studied by Moon and Weidner (2015) and Lu and Su (2016), among others. When $\rho_i > 1$ or =1 and the factors are nonstationary or mildly explosive, one may be tempted to estimate ρ_i^0 along with the factors and factor loadings based on (2.2 .8). But we find that it is extremely challenging, if possible at all, to study the asymptotic properties of such estimators. In this paper, we follow Bai and Ng (2004) and Lu and Su (2016) consider the estimation of the factors based on the first-diffference transformation of X_{it} instead. We allow the possible explosiveness of factors but do not need to assume the factors to be strong. We follow the lead of De Mol et al. (2008), Onatski (2012) and Freyaldenhoven (2019) and define the intensity of factors via the convergence rate of cummulative factor loadings and find that the factors can be estimated consistently as long as the intensity is not low.

Let $Z_{it} = X_{it} - X_{it-1}$. Noting that under $\mathbb{H}_0$, we have

$$\boldsymbol{f}_t^0 \equiv \boldsymbol{F}_t^0 - \boldsymbol{F}_{t-1}^0 = (\boldsymbol{\rho}_0 - \boldsymbol{I}_{R_0})\,\boldsymbol{F}_{t-1}^0 + \boldsymbol{u}_t \equiv \boldsymbol{B}_t^0 + \boldsymbol{u}_t, \tag{2.2 .9}$$

and

$$Z_{it} = \boldsymbol{\lambda}_i^{0\prime} \boldsymbol{f}_t^0 + \epsilon_{it} = \boldsymbol{\lambda}_i^{0\prime} \boldsymbol{B}_t^0 + \boldsymbol{\lambda}_i^{0\prime} \boldsymbol{u}_t + \epsilon_{it}. \tag{2.2 .10}$$

Based on (2.2 .10), we consider the PCA estimation of the factors $\boldsymbol{B}_t^0$ and factor loadings $\boldsymbol{\lambda}_i^0$ by assuming the true number of factors R_0 is known and by treating the sum, $\boldsymbol{\lambda}_i^{0\prime} \boldsymbol{u}_t + \epsilon_{it}$, as a pseudo error term. That is, we pretend to estimate $\boldsymbol{B}_t^0$ instead of $\boldsymbol{f}_t^0$ in the case of explosive factors. This will faciliate the analysis when the factors are mildly explosive so that $\boldsymbol{B}_t^0$ dominates $\boldsymbol{u}_t$. But such a treatment does not affect the immediate and final results as long as the rotation matrix for the PC estimates of factors and factor loadings is appropriately defined. For related work, see Peng et al. (2020) for a more general treatment for mixed and general factors in panel data models with IFEs. We will propose an information criterion to determine the working number of factors in Section 2.4.

Let $\boldsymbol{Z}_i = (Z_{i1}, \ldots, Z_{iT})'$ and $\boldsymbol{Z} = (\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_N)'$. Following Bai and Ng (2004), we apply the PC estimation to $\Sigma_{Z,N} \equiv N^{-1} \boldsymbol{Z}' \boldsymbol{Z}$. Let $(\widehat{\boldsymbol{B}}, \widehat{\boldsymbol{\Lambda}})$ denote the solution to the following minimization problem:

$$\min_{\boldsymbol{B},\boldsymbol{\Lambda}} \frac{1}{N} \left(\boldsymbol{Z} - \boldsymbol{\Lambda} \boldsymbol{B}'\right)' \left(\boldsymbol{Z} - \boldsymbol{\Lambda} \boldsymbol{B}'\right),$$

$$\text{s.t. } \mathcal{J}_B^{-2} \boldsymbol{B}' \boldsymbol{B} = \boldsymbol{I}_{R_0} \text{ and } \boldsymbol{\Lambda}' \boldsymbol{\Lambda} = \text{ diagnoal matrix}, \tag{2.2 .11}$$

where $\mathcal{J}_B$ is a user-specific choice normalization constant.

Remark 2.2. In the case where the factors follow a unit root process, the above PCA estimates the stationary factor difference $\boldsymbol{f}_t^0 (= \boldsymbol{u}_t)$ and the factor loadings, the ideal choice of $\mathcal{J}_B$ is $\sqrt{T}$ (see Bai and Ng (2004)). But we allow $\boldsymbol{F}_t^0$ to be mildly explosive here. In this case, we allow $\mathcal{J}_B$ to be any scalar that is larger than $\sqrt{T}$ in order (e.g., T). As shown in Appendix B.1, the convergence rate of factor loadings estimates will be affected by $\mathcal{J}_B$ while the convergence rate of factor estimate is not affected by $\mathcal{J}_B$ for the restricted model under the null.

Given $\widehat{\boldsymbol{B}}$ in (2.2 .11), we obtain the estimate of factor loadings $\boldsymbol{\lambda}_i^0$ by $\widehat{\boldsymbol{\lambda}}_i = (\widehat{\boldsymbol{B}}' \widehat{\boldsymbol{B}})^{-1} \widehat{\boldsymbol{B}}' \boldsymbol{Z}_i$ for

$i = 1, ..., N$. Define the residual $\hat{\epsilon}_{it} = Z_{it} - \widehat{\boldsymbol{\lambda}}_i' \widehat{\boldsymbol{B}}_t$. Define for $t = 2, \ldots, T$,

$$\hat{e}_{it} = \sum_{s=2}^{t} \hat{\epsilon}_{is}, \text{ and } \widehat{\boldsymbol{F}}_t = \sum_{s=2}^{t} \widehat{\boldsymbol{B}}_s. \tag{2.2 .12}$$

We will show that $\widehat{\boldsymbol{B}}_t$ estimates $\boldsymbol{f}_t^0$ consistently up to a well-defined rotation matrix $\boldsymbol{H}$. So $\widehat{\boldsymbol{F}}_t$ serves as an estimate of $\boldsymbol{F}_t^0$.

2.2.3 A test statistic for detecting bubbles in the idiosyncratic components

To proceed, we add some notations. Let $\hat{\boldsymbol{e}}_i = (\hat{\boldsymbol{e}}_{i1}, \ldots, \hat{\boldsymbol{e}}_{iT})'$ and $\hat{\boldsymbol{e}}_{i,-1} = (\hat{\boldsymbol{e}}_{i,0}, \ldots, \hat{\boldsymbol{e}}_{i,T-1})'$. Let $\hat{\boldsymbol{e}} = (\hat{\boldsymbol{e}}_1, \ldots, \hat{\boldsymbol{e}}_N)'$ and $\hat{\boldsymbol{e}}_{-1} = (\hat{\boldsymbol{e}}_{1,-1}, \ldots, \hat{\boldsymbol{e}}_{N,-1})'$. Define

$$\hat{\rho} = \frac{\text{tr}\left(\hat{\boldsymbol{e}}_{-1}'\hat{\boldsymbol{e}}\right)}{\text{tr}\left(\hat{\boldsymbol{e}}_{-1}'\hat{\boldsymbol{e}}_{-1}\right)}, \text{ and } \overline{\sigma_\epsilon^4} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \sigma_i^4, \text{ and } \overline{\sigma_\epsilon^2} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2,$$

where $\sigma_i^2 = \text{Var}(\epsilon_{it})$.

Following the lead of Bai and Ng (2010), we define the so-called PANIC Pooled Tests (PPT) statistic as follows:

$$PPT = \frac{\overline{\sigma_\epsilon^2}}{\sqrt{2}\left(\overline{\sigma_\epsilon^4}\right)^{1/2}} \sqrt{N} T \left(\hat{\rho} - 1\right). \tag{2.2 .13}$$

Under the null hypothesis $\mathbb{H}_0$, we will show that PPT is asymptotically standard normally distributed. This result is not surprising when both the common factors and idiosyncratic components follow unit-root processes: Bai and Ng (2010) obtain the same limiting null distribution for such as case. But our major focus is on the case where the common factors are mildly explosive. It is good that we still have the standard normal limiting null distribution in this case.

The test statistic PPT in (2.2 .13) is infeasible because $\overline{\sigma_\epsilon^2}$ and $\overline{\sigma_\epsilon^4}$ are unavailable in practice. To obtain a feasible test statistic, we can obtain the consistent estimates of $\overline{\sigma_\epsilon^2}$ and $\overline{\sigma_\epsilon^4}$ as follows:

$$\widehat{\overline{\sigma_\epsilon^2}} = \frac{2}{NT^2} \sum_{i=1}^{N} \sum_{t=2}^{T} \sum_{s=1}^{t-1} \hat{\epsilon}_{is}^2 \text{ and } \widehat{\overline{\sigma_\epsilon^4}} = \frac{2}{NT^2} \sum_{i=1}^{N} \sum_{t=2}^{T} \sum_{s=1}^{t-1} \hat{\epsilon}_{is}^2 \hat{\epsilon}_{it}^2.$$

A feasible version of PPT is defined as follows:

$$\widetilde{PPT} = \frac{\widehat{\overline{\sigma_\epsilon^2}}}{\sqrt{2}\left(\widehat{\overline{\sigma_\epsilon^4}}\right)^{1/2}} \sqrt{N}T\left(\hat{\rho}-1\right). \tag{2.2 .14}$$

$\widetilde{PPT}$ is asymptotically equivalent to PPT provided that we can establish the consistency of $\widehat{\overline{\sigma_\epsilon^2}}$ and $\widehat{\overline{\sigma_\epsilon^4}}$.

2.3 Asymptotic Properties

In this section, we first present a set of basic assumptions to derive the asymptotic properties of the statistic PPT. Then we study the asymptotic null distribution and local power properties of PPT.

2.3.1 Basic assumptions

Let $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \ldots, \epsilon_{Nt})'$, $\boldsymbol{u}_t = (u_{1t}, \ldots, u_{1R_0})'$, $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1', \ldots, \boldsymbol{\epsilon}_T')'$ and $\boldsymbol{u} = (\boldsymbol{u}_1', \ldots, \boldsymbol{u}_T')'$. Let $\mathcal{F}_{NT,t}^u \equiv \sigma\left(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_t, \boldsymbol{\epsilon}, \boldsymbol{\Lambda}^0\right)$ denote the minimal σ-algebra with $\left(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_t, \boldsymbol{\epsilon}, \boldsymbol{\Lambda}^0\right)$. Similarly, let $\mathcal{F}_{NT,t}^{\epsilon,i} \equiv \sigma(\{\epsilon_{jt}, \ldots, \epsilon_{j1}\}_{j=1}^N, \{\epsilon_{j,t+1}\}_{j=1,j\neq i}^N, \boldsymbol{u}, \boldsymbol{\Lambda}^0)$. Let $\gamma_N(s,t) = \sum_{i=1}^N E\left(\epsilon_{is}\epsilon_{it}\right)$ and $\zeta_{s,t} = E\left|N^{-1/2}\sum_{i=1}^N\left[\epsilon_{is}\epsilon_{it} - E\left(\epsilon_{is}\epsilon_{it}\right)\right]\right|^4$.

Assumption A2.1. *(a) The process $\{(\boldsymbol{\epsilon}_t, \boldsymbol{u}_t), t \geq 0\}$ is α-mixing across t with mixing coefficient $\alpha_{ij}(|t-s|)$ between $\{\epsilon_{it}\}$ and $\{\epsilon_{js}\}$ and mixing coefficient $\alpha_i(|t-s|)$ between $\{(\epsilon_{it}, \boldsymbol{u}_t)\}$ and $\{(\epsilon_{is}, \boldsymbol{u}_s)\}$. Assume that*

$$\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T\left(\alpha_{ij}(t)\right)^{\delta/(4+\delta)} = O(N), \quad \sum_{i=1}^N\sum_{j=1}^N\left(\alpha_{ij}(0)\right)^{\delta/(4+\delta)} = O(N), \text{ and } \max_i \alpha_i(t) = O\left(t^{-\theta}\right),$$

where $\delta > 0$ is chosen such that $E\left\|\omega_{it}\right\|^{4+\delta} < M$ with $\omega_{it} \in \{\boldsymbol{\lambda}_i^0, \boldsymbol{u}_t, \epsilon_{it}\}$ and $\theta > (4+\delta)/\delta$.

(b) $E\left(\boldsymbol{u}_t \mid \mathcal{F}_{NT,t-1}^u\right) = 0$ a.s., $E\left(\boldsymbol{u}_t\boldsymbol{u}_t' \mid \mathcal{F}_{NT,t-1}^u\right) = E\left(\boldsymbol{u}_t\boldsymbol{u}_t'\right) = \Sigma_u$, and $0 < \underline{c} < \varphi_{\min}\left(\Sigma_u\right) \leq \varphi_{\max}\left(\Sigma_u\right) < \bar{c} < \infty$, where $\underline{c}$ and $\bar{c}$ are some generic finite constants.

(c) $E\left(\varepsilon_{it} \mid \mathcal{F}_{NT,t-1}^{\epsilon,i}\right) = 0$ *a.s., and* $E\left(\varepsilon_{it}^2 \mid \mathcal{F}_{NT,t-1}^{\epsilon,i}\right) = E\left(\varepsilon_{it}^2\right) = \sigma_i^2 \leq M$ *a.s. for some finite constant* M*.*

(d) $E\left\|\boldsymbol{F}_0^0\right\|^2 \leq M$*.*

Assumption A2.2. *(a)* $\boldsymbol{\lambda}_i^0$ *is random such that* $E\left\|\boldsymbol{\lambda}_i^0\right\|^2 \leq M$*, (b)* $N^{-p}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0 \xrightarrow{p} \Sigma_\lambda$ *such that such that* $0 < \underline{c} < \varphi_{\min}\left(\Sigma_\lambda\right) \leq \varphi_{\max}\left(\Sigma_\lambda\right) < \bar{c} < \infty$*, (c)* $\frac{1}{2} < p \leq 1$*.*

Assumption A2.3. *(a)* $\max_t \sum_{s=1}^{T}\left|\gamma_N\left(s,t\right)\right| = O(N)$*, (b)* $\max_{s,t}\zeta_{s,t} \leq M$ *, (c)* $\max_i E\left|e_{i0}\right| \leq M$*, (d)* $\left\|\boldsymbol{\epsilon}\right\|_{sp} = O_p(\sqrt{N} + \sqrt{T})$*.*

Assumption A2.4. $\{\boldsymbol{\lambda}_i^0\}, \{\boldsymbol{u}_t\}$ *, and* $\{\epsilon_{js}\}$ *are three groups of mutually independent stochastic variables for every* (i, j, s, t)*.*

Assumption A2.5. *(a)* $\rho_{0,r}^0 = 1 + \frac{c_r}{\kappa_T}$ *for* $r = 1, ..., R_0$*, where* $c_r \geq 0$*,* $k_T \to \infty$*,* $k_T/T \to 0$ *and* $|c_{r,max} - c_{r,min}| = O_p\left(\frac{\kappa_T}{T}\right)$ *with* $c_{r,min} = \min\{c_1, \ldots, c_{R_0}\}$ *and* $c_{r,max} = \max\{c_1, \ldots, c_{R_0}\}$ *. (b)* $C_r(L) = 1$ *for* $\forall r$ *and* $D_i(L) = 1$ *for* $\forall i$*.*

Assumption A2.6. $\Upsilon_1 \equiv$*plim*$_{(N,T)\to\infty}\left(\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}\right)^{-1}\left(\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\right)\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0}{N^p}\left(\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}\right)\left(\boldsymbol{\rho}_0^0\right)^{-2T}$ *exists such that* $0 < \underline{c} < \varphi_{\min}\left(\Upsilon_1\right) < \varphi_{\max}\left(\Upsilon_1\right) < \bar{c} < \infty$ *for all* $\widehat{\boldsymbol{B}} \in \{\boldsymbol{B} : \mathcal{J}_B^{-1}\boldsymbol{B}'\boldsymbol{B} = I_{R_0}\}$*.*

Assumption A2.1, A2.2, and A2.3 impose moments conditions on the error terms, factors, factor loadings. They are widely used in the literature; see, e.g., Bai and Ng (2004) and Peng et al. (2020). In particular, the martingale difference sequence (m.d.s.) condition in Assumption A2.1(b)-(c) simplifies our theoretical analysis. It can be removed at the cost of more complicated analysis. In Assumption A2.2(b), the quantity p represents the strength of the factors. When $p = 1$, the factors are pervasive and thus termed as strong factors. This is the case that has been widely considered in the literature; see, e.g., Bai and Ng (2002), Bai (2003), and Fan et al. (2013). When $p < 1$, the factors are coined as weak factors; see, De Mol et al. (2008), Onatski (2012), and Freyaldenhoven (2019), among others.

Assumption A2.4 follows Assumption D in Bai and Ng (2004). Assumptions A2.5(a)-(b) are similar to Assumption 5 in Horie and Yamamoto (2016). Assumptions A2.5(b) will allow us to ease the exposition of our analysis. Moreover, Assumption A2.6 is a technical assumption for the theoretical analysis.

Remark 2.3. Assumption A2.5(a) allows for the general case that the degrees of explosiveness in factors are different. According to the results in the next subsection and the Appendix B, $\{\rho_{0r}^{0}\}_{r=1}^{R_0}$ plays no role in the asymptotic null distribution and the asymptotic power property of the proposed test statistic explicitly, although it plays a role in some immediate results. Thus, the different degrees of explosiveness in the factors do not affect the first order asymptotic properties of the proposed test statistic. However, in practice, heterogeneous explosiveness in factors will matter for the proposed test's finite-sample performance. The intuition behind this statement is straightforward as follows.

When the true number of factors R_0 is not available, we can obtain the working number of factors $\hat{R}$ by employing existing information criteria. Without loss of generality, under the alternatives, we suppose $\{\rho_{0r}^{0}\}_{r=1}^{R_0}$ are widespread in the interval $[1.01, 1.09]$, and $\{\rho_i^0\}_{i=1}^{N}$ are all larger than 1. In finite samples, the sample eigenvalues from the common components can be very close to the sample eigenvalues from idiosyncratic components. In the random matrix theory, this means that the spikes of the covariance matrix $(NT)^{-1}\boldsymbol{Z}'\boldsymbol{Z}$ are not well separated from the bulks. As pointed out by Dobriban (2017), such phenomena are universal in high dimensional data. Thus, $\hat{R}$ usually will be overestimated under the alternatives. The more factors are extracted from the data under the alternative, the more under-rejections the proposed test statistic might have. In a related study, Onatski and Wang (2020) document that the classical left-tail unit root test for panel data models severely overrejects the null of unit root because of the over-extraction of factors in their simulation studies.

2.3.2 Asymptotic null distribution

To establish the null distribution of PPT, we add the following assumption.

Assumption A2.7. *(a)* $\kappa_T N^{1/2}/T \to 0$.

(b) (1) $p > 1/2$ *if* $\rho_{0,r}^0 > 1$ *for all* $r = 1, \ldots, R_0$, *and (2)* $p = 1$ *if* $\rho_{0,r}^0 = 1$ *for all* $r = 1, \ldots, R_0$.

Assumption A2.7(a) impose conditions on the rates at which N and T pass to infinity. Assumption A2.7(a) is not restrictive compared to $N/T \to 0$ used in Bai and Ng (2010) and Westerlund (2015). We can even allow N and T to pass to infinity at identical rates. For example, Horie and Yamamoto (2016) mention that a typical form of κ_T is $\kappa_T = T^\tau$ with $0 < \tau < 1$, when τ is appropriately small, Assumption A2.7(a) holds obviously even if N and T pass to infinity at identical rates. Bai and Ng (2010) and Westerlund (2015) obtain the asymptotic normality of PPT through the asymptotics of linear processes despite they require $N/T \to 0$ to deal with additional terms in the the Beveridge-Nelson decomposition. Similarly, Assumption A2.7(a) above is the cost we pay to the additional terms due to the presence the mildly explosive factors in our theoretical analysis. Besides, our Monte Carlo simulations in Section 2.5 show that our bootstrap-based test performs well in terms of well-controlled size and high power against the alternatives for various combinations of N and T.

Assumption A2.7(b) imposes the lower limit $1/2$ for p when factors are mildly explosive. This condition is necessary for deriving the null distribution of PPT. Assumption A2.7(b) implies that the intensity of signals from the common components should be moderately strong compared to that of signals from the idiosyncratic components, although unobserved common factors are mildly explosive. Under Assumption A2.7(b), we do not need to distinguish whether factors are mildly explosive or nonstationary when we implement the proposed test in practice. This greatly enlarges the potential scope for the application of our test.

The following theorem states the asymptotic null distribution of PPT.

Theorem 2.3.1. *Suppose that Assumptions A2.1–A2.7 hold. Then as* $(N,T) \to \infty$, $PPT \xrightarrow{d} \mathcal{N}(0,1)$ *under* $\mathbb{H}_0$.

As a direct consequence of Theorem 2.3.1, we can readily obtain the asymptotic normality of $\widetilde{PPT}$ in (2.2 .14). The corollary below states this result formally.

Corollary 2.3.2. *Suppose that Assumptions A2.1–A2.7 hold. Then as* $(N,T) \to \infty$, $\widetilde{PPT} \xrightarrow{d} \mathcal{N}(0,1)$ *under* $\mathbb{H}_0$.

Given the result in Corollary 2.3.2 and noticing that our test is a one-sided test, we can reject the null when $\widetilde{PPT}$ is sufficiently large, say larger than the associated normal critical value at a given significance level.

2.3.3 Asymptotic power property

To analyze the asymptotic local power property of PPT, we consider the following sequence of Pitman local alternatives:

$$\mathbb{H}_{1NT} : \rho_i = 1 + \frac{c_i}{TN^{1/2}} \text{ for } i = 1, ..., N.$$

Assumption A2.8 specifies conditions on c_i's that contribute to the nontrivial local power of our test.

Assumption A2.8. *(a)* $c_i \geq 0$ *for each* i, *(b)* $\mu \equiv \lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} c_i \sigma_i^2 > 0$.

Assumption A2.8(a) implies that we restrict our attention to the one-sided local laternative, namely, we only consider mildly explosive idiosycratic error terms under the local alternative. Assumption A2.8(b) implies that the alternatives cannot be sparse in order for the concentration parameter μ to be positive. If there are only $o(N)$ individual time series $\{e_{it}, t \geq 0\}$ that exhibit mildly explosiveness with $c_i > 0$, the concentration parameter μ is zero and our test will lose power in this case. In the case of such sparse alternatives, one could consider alternative supremum-type of test, or augment our test with such a sup-test. See, e.g., Fan et al. (2015). We leave this as a future work.

Theorem 2.3.3 below studies the asymptotic local power property of our test.

Theorem 2.3.3. *Suppose that Assumptions A2.1-A2.8 hold. Then as* $(N, T) \to \infty$, *we have* $\widetilde{PPT} \xrightarrow{d} \mathcal{N}(\bar{\mu}, 1)$ *under* $\mathbb{H}_{1NT}$, *where* $\bar{\mu} = \mu \frac{\overline{\sigma_\epsilon^2}}{\sqrt{2}\left(\overline{\sigma_\epsilon^4}\right)^{1/2}}$.

Theorem 2.3.3 indicates that our PPT test has nontrivial power to detect local alternative converging to the null at rate $T^{-1}N^{-1/2}$. In contrast, the series-by-series test of Horie and Yamamoto (2016) can only detect local alternatives converging to the null at rate T^{-1}. This indicates the benefit of pulling all cross-section units together to conduct a joint test for the possible explosive behavior in the idiosyncratic errors.

It is worth mentioning the non-centrality parameter $\bar{\mu}$ in Theorem 2.3.3 is different from that in Westerlund (2015). The main reasons are as follows. In the current paper, $\{c_i\}$ are treated as the constant, and further we allow for the Pitman's local alternatives converging at the parametric rate $T^{-1}N^{-1/2}$. Contrarily, Westerlund (2015) assumes $\{c_i\}$ are random variables, and accommodates Pitman's local alternatives converging at the parametric rate $T^{-1}N^{-v}$ with $v \geq 0$.

2.3.4 A bootstrap version of the test

The estimates of the factors and factor loadings may converge to the null slowly in the presence of weak factors. To improve the finite sample performance of our test, we propose a bootstrap version of our test. The test is in the spirit of fixed-regressor wild bootstrap that has been widely used in the literature. The algorithm is given below.

Algorithm 2.1: Wild Bootstrap

1. Given R_0, obtain the PC estimator $(\widehat{\boldsymbol{B}}, \widehat{\boldsymbol{\Lambda}})$ under (2.2 .10), and calculate $\widetilde{PPT}$ based on (2.2 .14).

2. Obtain the bootstrap errors $\epsilon_{it}^* = \hat{\epsilon}_{it}\varsigma_{it}$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, where ς_{it}'s are i.i.d. Racdmercher sequences such that ς_{it} has a 0.5 chance of being 1 and a 0.5 chance of being -1. Generate the bootstrap analogue Z_{it}^* of Z_{it} by holding $(\widehat{\boldsymbol{B}}, \widehat{\boldsymbol{\Lambda}})$ as fixed in the bootstrap world: $Z_{it}^* = \widehat{\boldsymbol{\lambda}}_i' \widehat{\boldsymbol{B}}_t + \epsilon_{it}^*$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$.

3. Obtain the PC estimator $(\widehat{\boldsymbol{B}}^*, \widehat{\boldsymbol{\Lambda}}^*)$ based on $\{Z_{it}^*\}$ and calculate $\widetilde{PPT}^*$, the bootstrap analogoue of $\widetilde{PPT}$, based on this bootstrap estimate and $\{Z_{it}^*\}$.

4. Repeat Step 2 and 3 B times and denote the resulting bootstrap test statistics as $\{\widetilde{PPT}_b^*\}_{b=1}^B$. Obtain the bootstrap $p-$value as $p^* = B^{-1}\sum_{b=1}^B \mathbb{I}\left\{\widetilde{PPT}_b^* > \widetilde{PPT}\right\}$, where $\mathbb{I}\{\cdot\}$ is the usual indicator function.

When R_0 is unknown, we can estimate it by $\hat{R}$ by using Algorithm 2.2 in the next section. Then we use $\hat{R}$ to replace R_0 in the above algorithm. This will not create any theoretical problem under the null but may cause power loss under the alternative.

The following theorem establishes the asymptotic validity of the above bootstrap test.

Theorem 2.3.4. *Suppose that Assumptions A2.1–A2.7 hold. Then* $\widetilde{PPT}^* \xrightarrow{D^*} N(0,1)$ *in probability, where* $\xrightarrow{D^*}$ *denotes weak convergence under the bootstrap probability measure conditional on the observed sample* $\boldsymbol{Z} = \{Z_{it}\}$.

Theorem 2.3.4 shows that the bootstrap provides an asymptotic valid approximation to the limit null distribution of $\widetilde{PPT}$. This holds because we generate the bootstrap data by imposing the null hypothesis in Step 2.

Remark 2.4. As indicated by Su and Wang (2017), the fixed-design wild bootstrap scheme may only apply to the case that $\{e_{it}\}$ have no cross-sectional dependence or only exhibit weak cross-sectional dependence. Hence, Su and Wang (2017) propose an alternative bootstrap scheme to accommodate mildly or strong cross-sectional dependence among $\{e_{it}\}$. The key idea is to use $\Sigma_{\hat{e}}$, defined as $\Sigma_{\hat{e}} = \frac{1}{T}\sum_t \hat{\boldsymbol{e}}_t\hat{\boldsymbol{e}}_t'$, to capture the cross-sectional dependence in bootstrap world. However, such treatment may not apply to our current study because $\Sigma_{\hat{e}}$ fails to intimate the original cross-sectional dependence accurately when common factors $\{\boldsymbol{F}_t\}$ or idiosyncratic components $\{e_{it}\}$ follow explosive processes. We leave it for our future work to extend the alternative bootstrap procedure in Su and Wang (2017) to panel factor models with mildly explosive factors.

2.4 Determination of the Working Number of Factors

In this section, we first present an information criterion to determine the working number of factors for the proposed test and then discuss the effects of the working number of factors on the size and power of our proposed test.

2.4.1 Transformed double ridge ratio criterion

To implement the PPT test in practice, one generally needs to estimate the true number of factors R_0. We aim to propose a criterion function based on the eigenvalues associated with the first-differenced data matrix $\boldsymbol{Z}$. Let $R_{max} \geq R_0$ denote the largest possible number of factors. Let c_{ridge} and δ_{ridge} be two small but positive ridge parameters.

We follow the lead of Xia et al. (2017) and Zhu et al. (2020) and propose a *transformed double ridge ratio* (TDRR) criterion to determine the working number of factors. The procedure is summarized in Algorithm 2.2 below.

Algorithm 2.2: Transformed Double Ridge Ratio (TDRR) Criterion

1. Let $\varphi_{NT,k}$ be the k-th eigenvalue of the matrix $(NT)^{-1}\sum_{i=1}^{N}\boldsymbol{Z}_i'\boldsymbol{Z}_i$. Calculate

$$\varphi_{NT,k}^{*} = \ln\left(1 + \frac{\varphi_{NT,k}}{W_{k-1}}\right)$$

   where $W_k = \sum_{r=k+1}^{\min(N,T)} s_{NT,r}$.

2. Calculate

$$\varphi_{NT,k}^{**} = \frac{\varphi_{NT,k}^{*} + c_{ridge}}{\varphi_{NT,k+1}^{*} + c_{ridge}} - 1 \text{ and } DRR_k = \frac{\varphi_{NT,k+1}^{**} + \delta_{ridge}}{\varphi_{NT,k}^{**} + \delta_{ridge}},$$

   where c_{ridge} and δ_{ridge} are two small positive ridge parameters to avoid devision by zero.

3. Estimate R_0 by $\hat{R} = \tilde{R} + 1$ where $\tilde{R} = \arg\max_{1\leq k\leq R_{max}} DRR_k$.

It is a practical challenge to avoid underestimation of the working number of factors in the current framework. If some genuine factors are overwhelmingly strong, other geniune ones become relatively weak so that it is extremely hard for them to be detected by some prevailing methods such as the variants of the scree test that search for a gap separating small from large eigenvalues. Dobriban and Owen (2019) coin this phenomenon as *shadowing* and indicate that shadowing is common in financial data like stock returns. We also witness shadowing in our unreported simulations when mildly explosive common factors $\{\boldsymbol{F}_t^0\}$ or idiosyncratic components $\{e_{it}\}$ are present.

Algorithm 2.2 is motivated from the thresholding double ridge ratio (TDRR) approach of Zhu et al. (2020). To alleviate the longstanding problem of underestimation of the dimensionality of a model via existing eigenvalue-based criteria, Zhu et al. (2020) propose the TDRR criterion that can proivde a consistent estimate even when there are several local minima or weak factors. A reason behind the underestimation is that some largest estimated eigenvalues are often dominating and the other estimated eigenvalues are close to each other in magnitude, no matter whether they are nonzero or not at the population level. The TDRR criterion relies on the eigenvalue compression that can, to certain extent, alleviate the eigenvalue domination effect. It can further make the ratio at the true order be well separated from the other ratios by better exaggerating the difference beween the ratio at the true dimension and the other ratios. In comparison with the TDRR approach in Zhu et al. (2020), the main modification in Algorithm 2.2 lies in the monotonic transformation of eigenvalues in Step 1. Specifically, we adopt the monotonic transformation suggested by Xia et al. (2017) to narrow the difference between adjacent eigenvalues sufficiently in the absolute magnitude. Besides, unlike the approach in Zhu et al. (2020), we find the location of maximum value among $\{DDR_k\}$ because such treatment leads to more reliable choices of the working number of factors in our unreported simulations.

Under the null, we can prove the consistency of $\hat{R}$ in Algorithm 2.2 for $R_0 \geq 2$ by following the proof strategy in Zhu et al. (2020). To achieve compression, the monotonic transformation of eigenvalues in Zhu et al. (2020) is deliberately chosen, our choice of deliberately-chosen monotonic transformation in Algorithm 2.2 plays the same role in dealing with the effect of overwhelmingly dominant factors. We refer the readers to Zhu et al. (2020) for details.

Remark 2.5. We have some remarks on Algorithm 2.2 as follows.

(1) The TDRR is designed to estimate the true number of factors when $R_0 \geq 2$. In case of $R_0 = 1$, Step 3 in Algorithm 2.2 implies $\hat{R} \geq 2$ holds and thus $\hat{R} > R_0$ for sure. But this is not a severe issue under the null, although the over-extraction of factors is likely to spoil the power of the proposed test statistic under the alternatives. Our Monte Carlo simulations in Section

2.5 indicate that a wild bootstrap implementation of the proposed test statistic still exhibits good size and power properties in almost all cases under our investigation.

(2) In our unreported simulations, except for the monotonic transformation used in Step 1 in Algorithm 2.2, other monotonic transformation functions such as $arctan(\cdot)$ and standard normal distribution function $\Phi(\cdot)$ are also considered. When other functions are used, under the null, the test statistic also has desired size and the rate of choosing R_0 is very high. However, the power of the proposed test statistic is not as satisfactory as that with the current choice of monotonic transformation under the alternatives. Therefore, Algorithm 2.2 is proposed to balance the trade-offs between the size and power of the proposed test statistic.

(3) A user-specified parameter R_{max} is set to avoid outliers in DRR_k for $k = 1, \ldots, \min(N, T)$ in Algorithm 2.2 like other information criteria such as PC and IC in Bai and Ng (2002).

(4) Following Zhu et al. (2020), we recommend setting the two ridge parameters as follows: $c_{ridge} = \ln T/(10\sqrt{T})$ and $\delta_{ridge} = \ln T/(5\sqrt{T})$. The optimal choices of ridge parameters or the data-driven choices of ridge parameters are left for future research.

2.4.2 The effects of the working number of factors on the size and power of the proposed test

In this subsection, we reason that it is unnecessary to know the true number of factors (R_0) to implement our proposed test in Algorithm 2.1 for detecting the bubble's explosiveness in idiosyncratic components in practice. When we use the working number of factors $\hat{R}$ in Algorithm 2, we expect that our proposed test still performs well with the correct size under the null and non-trivial power against the alternatives.

Let $\Delta e_{it} = e_{it} - e_{it-1} = (\rho_i - 1)e_{it-1} + \epsilon_{it}$. Under the null hypothesis, $\Delta e_{it} = \epsilon_{it}$ and by by (2.2 .10), thus

$$Z_{it} = \underbrace{\boldsymbol{\lambda}_i^{0\prime}\left(\boldsymbol{B}_t^0 + \boldsymbol{u}_t\right)}_{R_0 \text{ true factors}} + \epsilon_{it}. \qquad (2.4\ .1)$$

In practices, the true number of factors R_0 is usually unknown. Then we can apply Algorithm 2.2 to obtain the working number $\hat{R}$ of factors. Obviously, when $\hat{R} = R_0$ holds, PPT is supposed to have desired size and non-trivial power against the alternatives. When $\hat{R} > R_0$, we conjecture that we can extend the work of Moon and Weidner (2015) and Lu and Su (2016) to our framework and show that the null distribution of PPT is unchanged. Intuitively, when $\hat{R} > R_0$, the residuals $\{\hat{\epsilon}_{it}\}$ still maintain the same first order statistical properties as the original sequences $\{\epsilon_{it}\}$, and as result, the PPT statistic that is based on the residuals $\{\hat{\epsilon}_{it}\}$ also follows the same null distribution. On the other hand, if $\hat{R} < R_0$ is chosen under the null, as long as the factors are mildly explosive, the residuals $\{\hat{e}_{it}\}$ will contain information on omitted factors and thus tend to be mildly explosive too, leading to the over-rejections of the test.

Under the alternative hypothesis, $\rho_i > 1$ for some i, the exact determination of the number of true factors becomes difficult due to the occurrence of spurious factors as documented by Onatski and Wang (2020). Note that $\Delta e_{it} = (\rho_i - 1)e_{it-1} + \epsilon_{it} \neq \epsilon_{it}$ when $\rho_i > 1$ and

$$Z_{it} = \underbrace{\boldsymbol{\lambda}_i^{0\prime}\left(\boldsymbol{B}_t^0 + \boldsymbol{u}_t\right)}_{R_0 \text{ true factors}} + \underbrace{\Delta e_{it}}_{\text{error term with high persistency}}. \tag{2.4 .2}$$

It is easy to see $\{\Delta e_{it}, t \geq 1\}$ may be highly persistent when $\{e_{it}, t \geq 0\}$ is mildly explosive. This can also be verified by regressing Δe_{it} on e_{it-1}. As a result, the PC estimation suffers from the serious issue of spurious factors. As Onatski and Wang (2020) argue, difference-stationary series can be approximated by Wiener processes, and much of the variation in Δe_{it} can be captured by a few of the trigonometric functions corresponding to the first few terms in the Karhunen-Loève (KL) expansion. Following Onatski and Wang (2020), we can rewrite (2.4 .2) as follows

$$Z_{it} = \underbrace{\boldsymbol{\lambda}_i^{0\prime}\left(\boldsymbol{B}_t^0 + \boldsymbol{u}_t\right)}_{R_0 \text{ true factors}} + \underbrace{\boldsymbol{\Gamma}_i^{\prime}\boldsymbol{W}_t}_{\text{spurious factor structure with } K \text{ factors}} + \text{spurious error terms}, \tag{2.4 .3}$$

where $\boldsymbol{W}_t$ represents a few of the trigonometric functions. By the extensive Monte Carlo simulations in Onatski and Wang (2020) for highly persistent time series, K often equals to 2 or 3 in practice.

Under the alternatives, once the working number of factors $\hat{R}$ satisfies $\hat{R} < R_0 + K$, the residuals $\{\hat{e}_{it}\}$ defined in (2.2 .12) are still explosive to drive the PPT test statistic to diverge to infinity. In general, as along as $\hat{R} < R_0 + K$, PPT still has non-trivial power against the alternatives. Our simulation results in Section 2.5 imply $\hat{R}$ obtained from Algorithm 2.2 typically meets such a requirement, which ensures the good finite sample power property of our test.

2.5 Monte Carlo Simulations

In this section we evaluate the finite sample performance of the proposed test, we explore different setups of factors, such as, mildly explosive common factors with the same degree of explosiveness, mildly explosive factors with the different degrees of explosiveness, and factors that follow unit root processes.

2.5.1 Data generating processes

In this section, we consider the following data generating process (DGP):

$$\begin{aligned} X_{it} &= \boldsymbol{\lambda}_i^{0\prime} \boldsymbol{F}_t^0 + e_{it}, \\ \boldsymbol{F}_t^0 &= \boldsymbol{\rho}_0^0 \boldsymbol{F}_{t-1}^0 + \boldsymbol{u}_t, \\ e_{it} &= \rho_i^0 e_{it-1} + \epsilon_{it} \end{aligned} \tag{2.5 .1}$$

where $\boldsymbol{\rho}_0^0 \equiv diag\left(\rho_{0,1}^0, \ldots, \rho_{0,R_0}^0\right)$ is an $R_0 \times R_0$ diagonal matrix. All data are generated by (2.5 .1) with different settings of parameters. the nominal level is 0.05.

DGP 1 $\lambda_i \overset{i.i.d}{\sim} \mathcal{N}(0,1)$, $\epsilon_{it} \overset{i.i.d}{\sim} \mathcal{N}(0,1)$, $\boldsymbol{u}_t \overset{i.i.d}{\sim} \mathcal{N}(0, \boldsymbol{I}_{R_0})$, and $\rho_i = 1$ for $i = 1, \ldots, N$.

DGP 2 $\lambda_i \overset{i.i.d}{\sim} \mathcal{N}(0,1)$, $\epsilon_{it} \overset{i.i.d}{\sim} \mathcal{N}(0,1)$, $\boldsymbol{u}_t \overset{i.i.d}{\sim} \mathcal{N}(0, \boldsymbol{I}_{R_0})$, and $\rho_i \overset{i.i.d}{\sim}$ *Uniform* $(1.08, 1.10)$ for $i = 1, \ldots, N$.

Apparently, DGPs 1 and 2 are used to evaluate the size and power of the PPT test respectively. For the true number of factors, we consider two cases: (a) $R_0 = 3$, and (b) $R_0 = 1$, which correspond to the 3-factor Fama-French factor model and the 1-factor capital asset pricing model (CAPM) in financial studies, respectively.

In case (a), we consider the following specifications of $\boldsymbol{\rho}_0^0$ in both DGPs:

(1) $\boldsymbol{\rho}_0^0 = diag\,(1,1,1)$; (2) $\boldsymbol{\rho}_0^0 = diag\,(1.02, 1, 02, 1, 02)$; (3) $\boldsymbol{\rho}_0^0 = diag\,(1.04, 1.04, 1.04)$; (4) $\boldsymbol{\rho}_0^0 = diag\,(1.08, 1.08, 1.08)$; (5) $\boldsymbol{\rho}_0^0 = diag\,(1.02, 1.04, 1.06)$; (6) $\boldsymbol{\rho}_0^0 = diag\,(1.02, 1.05, 1.08)$.

In case (b), $\boldsymbol{\rho}_0^0$ degenerates to a scalar ρ_0^0 and we consider the the following four specifications of $\boldsymbol{\rho}_0^0$ in both DGPs: (1) $\rho_0^0 = 1$, (2) $\rho_0^0 = 1.02$, (3) $\rho_0^0 = 1.04$, (4) $\rho_0^0 = 1.08$.

When we estimate R_0 by Algorithm 2.2, we set $R_{\max} = 6$ when $R_0 = 3$ and $R_{\max} = 3$ when $R_0 = 1$. For each scenario, we conduct 500 replications with $B = 300$ bootstrap resamples in each replication. We consider the 5% nominal level in all cases.

2.5.2 Simulation Results

Table 2.1: Finite sample properties of the PPT under the null and alternatives when $R_0 = 3$ that is taken as known.

	Size of the wild bootstrap implementation of the proposed test					
DGP 1	$\rho_i = 1$ for $i = 1, \ldots, N$					
$\boldsymbol{\rho}_0^0 \backslash (N,T)$	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
$diag\,(1.00, 1.00, 1.00)$	0.052	0.048	0.048	0.054	0.048	0.042
$diag\,(1.02, 1.02, 1.02)$	0.044	0.048	0.048	0.042	0.048	0.042
$diag\,(1.04, 1.04, 1.04)$	0.048	0.058	0.054	0.058	0.048	0.044
$diag\,(1.08, 1.08, 1.08)$	0.058	0.040	0.056	0.048	0.054	0.042
$diag\,(1.02, 1.04, 1.06)$	0.044	0.046	0.042	0.040	0.044	0.044
$diag\,(1.02, 1.05, 1.08)$	0.056	0.060	0.046	0.048	0.044	0.044
	Power of the wild bootstrap implementation of the proposed test					
DGP 2	$\rho_i \overset{i.i.d}{\sim}$ *Uniform* $(1.08, 1.10)$ for any $i \in \{1, \ldots, N\}$					
$\boldsymbol{\rho}_0^0 \backslash (N,T)$	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
$diag\,(1.00, 1.00, 1.00)$	0.878	1.000	1.000	0.990	1.000	1.000
$diag\,(1.02, 1.02, 1.02)$	0.530	0.918	1.000	0.556	0.902	1.000
$diag\,(1.04, 1.04, 1.04)$	0.506	0.930	1.000	0.544	0.930	1.000
$diag\,(1.08, 1.08, 1.08)$	0.774	0.994	1.000	0.792	0.994	1.000
$diag\,(1.02, 1.04, 1.08)$	0.470	0.916	1.000	0.590	0.950	1.000
$diag\,(1.02, 1.05, 1.08)$	0.546	0.948	1.000	0.576	0.964	1.000

Tables 2.1 and 2.3 report the performance of the PPT test under various scenarios when the true

number of factors R_0 is known and given by 3 and 1, respectively. For each setup of the common factors, the tables exhibit the results for each combination of N and T. The results suggest that the wild bootstrap implementation of our proposed test has decent power and well-controlled size in all cases. It is worth mentioning that we also report the performance of the proposed test based on the asymptotic critical value in the Appendix B, the proposed test based on the asymptotic critical value suffers from downward size distortions in most cases. These results indicate that the bootstrap version of the proposed test can avoid possible size distortions.

Table 2.2: Finite sample properties of the PPT under the null and alternatives when $R_0 = 3$ that is taken as unknown and estimated by Algorithm 2.2.

	Size of the wild bootstrap implementation of the proposed test					
DGP 1	$\rho_i = 1$ for $i = 1, \ldots, N$					
$\boldsymbol{\rho}_0^0 \backslash (N,T)$	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
$diag\,(1.00, 1.00, 1.00)$	0.046	0.046	0.048	0.046	0.052	0.054
$diag\,(1.02, 1.02, 1.02)$	0.042	0.048	0.044	0.044	0.058	0.044
$diag\,(1.04, 1.04, 1.04)$	0.050	0.052	0.044	0.044	0.046	0.044
$diag\,(1.08, 1.08, 1.08)$	0.046	0.058	0.052	0.040	0.052	0.046
$diag\,(1.02, 1.04, 1.06)$	0.044	0.044	0.050	0.044	0.040	0.052
$diag\,(1.02, 1.05, 1.08)$	0.048	0.040	0.054	0.040	0.044	0.042
	Power of the wild bootstrap implementation of the proposed test					
DGP 2	$\rho_i \overset{i.i.d}{\sim}$ *Uniform* $(1.08, 1.10)$ for any $i \in \{1, \ldots, N\}$					
$\boldsymbol{\rho}_0^0 \backslash (N,T)$	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
$diag\,(1.00, 1.00, 1.00)$	0.800	0.840	0.948	0.822	0.882	0.992
$diag\,(1.02, 1.02, 1.02)$	0.694	0.872	0.998	0.780	0.876	0.998
$diag\,(1.04, 1.04, 1.04)$	0.602	0.776	0.876	0.760	0.786	0.920
$diag\,(1.08, 1.08, 1.08)$	0.548	0.704	1.000	0.636	0.770	0.994
$diag\,(1.02, 1.04, 1.06)$	0.698	0.650	0.916	0.628	0.706	0.926
$diag\,(1.02, 1.05, 1.08)$	0.512	0.780	0.888	0.686	0.736	0.924

Tables 2.2 and 2.4 show the performance of the proposed PPT test under various scenarios when the true number of factors R_0 is estimated by Algorithm 2.2 for the cases with $R_0 = 3$ and 1, respectively. For each setup of the common factors, the tables present the results for each combination of N and T. According to these results, we see that the wild bootstrap implementation of our proposed test still has satisfactory performances in terms of decent power and reasonable size. This implies that the devised criterion presented in Algorithm 2.2 works well in a variety of scenarios in the context

of the current paper. As pointed out in the remark 2.5, Algorithm 2.2 overestimates R_0 for sure with $R_0 = 1$, however, those results in Tables 2.2 and 2.4 show the wild bootstrap implementation of our proposed test still maintains appropriate size and has high power in most cases. In addition, even if it is pervasive that only one common factor plays a role in economic or finance models, the results in Table 2.4 imply our proposed test is still applicable in spite of the overestimation of the true number of factors by Algorithm 2.2.

Table 2.3: Finite sample properties of the PPT under the null and alternatives when $R_0 = 1$ that is taken as known.

	Size of the wild bootstrap implementation of the proposed test					
DGP 1	$\rho_i = 1$ for $i = 1, \ldots, N$					
$\rho_0^0 \backslash (N, T)$	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
1.00	0.052	0.046	0.046	0.050	0.054	0.046
1.02	0.040	0.056	0.048	0.054	0.044	0.044
1.04	0.040	0.040	0.050	0.048	0.040	0.050
1.08	0.046	0.044	0.044	0.052	0.044	0.056
	Power of the wild bootstrap implementation of the proposed Test					
DGP 2	$\rho_i \overset{i.i.d}{\sim}$ *Uniform* $(1.08, 1.10)$ for any $i \in \{1, \ldots, N\}$					
$\rho_0^0 \backslash (N, T)$	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
1.00	0.478	1.000	1.000	0.566	1.000	1.000
1.02	0.430	1.000	1.000	0.452	1.000	1.000
1.04	0.558	1.000	1.000	0.606	1.000	1.000
1.08	0.438	1.000	1.000	0.520	1.000	1.000

2.6 Empirical Studies

The first cryptocurrency, namely Bitcoin, was invented by Satoshi Nakamoto during the 2008 subprime crash and gained much attention from the public, regulators, and academic scholars since 2010. For this past decade, the total number of cryptocurrencies has risen to more than 2500 due to Bitcoin's success. The first wave of bubbles in Bitcoin price popped in 2011, and then several speculative bubbles in cryptocurrencies' daily prices are observed and discussed at length. In very recent years, a growing number of studies in finance and economics focus on the detection of bubble behaviors in cryptocurrencies' daily prices by utilizing time-series data of Bitcoin or other cryptocurrencies'

Table 2.4: Finite sample properties of the PPT under the null and alternatives when $R_0 = 1$ that is taken as unknown and estimated by Algorithm 2.2.

	Size of the wild bootstrap implementation of the proposed test					
DGP 1	$\rho_i = 1$ for $i = 1, \dots, N$					
$\rho_0^0 \backslash (N, T)$	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
1.00	0.040	0.056	0.054	0.048	0.042	0.044
1.02	0.042	0.056	0.048	0.040	0.046	0.044
1.04	0.042	0.048	0.040	0.044	0.044	0.056
1.08	0.040	0.044	0.048	0.054	0.044	0.044
	Power of the wild bootstrap implementation of the proposed test					
DGP 2	$\rho_i \overset{i.i.d}{\sim}$ *Uniform* $(1.08, 1.10)$ for any $i \in \{1, \dots, N\}$					
$\rho_0^0 \backslash (N, T)$	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
1.00	1.000	1.000	1.000	1.000	1.000	1.000
1.02	0.904	1.000	1.000	0.982	0.998	1.000
1.04	0.898	0.952	1.000	0.934	0.950	1.000
1.08	0.780	0.998	1.000	0.952	1.000	1.000

daily prices based on the right-tail unit root test like the PWY approach proposed in Phillips et al. (2011), and the PSY approach developed in Phillips et al. (2015). Cheung et al. (2015) identify both short-lived and giant bubbles from 2011 to 2014. Hafner (2020) also confirms explosive speculative bubbles in cryptocurrencies when the time-varying volatility exists. Harvey et al. (2020) locate the staring and end date of bubbles in Bitcoin daily prices around the end of 2017 by a variant of the PSY approach. Enoksen et al. (2020) disclose the recent bubbles in several cryptocurrencies daily prices during the mid of November of 2017 to January of 2018. Geuder et al. (2019) also find similar bubbles' explosiveness in Bitcoin daily prices in such a time period.

Thus, in this section, we apply our test procedure to examine whether or not idiosyncratic bubbles exist during the period of bubbles' explosiveness in cryptocurrencies' daily prices. Enoksen et al. (2020) point out the lack of empirical studies on the predictors of bubbles in cryptocurrencies' prices. Therefore, it is natural to treat the fundamentals in pricing the cryptocurrencies as unobserved common factors when we study the speculative bubbles in cryptocurrencies' daily prices.

2.6.1 Data

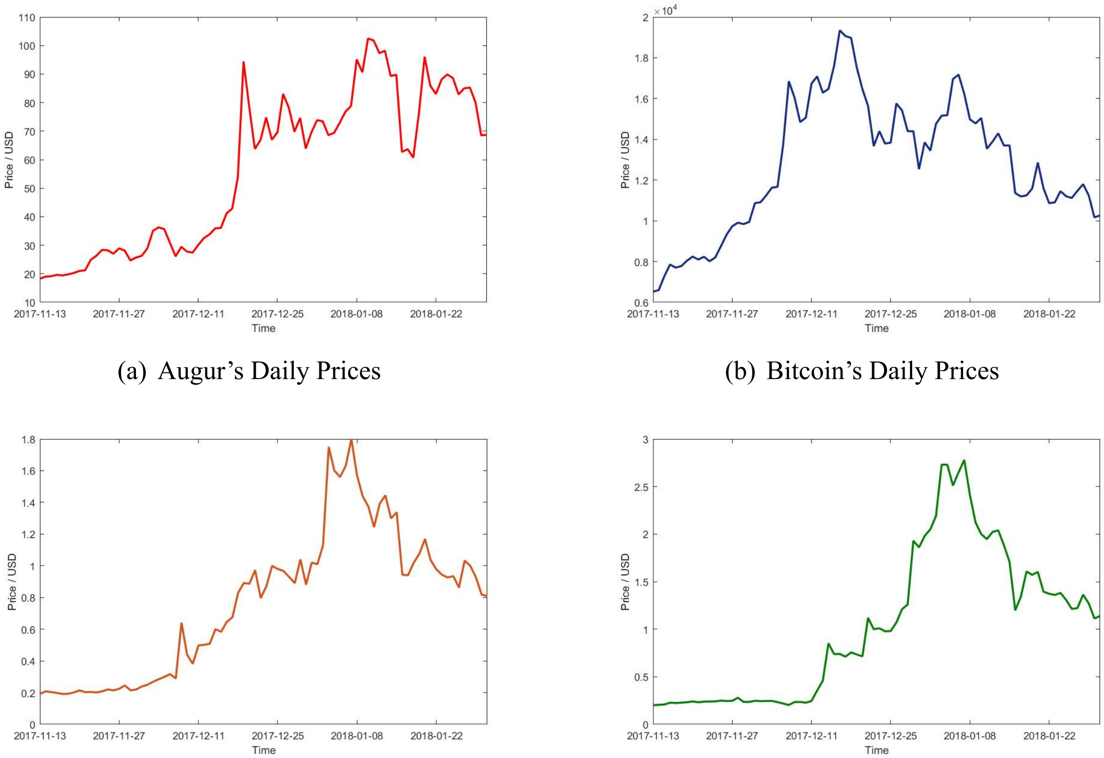
We use cryptocurrencies' daily prices collected from Investing.com. Data consist of 26 cross-section units covering 77 time-series observations between November 15, 2017, to January 31, 2018[1]. The collected cryptocurrencies are chosen from top 100 cryptocurrencies listed in investing.com based on their total market capitalization and data availability. We list names of these cryptocurrencies in the alphabetic order and enumerate them by numbers as follows: (1) 0x, (2) Augur, (3) Basic Attention Token, (4) Binance Coin (5) Bitcoin Case, (6) Bitcoin Gold, (7) Bitcoin, (8) Dash, (9) DigiByte, (10) Dogecoin, (11) EOS, (12) Ethereum, (13) ICON, (14) IOTA, (15) Litecoin, (16) Monero, (17) NEM, (18) Neo, (19) OMG, (20) Qtum (21) Stellar, (22) Tether, (23) VeChain, (24) Waves, (25) XRP, (26) Zcash.

As the preliminary inspect below, we present the time-series plot of daily prices for several cryptocurrencies from the mid of November 2017 to the end of January 2018 in Figure 1. Moreover, we regress $\Delta Price_{it}$ on $Price_{it-1}$ to obtain the AR coefficients for each cryptocurrency. These coefficients range from 0.9636 to 1.0075. These initial results imply speculative explosive bubbles in cryptocurrencies' daily prices.

2.6.2 Testing for cryptocurrencies' prices with existing approaches

Thanks to the contributions by Enoksen et al. (2020) and Harvey et al. (2020), we can use the PWY approach by Phillips et al. (2011) to double-check the explosiveness in each cryptocurrency's prices for the given period and eliminate those cryptocurrencies that exhibit no bubbles' behaviors in daily prices by the series-by-series checks. Then, to verify the explosiveness for the panel of remaining cryptocurrencies' daily prices, we adapt the approach used in Chen et al. (2019) to have a double-check. Specifically, we apply the PC estimation to the logarithm of the remaining eleven cryptocurrencies' daily prices under the normalization condition that satisfies $N^{-1}\,\boldsymbol{\Lambda}'\boldsymbol{\Lambda} = \boldsymbol{I}$, and then use the PWY approach again for the first component in estimated common factors $\hat{\boldsymbol{F}}$.

[1] We follow findings in Harvey et al. (2020) and Enoksen et al. (2020) to collect cryptocurrencies' daily prices for such a period



(a) Augur's Daily Prices

(b) Bitcoin's Daily Prices

(c) NEM's Daily Prices

(d) XRP's Daily Prices

Figure 2.1: Cryptocurrencies' Daily Prices

The upper part of Table 2.5 displays the values of PWY test statistics for each cryptocurrency. We also compute and report the finite-sample 5% significance-level critical value for the collected panel of cryptocurrencies' daily prices based on 2000 repetitions in the last row of Table 2.5. We use the star in the bracket behind values to indicate this value exceeds the critical value. From tests results in the upper part of the table, we can see that 11 out of 26 cryptocurrencies' daily prices have bubbles' explosiveness in the given period. The lower part of Table 2.5 shows that the data are still explosive as the whole for the panel of remaining cryptocurrencies' daily prices. We can then apply our proposed test to the panel of remaining 11 cryptocurrencies' daily prices in the next subsection.

2.6.3 Empirical results and the robustness check

Following existing studies on cryptocurrencies' daily prices, we apply our proposed test to the logarithm of daily prices. As shown in Table 2.6, we report testing results based on the working numbers

Table 2.5: Test statistics based on the PWY approach and the finite-sample critical value

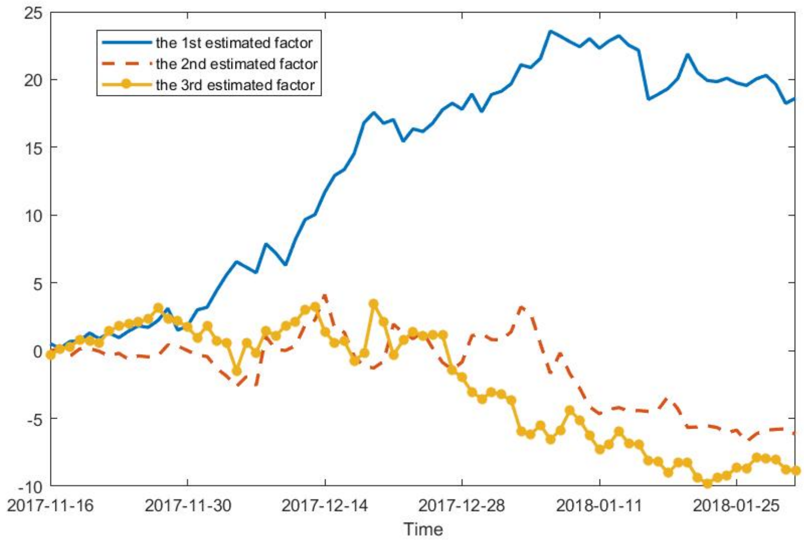
Name Order	Test Statistics	Name Order	Test Statistics	Name Order	Test Statistics
(1)	1.8984(*)	(10)	0.9308	(19)	1.1704
(2)	2.1408(*)	(11)	0.7824	(20)	2.4668(*)
(3)	0.8557	(12)	-0.4676	(21)	1.0480
(4)	2.6428(*)	(13)	-0.5231	(22)	-2.7338
(5)	0.9947	(14)	2.3078(*)	(23)	2.5009(*)
(6)	-1.2213	(15)	3.1612(*)	(24)	3.0959(*)
(7)	2.5281(*)	(16)	0.5027	(25)	4.5272(*)
(8)	0.4006	(17)	1.4546(*)	(26)	0.9568
(9)	0.9428	(18)	0.4818		
critical value	1.1912		**# explosive prices**	11	
Test Statistic of the First Factor for the Panel of Remaining 11 Cryptocurrencies' Daily Prices					
Test statistic for the first factor			1.9616(*)		

of factors estimated by Algorithm 2.2. Additionally, for robustness checks, namely, to avoid overestimating the true number of factors, we also report results for the decreased working number of factors. For our empirical application here, let $k_{max} = \lceil 2 \times \min(\sqrt{N}, \sqrt{T}) \rceil$, where $\lceil M \rceil$ represents the integer that is nearest to M. Specifically, in our empirical studies, $k_{max} = 10$ when we use all 26 cryptocurrencies, and $k_{max} = 7$ when we use remaining 11 cryptocurrencies. We set the bootstrap sample $B = 1000$.

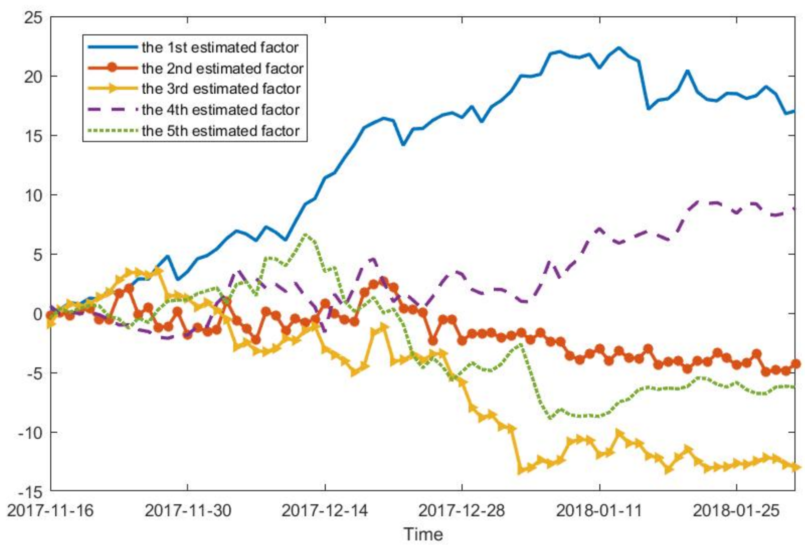
According to testing results by our proposed test statistic PPT in Table 2.6, we do not reject the null of unit root for series in idiosyncratic components. In other words, conditional on collected data, we found no idiosyncratic bubble's explosiveness for the panel of cryptocurrencies' daily prices from November 15, 2017, to January 31, 2018. Robustness checks in Table 2.6 imply the bubbles' explosiveness in the panel of cryptocurrencies' daily prices is driven by the fundamentals, namely, the unobserved common factors. PC estimates of unobserved common factors are consistent in this case, as indicated in Section 2.2.2. To verify the implication from robustness checks, we can apply the PWY approach to estimated common factors. The testing results for the PWY test in Table 2.6 shows that bubbles' explosiveness exists in the estimated first common factor only. Further, to show the time-series behavior of estimated factors, we depict the time-series plot for these estimated factors as below.

Table 2.6: Testing Results for the panel of cryptocurrencies' daily prices

Testing results for remaining 11 cryptocurrencies' explosive daily prices					
PPT for idiosyncratic components	$\hat{R}$	**Robustness Checks**			
The Working number of factors	3	2	1	-	-
Test Statistics	-1.1384	-0.8975	-1.2157	-	-
Bootstrap p-value	0.7160	0.6300	0.6330	-	-
PWY Test Statistic of Estimated Factors					
Test Statistic for the 1st Factor					2.2061(*)
Test Statistic for the 2nd Factor					0.8341
Test Statistic for the 3rd Factor					-0.1274
critical value					1.1912
Testing results for all 26 cryptocurrencies' daily prices					
PPT for idiosyncratic components	$\hat{R}$	**Robustness Checks**			
The Working number of factors	5	4	3	2	1
Test Statistics	-2.8940	-2.9098	-2.5599	-1.0124	-1.8645
Bootstrap p-value	0.9770	0.9897	0.9380	0.6530	0.8350
PWY Test Statistic of Estimated Factors					
Test Statistic for the 1st Factor					1.6420(*)
Test Statistic for the 2nd Factor					-2.0372
Test Statistic for the 3rd Factor					1.1589
Test Statistic for the 4th Factor					1.0344
Test Statistic for the 5th Factor					0.0763
critical value					1.1912



(a) Estimated Factors for remaining 11 cryptocurrencies' explosive daily prices



(b) Estimated Factors for all 26 cryptocurrencies' daily prices

Figure 2.2: Estimated Factors

Figures 2.2(a) and 2.2(b) clearly reveal that the first estimated factor behaves explosively and thus servers as the fundamental to trigger the explosive increases in cryptocurrencies' daily prices for the period we study. This is a further investigation and extension of findings in Enoksen et al. (2020) and Harvey et al. (2020).

2.7 Concluding Remarks

In this paper, we propose a consistent test for detecting bubbles in idiosyncratic components in the presence of nonstationary or mildly explosive factors in common components in panel factor models. We establish the limit null distribution, asymptotic power property, and the consistency of our test under mild conditions. Monte Carlo simulations demonstrate that our proposed test has approximately correct size and discriminatory power against the alternatives. The most important takeaway from the analysis in this paper is that it should be the first and essential step to test the bubbles' explosiveness in idiosyncratic components if researchers try to use PC estimation to disentangle the sources of the bubbles' explosiveness in data. Our proposed test provides a simple and effective approach to detect bubbles' explosiveness from idiosyncratic components by exploiting the cross-sectional information in panel data, which is parallel with the classical panel analysis of nonstationarity in idiosyncratic and common components. As an empirical illustration, We apply our proposed test to cryptocurrencies' daily prices. We find no idiosyncratic bubbles in cryptocurrencies' daily prices by our proposed test and verify the bubbles' explosiveness in unobserved common factors by the classical time-series approach.

Chapter 3

Detection of Bubbles in Common Factors with Local-to-unity Errors

3.1 Introduction

Detecting the bubbles in financial assets or portfolios is of great interest for both practitioners and researchers. The local-to-unity explosive bubbles play an essential role in related studies since the seminal work by Phillips (1987) and Phillips (1988) over the past decades; for recent studies, see e.g., Harvey et al. (2015), Dou and Müller (2019), Whitehouse (2019), and Bykhovskaya and Phillips (2020). On the other hand, it is a common practice to employ panel factor models to uncover the sources of nonstationarity of data since the high dimensional factor analysis introduced by Bai (2003), Bai (2004) and Bai and Ng (2004). We aim to bring these two lines of research together and propose an easy-to-implement test to detect bubbles in common factors for the explosive panel data.

This paper portrays fundamentals and peculiar characteristics as common factors and idiosyncratic components in the framework of large dimensional panel factor models as done in Chapter 2. Despite that the intensity of signals from the fundamentals and individual features is similar to each other when both the common factors and idiosyncratic components follow the local-to-unity process, we find that we can still obtain the consistent estimator of unobserved factors and factor loadings via the PC estimation for the first difference form of data. Therefore, our findings enable practitioners to delve into the local-to-unity explosive bubbles in financial and economic data in the context of panel factor models.

Although the genuine factors are unobserved in practice, the PC estimates of common factors motivate us to devise a bubble test for detecting local-to-unity explosiveness in fundamentals. To be concrete, we construct the right-tailed unit root tests built upon those estimated common factors for detecting bubbles in fundamentals when the idiosyncratic error terms can follow a unit-root or local-to-unity process. We establish the limiting null distribution and the asymptotic local power property of the proposed tests. In particular, we show that our tests have non-trivial power to detect those bubbles in unobserved common factors under the alternative of local-to-unity.

However, unlike the Single-Factor model discussed in Chen et al. (2019), for the multi-factor structure, the proposed test based on the PC estimation of common factors is generally not asymptotically pivotal under the null because of an array of nuisance parameters. We show that the proposed test involves the weighted sum of independent standard Brownian motion, where the weights rely upon the asymptotic matrix of the rotation matrix arising from the PC estimation. The presence of such unknown nuisance parameters is the price we pay to replace those unobserved common factors with estimated ones. To implement the proposed test in practice, we can follow the literature and propose using the dependent wild bootstrap (DWB) method to simulate the critical values and improve our tests' finite sample performance. We justify the DWB method's validity in the sense of bootstrap consistency and power under the local-to-unity explosive common factors settings.

Accordingly, our findings contribute to following two lines of current literature.

First, our paper follows lines of classical studies on factor analysis of nonstationary panel data; see e.g., Bai (2004) and Bai and Ng (2004). In these studies, the validity of PC estimations holds when unobserved common factors follow unit-root processes, whereas the idiosyncratic error terms remain stationary. In particular, to disentangle the sources of nonstationarity of data, related studies prompt the PC estimation for the first-differenced data. As a natural extension of Bai and Ng (2004) to accommodate local-to-unity explosiveness in panel data, our paper can allow both the common factors and the idiosyncratic error components to follow the local-to-unity process. Furthermore, we present the consistency of the PC estimations when both the common factors and idiosyncratic components follow the local-to-unity process.

Second, the present paper relates to the growing literature on the high dimensional factor analysis of the explosive data; see e.g., Horie and Yamamoto (2016), Chen et al. (2019), and Chapter 2. For example, Chen et al. (2019) considers a Single-Factor model to use Phillips, Shi and Yu's (2015, PSY hereafter) procedure directly to detect speculative bubbles in the first estimated factor series when the common factors are mildly explosive and idiosyncratic error terms are stationary. However, on the one hand, researchers have to impose a Single-Factor model by faith and restrict their attentions on the one fundamental only in empirical studies; on the other hand, it can be unnecessary to rule out the presence of local-to-unity explosive bubbles in the idiosyncratic error terms. Although Feng and Su (2020) propose to test bubbles in the idiosyncratic error terms based on the PC estimation in the presence of unobserved nonstationary or explosive factors and use wild bootstrap to improve the finite-sample performance of their test, the detection of bubbles in common factors is not covered formally. As a necessary and useful complement to the bubbles detection in panel data, our current paper concentrates on the case that both common factors and idiosyncratic error terms follow the local-to-unity process. Therefore, this paper fills in the gap in the literature to address the impacts of the local-to-unity explosive process on the PC estimation and bubble detection in panel factor models.

The remainder of the paper is structured as follows. In Section 3.2, we formally introduce our model, hypotheses, and estimation strategy. Section 3.3 reports the main theoretical results. Section 3.4 discusses the model selection issues concerning the implementation of the proposed tests. In Section 3.5, we conduct Monte Carlo experiments to evaluate the finite sample performance of the proposed test. Section 3.6 concludes. Proofs of the main results in the paper are relegated to the Appendix C. Further technical details are also provided in the Appendix C.

NOTATION. Throughout the paper we adopt the following notation. For an rectangular real matrix $A \in \mathbb{R}^{m\times n}$, we denote its transpose as A', its trace as tr$A$, its Frobenius norm as $\|A\|$

$\left(\equiv [\text{tr}(A'A)]^{1/2}\right)$, and its spectral norm as $\|A\|_{\text{sp}} \left(\equiv \sqrt{\varphi_1(A'A)}\right)$, where $\equiv$ means "is defined as" and $\varphi_k(\cdot)$ denotes the k-th largest eigenvalue of a real symmetric matrix by counting eigenvalues of multiplicity multiple times. We also use $\varphi_{\min}$ and $\varphi_{\max}$ to stand for the minimum and maximum eigenvalues of a symmetric real matrix. Let diag$(a_1, \ldots, a_m)$ represent a $\mathbb{R}^{m\times m}$ diagonal matrix with entries $a_1, \ldots, a_m$ on its diagonal. We write $A \asymp B$ if there exist some finite positive constants c and C such that $c|A| \leq B \leq C|A|$. Besides, M stands for a generic large positive constant that may vary across lines. The operator $\xrightarrow{p}$ denotes convergence in probability, $\Rightarrow$ weak convergence, and plim probability limit. We use $\mathbb{I}(\cdot)$ to denote the usual indicator function. For a full rank $N \times R$ matrix $\boldsymbol{F}$ with $N > R$, we denote the corresponding orthogonal projection matrices as $P_{\boldsymbol{F}} = \boldsymbol{F}(\boldsymbol{F}'\boldsymbol{F})^{-1}\boldsymbol{F}$ and $M_{\boldsymbol{F}} = \boldsymbol{I}_N - P_{\boldsymbol{F}}$, where $\boldsymbol{I}_N$ denotes the $N \times N$ identity matrix. Besides, let all the time-series observations for each individual are available from period 1 to period T.

3.2 Basic Framework

3.2.1 The model

We consider the following panel factor model

$$X_{it} = \boldsymbol{\lambda}_i^{0\prime}\boldsymbol{F}_t^0 + e_{it}, \tag{3.2 .1}$$

where $i = 1, \ldots, N$, $t = 1, \ldots, T$, $\boldsymbol{F}_t^0$ and $\boldsymbol{\lambda}_i^0$ are $R_0 \times 1$ vectors of factors and factor loadings, respectively, and e_{it} is the idiosyncratic error term. (3.2 .1) also can be rewritten in matrix form:

$$\boldsymbol{X} = \boldsymbol{\Lambda}^0\boldsymbol{F}^{0\prime} + \boldsymbol{e}, \tag{3.2 .2}$$

where $\boldsymbol{X} = (X_{it})$ denotes a $N \times T$ matrix. $\boldsymbol{F}^0 = (\boldsymbol{F}_1^0, \ldots, \boldsymbol{F}_T^0)'$ is a $T \times R_0$ matrix of unobserved factors, $\boldsymbol{\Lambda}^0 = (\boldsymbol{\lambda}_1^0, \ldots, \boldsymbol{\lambda}_N^0)'$ is an $N \times R_0$ matrix of factor loadings, and $\boldsymbol{e}$ represents an $N \times T$ matrix of idiosyncratic shocks.

We assume the unobserved common factors are generated as follows:

$$(1-\boldsymbol{\rho}_0^0 L)\boldsymbol{F}_t^0 = \boldsymbol{u}_t, \tag{3.2 .3}$$

where $\boldsymbol{\rho}_0^0$ =diag$(\rho_{0,1}^0, \ldots, \rho_{0,R_0}^0)$, L is lag operator, and $\boldsymbol{u}_t$ is the error term.

The idiosyncratic components are generated by similar autoregressive processes,

$$(1-\rho_i^0 L)e_{it} = \epsilon_{it}, \tag{3.2 .4}$$

where $\{\rho_i^0\}_{i=1}^N$ are the AR coefficients, and ϵ_{it} is the error term. Throughout this paper, for $\{e_{it}\}_{t=1}^T$ with $i = 1, \ldots, N$, we focus on the following case: e_{it} is local-to-unity explosive so that (3.2 .4) is a near-to-unit-root model from the explosive side.

Remark 3.1. Except for the local-to-unity explosive case above, there are another two typical cases: (i) e_{it} has a unit root; (ii) e_{it} is near-stationary and (3.2 .4) is a near-to-unit-root model from the stationary side. Both of these two cases can be treated as the special cases we consider above.

3.2.2 Hypotheses and estimation of the restricted model

The null hypothesis of interest is

$$\mathbb{H}_0 : \rho_{0,r}^0 = 1, \text{ for all } r \tag{3.2 .5}$$

The alternative hypothesis is

$$\mathbb{H}_1 : \rho_{0,r}^0 > 1. \text{ for some } r \tag{3.2 .6}$$

That is, we allow for explosive behavior under the alternative but not under the null. In the case of rejections of the null, we can conclude the explosive behavior in the data is driven by some of unobserved common factors.

To construct a test statistic for $\mathbb{H}_0$, we propose to estimate the factors and factor loadings under the null and obtain the residuals based on (3.2 .1). Under the null, we can rewrite the restricted model as follows:

$$\begin{aligned} X_{it} &= \boldsymbol{\lambda}_i^{0\prime} \boldsymbol{F}_t^0 + e_{it}, \\ \boldsymbol{F}_t^0 &= \boldsymbol{\rho}_0^0 \boldsymbol{F}_{t-1}^0 + \boldsymbol{u}_t, \\ e_{it} &= \rho_i^0 e_{it-1} + \epsilon_{it}. \end{aligned} \tag{3.2 .7}$$

Remark 3.2. As mentioned in Feng and Su (2020), the model under study is closely linked with panel data models with interactive fixed effects (IFEs). To see this point, we make a quasi-differencing transformation with respect to X_{it} in (3.2 .7) as follows

$$\begin{aligned} X_{it} &= \rho_i^0 X_{it-1} + \boldsymbol{\lambda}_i^{0\prime} \boldsymbol{F}_t^0 - \rho_i^0 \boldsymbol{\lambda}_i^{0\prime} \boldsymbol{F}_{t-1}^0 + \epsilon_{it} \\ &= \rho_i^0 X_{it-1} + \boldsymbol{\delta}_i^{0\prime} \boldsymbol{G}_t^0 + \epsilon_{it}, \end{aligned} \tag{3.2 .8}$$

where $\boldsymbol{\delta}_i^0 = (\boldsymbol{\lambda}_i^{0\prime}, \rho_i^0 \boldsymbol{\lambda}_i^{0\prime})'$ and $\boldsymbol{G}_t^0 = (\boldsymbol{F}_t^0, \boldsymbol{F}_{t-1}^{0\prime})'$. When $\rho_i = \rho < 1$ for all i and all factors in $\boldsymbol{F}_t^0$ are stationary, the above model specifies a linear dynamic panel data model with IFEs studied by Moon and Weidner (2015) and Lu and Su (2016), among others. When $\rho_i = 1$, (3.2 .8) turns out to be a nonstaioanry dynamic panel data model with IFEs under additional and necessary conditions, for a closely-related study, we refer readers to Huang et al. (2020).

Throughout this paper, we focus on the local-to-unity explosiveness among $\{\rho_i^0\}_{i=1}^N$ and set them to be $\rho_i^0 = 1 + c_i/T$ for $i = 1, \ldots, N$ with the finite and fixed parameters $\{c_i\}_{i=1}^N$. We can treat the unit-root case as a special case with $c_i = 0$ for all i. Following Bai and Ng (2004) and Lu and Su (2016), we consider the PC estimation of the factors based on the first-diffference transformation of X_{it} instead.

Let $Z_{it} = X_{it} - X_{it-1}, \boldsymbol{B}_t^0 \equiv \boldsymbol{F}_t^0 - \boldsymbol{F}_{t-1}^0$. Noting that under $\mathbb{H}_0$, we have

$$\boldsymbol{B}_t^0 = \left(\boldsymbol{\rho}_0^0 - \boldsymbol{I}_{R_0}\right) \boldsymbol{F}_{t-1}^0 + \boldsymbol{u}_t = \boldsymbol{u}_t, \tag{3.2 .9}$$

and

$$Z_{it} = \boldsymbol{\lambda}_i^{0\prime} \boldsymbol{B}_t^0 + e_{it} - e_{it-1} = \boldsymbol{\lambda}_i^{0\prime} \boldsymbol{u}_t + \left(\rho_i^0 - 1\right) e_{it-1} + \epsilon_{it}. \tag{3.2 .10}$$

Based on (3.2 .10), we consider the PC estimation of the factors $\boldsymbol{u}_t$ and factor loadings $\boldsymbol{\lambda}_i^0$ by assuming the true number of factors R_0 is known. Particularly, under (3.2 .5), according to related studies on panel factor models, we can obtain the consistent estimation of the common factors by employing the PC estimation when e_{it} in (3.2 .7) follows a unit-root model, that is, $\{\rho_i^0\}_{i=1}^N$ in (3.2 .7) equal to one; see e.g., Bai and Ng (2004). Under the alternatives (3.2 .6), if the common factors are mildly explosive or nonstationary, Feng and Su (2020) shows that the PC estimation also can yield the consistent estimation of the common factors when $\{\rho_i^0\}_{i=1}^N$ in (3.2 .7) are in the very near vicinity of unity. In this paper, we demonstrate that the PC estimation is still valid if both common factors and idiosyncratic error terms in (3.2 .7) follow a local-to-unity explosive model, that is, both $\{\rho_i^0\}_{i=1}^N$ and $\{\rho_{0,r}^0\}_{r=1}^{R_0}$ in (3.2 .7) lie in the vicinity of unity. Detailed proofs are relegated to the Appendix C due to space limitation.

Let $\boldsymbol{Z}_i = (Z_{i1}, \ldots, Z_{iT})'$ and $\boldsymbol{Z} = (\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_N)'$. Following Bai and Ng (2004), we apply the PC estimation to $\Sigma_{Z,N} \equiv N^{-1}\boldsymbol{Z}'\boldsymbol{Z}$. Let $(\widehat{\boldsymbol{B}}, \widehat{\boldsymbol{\Lambda}})$ denote the solution to the following minimization problem:

$$\min_{\boldsymbol{B},\boldsymbol{\Lambda}} \frac{1}{N} \left(\boldsymbol{Z} - \boldsymbol{\Lambda}\boldsymbol{B}'\right)' \left(\boldsymbol{Z} - \boldsymbol{\Lambda}\boldsymbol{B}'\right),$$

$$\text{s.t. } T^{-1}\boldsymbol{B}'\boldsymbol{B} = \boldsymbol{I}_{R_0} \text{ and } \boldsymbol{\Lambda}'\boldsymbol{\Lambda} = \text{ diagnoal matrix}, \tag{3.2 .11}$$

Remark 3.3. In the case where the factors and idiosyncratic error terms follow a unit root process, the above PC estimates the stationary factor difference $\boldsymbol{B}_t^0 (= \boldsymbol{u}_t)$ and the factor loadings, and the rescale parameter T^{-1} in (3.2 .11) is necessary (e.g., see Bai (2003)). However, Feng and Su (2020) demonstrate that the rescale parameter T^{-1} in (3.2 .11) can be replaced by any polynomial functions of T when the unobserved common factors are mildly explosive. Here, we continue to set the rescale parameter in (3.2 .11) to be T^{-1} because technical treatments of the local-to-unity case is similar to those of the unit-root case in Bai and Ng (2004).

Given $\widehat{\boldsymbol{B}}$ in (3.2 .11), we obtain the estimate of factor loadings $\boldsymbol{\lambda}_i^0$ by $\widehat{\boldsymbol{\lambda}}_i = (\widehat{\boldsymbol{B}}'\widehat{\boldsymbol{B}})^{-1}\widehat{\boldsymbol{B}}'\boldsymbol{Z}_i$ for $i = 1, ..., N$. Define the residual $\widehat{\epsilon}_{it} = Z_{it} - \widehat{\boldsymbol{\lambda}}_i'\widehat{\boldsymbol{B}}_t$. Define for $t = 2, \ldots, T$,

$$\widehat{e}_{it} = \sum_{s=2}^{t} \widehat{\epsilon}_{is}, \text{ and } \widehat{\boldsymbol{F}}_t = \sum_{s=2}^{t} \widehat{\boldsymbol{B}}_s. \tag{3.2 .12}$$

In the Appendix C, we will show that $\widehat{\boldsymbol{B}}_t$ estimates $\boldsymbol{B}_t^0$ consistently up to a well-defined rotation matrix $\boldsymbol{H}$. So $\widehat{\boldsymbol{F}}_t$ serves as an estimate of $\boldsymbol{F}_t^0$.

3.2.3 A test statistic for detecting bubbles in the idiosyncratic components

To proceed, we add some notations. Let $\widehat{\boldsymbol{B}}^{(r)}$ and $\widehat{\boldsymbol{F}}^{(r)}$ be the r-th column of $\widehat{\boldsymbol{B}}$ and $\widehat{\boldsymbol{F}}$ such that $\widehat{\boldsymbol{B}}^{(r)} = \left(\widehat{\boldsymbol{B}}_2^{(r)}, \ldots, \widehat{\boldsymbol{B}}_T^{(r)}\right)'$ and $\widehat{\boldsymbol{F}}_{-1}^{(r)} = \left(\widehat{\boldsymbol{F}}_2^{(r)}, \ldots, \widehat{\boldsymbol{B}}_T^{(r)}\right)'$.

With estimated factors in (3.2 .12), we can use them to devise the following test statistic to detect the bubbles in common factors.

Specifically, denote $T_p = T - p$ for some $0 < p < T$, let $\widehat{\boldsymbol{B}}_t^{(r),c} = \widehat{\boldsymbol{B}}_t^{(r)} - T_2^{-1}\sum_{s=3}^{T}\widehat{\boldsymbol{B}}_{s-1}^{(r)}$ and $\widehat{\boldsymbol{F}}_{t-1}^{(r),c} = \widehat{\boldsymbol{F}}_{t-1}^{(r)} - T_2^{-1}\sum_{s=3}^{T}\widehat{\boldsymbol{F}}_{s-1}^{(r)}$, we run time-series regression based on following equation

$$\widehat{\boldsymbol{B}}_t^{(r),c} = \widehat{\nu}_r \widehat{\boldsymbol{F}}_{t-1}^{(r),c} + \phi_t \text{ for } t = 3, \ldots, T \tag{3.2 .13}$$

where ϕ_t is the regression residual, and the OLS estimator $\widehat{\nu}_r$ is obtained as follows

$$\begin{aligned} \widehat{\nu}_r \;=\; & \left[T_2 \sum_{t=3}^{T} \widehat{\boldsymbol{B}}_t^{(r)} \widehat{\boldsymbol{F}}_{t-1}^{(r)\prime} - \sum_{t=3}^{T} \widehat{\boldsymbol{B}}_t^{(r)} \sum_{t=3}^{T} \widehat{\boldsymbol{F}}_{t-1}^{(r)\prime}\right] \\ & \times \left[T_2 \sum_{t=3}^{T} \widehat{\boldsymbol{F}}_{t-1}^{(r)} \widehat{\boldsymbol{F}}_{t-1}^{(r)\prime} - \left(\sum_{t=3}^{T} \widehat{\boldsymbol{F}}_{t-1}^{(r)}\right)\left(\sum_{t=3}^{T} \widehat{\boldsymbol{F}}_{t-1}^{(r)}\right)'\right]^{-1}. \end{aligned} \tag{3.2 .14}$$

And we then define and calculate $\widehat{\varpi}_{\widehat{\nu}_r}$ as follows,

$$\widehat{\varpi}_{\widehat{\nu}_r} = \widehat{\sigma}_{(r)}^2 \left[\sum_{t=3}^{T} \widehat{\boldsymbol{F}}_{t-1}^{(r)} \widehat{\boldsymbol{F}}_{t-1}^{(r)\prime} - T_2^{-1}\left(\sum_{t=3}^{T} \widehat{\boldsymbol{F}}_{t-1}^{(r)}\right)\left(\sum_{t=3}^{T} \widehat{\boldsymbol{F}}_{t-1}^{(r)}\right)'\right]^{-1},$$

where $\widehat{\sigma}_{(r)}^2 = T_2^{-1}\sum_{t=3}^{T}\left(\widehat{\boldsymbol{B}}_t^{(r)}\right)^2$.

Then, we can implement the feasible proposed test statistic for r-th estimated factor as follows,

$$\mathcal{DF}^{(r),f} = \widehat{\varpi}_{\widehat{\nu}_r}^{-1/2}\widehat{\nu}_r. \tag{3.2 .15}$$

Under $\mathbb{H}_0$ in (3.2 .5), we will show that for $r = 1, \ldots, R_0$, the limiting distribution of $\mathcal{DF}^{(r),f}$ will involve a weighted sum of standard Brownian motion. As a special case, when $R_0 = 1$, our result is consistent with that of Chen et al. (2019) in the Single-Factor model, though their setups are slightly different. We will have more detailed discussions on limiting distributions of (3.2 .15) under the null and alternative of local-to-unity in next section.

Furthermore, we propose to test (3.2 .5) jointly as follows. First, regressing $\widehat{\boldsymbol{B}}_t^{R_0,c}$ on $\widehat{\boldsymbol{F}}_{t-1}^{R_0,c}$, where $\widehat{\boldsymbol{B}}_t^{R_0,c} = \sum_{r=1}^{R_0}\widehat{\boldsymbol{B}}_t^{(r),c}$ and $\widehat{\boldsymbol{F}}_t^{R_0,c} = \sum_{r=1}^{R_0}\widehat{\boldsymbol{F}}_t^{(r),c}$, and denote the OLS estimate of the regression coefficient by $\widehat{\nu}_{R_0}$,

$$\widehat{\nu}_{R_0} = \left[T_2\sum_{t=3}^{T}\widehat{\boldsymbol{B}}_t^{R_0}\widehat{\boldsymbol{F}}_{t-1}^{R_0\prime} - \sum_{t=3}^{T}\widehat{\boldsymbol{B}}_t^{R_0}\sum_{t=3}^{T}\widehat{\boldsymbol{F}}_{t-1}^{R_0\prime}\right]\left[T_2\sum_{t=3}^{T}\widehat{\boldsymbol{F}}_{t-1}^{R_0}\widehat{\boldsymbol{F}}_{t-1}^{R_0\prime} - \left(\sum_{t=3}^{T}\widehat{\boldsymbol{F}}_{t-1}^{R_0}\right)\left(\sum_{t=3}^{T}\widehat{\boldsymbol{F}}_{t-1}^{R_0}\right)'\right]^{-1}.$$

Similarly, we define and calculate $\widehat{\varpi}_{\widehat{\nu}_{R_0}}$ as follows,

$$\widehat{\varpi}_{\widehat{\nu}_{R_0}} = \widehat{\sigma}_{R_0}^2\left[\sum_{t=3}^{T}\widehat{\boldsymbol{F}}_{t-1}^{R_0}\widehat{\boldsymbol{F}}_{t-1}^{R_0\prime} - T_2^{-1}\left(\sum_{t=3}^{T}\widehat{\boldsymbol{F}}_{t-1}^{R_0}\right)\left(\sum_{t=3}^{T}\widehat{\boldsymbol{F}}_{t-1}^{R_0}\right)'\right]^{-1},$$

where $\widehat{\sigma}_{R_0}^2 = T_2^{-1}\sum_{t=3}^{T}\left(\widehat{\boldsymbol{B}}_t^{R_0}\right)^2$, and $\widehat{\boldsymbol{B}}_t^{R_0} = \sum_{r=1}^{R_0}\widehat{\boldsymbol{B}}_t^{(r)}$ and $\widehat{\boldsymbol{F}}_t^{R_0} = \sum_{r=1}^{R_0}\widehat{\boldsymbol{F}}_t^{(r)}$. Finally, we can test (3.2 .5) jointly by implementing

$$\mathcal{DF}^{R_0,f} = \widehat{\varpi}_{\widehat{\nu}_{R_0}}^{-1/2}\widehat{\nu}_{R_0}. \tag{3.2 .16}$$

Remark 3.4. When no serial correlation exists among $\{\boldsymbol{u}_t\}$, (3.2 .15) is rescaled by the estimates of the variance of $\widehat{\nu}_r$ up to a rotation matrix. In general case that allows for the serial correlation among

$\{\boldsymbol{u}_t\}$, (3.2 .15) is rescaled by the estimates of contemporaneous variance of $\widehat{\nu}_r$ up to the rotation matrix. Similar arguments apply to (3.2 .16). By doing so, we can alleviate the adverse impacts of the nuisance parameters due to the presence of asymptotic matrix of the rotation matrix in the limiting distributions of proposed test statistics and the serial correlations in practice, which can lead to the over-rejections of the null.

3.3 Asymptotic Properties

In this section, we first present a set of basic assumptions to derive the consistency of the PC estimation and the asymptotic properties of the statistic $\{\mathcal{DF}^{(r),f}\}_{r=1}^{R_0}$ and $\mathcal{DF}^{R_0,f}$. Then we study the asymptotic null distribution and local power properties of $\{\mathcal{DF}^{(r),f}\}_{r=1}^{R_0}$ and $\mathcal{DF}^{R_0,f}$.

3.3.1 Basic assumptions

Let $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \ldots, \epsilon_{Nt})'$, $\boldsymbol{u}_t = (u_{1t}, \ldots, u_{1R_0})'$, $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1', \ldots, \boldsymbol{\epsilon}_T')'$ and $\boldsymbol{u} = (\boldsymbol{u}_1', \ldots, \boldsymbol{u}_T')'$. Let $\gamma_N(s,t) = \sum_{i=1}^N E(\epsilon_{is}\epsilon_{it})$ and $\zeta_{s,t} = E\left|N^{-1/2}\sum_{i=1}^N [\epsilon_{is}\epsilon_{it} - E(\epsilon_{is}\epsilon_{it})]\right|^4$.

Assumption A3.1. *(a) The process $\{(\boldsymbol{\epsilon}_t, \boldsymbol{u}_t), t \geq 0\}$ is α-mixing across t with mixing coefficient $\alpha_{ij}(|t-s|)$ between $\{\epsilon_{it}\}$ and $\{\epsilon_{js}\}$ and mixing coefficient $\alpha_i(|t-s|)$ between $\{(\epsilon_{it}, \boldsymbol{u}_t)\}$ and $\{(\epsilon_{is}, \boldsymbol{u}_s)\}$. Assume that*

$$\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T(\alpha_{ij}(t))^{\delta/(4+\delta)} = O(N), \quad \max_i \sum_{j=1}^N (\alpha_{ij}(0))^{\delta/(4+\delta)} = O(1), \text{ and } \max_i \alpha_i(t) = O\left(t^{-\theta}\right),$$

where $\delta > 0$ is chosen such that $E\|\omega_{it}\|^{4+\delta} < M$ with $\omega_{it} \in \{\boldsymbol{\lambda}_i^0, \boldsymbol{u}_t, \epsilon_{it}\}$ and $\theta > (4+\delta)/\delta$.

(b) $E(\boldsymbol{u}_t) = 0$, $E(\boldsymbol{u}_t\boldsymbol{u}_t') = \Sigma_u$, and $0 < \underline{c} < \varphi_{\min}(\Sigma_u) \leq \varphi_{\max}(\Sigma_u) < \bar{c} < \infty$, where $\underline{c}$ and $\bar{c}$ are some generic finite constants, and $\Phi_u = Var\left(T^{-1/2}\sum_t \boldsymbol{u}_t\right)$ exists and is positive definite.

(c) $E(\varepsilon_{it}) = 0$, and $E(\varepsilon_{it}^2) = \sigma_i^2 \leq M$ for some finite constant M, and $\Phi_{\epsilon i} = Var\left(T^{-1/2}\sum_t \epsilon_{it}\right)$ exists and is positive definite.

(d) $E\|\boldsymbol{F}_0^0\|^2 \leq M$.

Assumption A3.2. *(a)* $\boldsymbol{\lambda}_i^0$ *is random such that* $E\left\|\boldsymbol{\lambda}_i^0\right\|^2 \leq M$*, (b)* $N^{-1}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0 \xrightarrow{p} \Sigma_\lambda$ *such that* $0 < \underline{c} < \varphi_{\min}\left(\Sigma_\lambda\right) \leq \varphi_{\max}\left(\Sigma_\lambda\right) < \bar{c} < \infty$.

Assumption A3.3. *(a)* $\max_t \sum_{s=1}^T \left|\gamma_N\left(s,t\right)\right| = O(N)$*, (b)* $\max_{s,t}\zeta_{s,t} \leq M$*, (c)* $\max_i E\left|e_{i0}\right| \leq M$*, (d)* $\left\|\boldsymbol{\epsilon}\right\|_{sp} = O_p(\sqrt{N}+\sqrt{T})$.

Assumption A3.4. $\left\{\boldsymbol{\lambda}_i^0\right\}, \left\{\boldsymbol{u}_t\right\}$*, and* $\left\{\epsilon_{js}\right\}$ *are three groups of mutually independent stochastic variables for every* (i,j,s,t).

Assumption A3.5. $\rho_i^0 = 1 + c_i/T$ *for* $i = 1, ..., N$*, where* $c_i \geq 0$ *is fixed and finite.*

Assumption A3.1, A3.2, and A3.3 impose moments conditions on the error terms, factors, factor loadings. Assumption A3.4 follows Assumption D in Bai and Ng (2004). Assumptions A3.5 accommodates unit-root and local-to-unity process in the idiosyncratic error terms specified in (3.2 .7). Assumption A3.1-A3.5 are almost the same as those assumptions in Feng and Su (2020), they are widely used in the literature; see e.g., Bai and Ng (2004) and Peng et al. (2020).

3.3.2 Asymptotic behavior of the proposed test

Let $\boldsymbol{V}_{Z,N}$ be the diagonal matrix with entries consisting of first R_0 largest eigenvalues of sample covariance matrix $N^{-1}\boldsymbol{Z}'\boldsymbol{Z}$, and the rotation matrix $\boldsymbol{H} \equiv \left(T^{-1}\boldsymbol{V}_{Z,N}\right)^{-1}\left(T^{-1}\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\right)\left(N^{-1}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\right)$, then,

Proposition 3.3.1. *Suppose that Assumptions A3.1–A3.5 hold, then,* $\boldsymbol{H}_\infty \equiv \lim_{N,T\to\infty} \boldsymbol{H}$ *exists and nonsingular.*

Proposition 3.3.1 states the result on asymptotic matrix of $\boldsymbol{H}$ as N and T pass to infinity simultaneously. This result is critical in deriving the limiting distribution of the proposed test under the null and the alternative of local-to-unity when we use the estimated factors to construct the test statistics presented in (3.2 .15) and (3.2 .16), because it is well-known that those unobserved common factors and factor loadings are identifiable only up to a rotation matrix when we apply the PC estimation.

Proposition 3.3.1 holds as the consequence of Assumption A3.2, Lemma C.1.1, and Lemma C.3.3 in the Appendix C.

Furthermore, to analyze the asymptotic local power property of $\mathcal{DF}^{(r),f}$ and $\mathcal{DF}^{R_0,f}$, we consider the following sequence of Pitman's local alternatives:

$$\mathbb{H}_{1T} : \rho_{0,r}^0 = 1 + \frac{\nu_r}{T} \text{ for } r = 1, ..., R_0.$$

Namely, we consider the alternative of local-to-unity as displayed above. Besides, we also need following technical assumption to derive the asymptotic local power of the test statistic. Denote $\boldsymbol{\nu} = diag\,(\nu_1, \ldots, \nu_{R_0})$,

Assumption A3.6. *(a) $\nu_r \geq 0$ for each r, (b) there exist nonzero R_0-dimensional vector $\boldsymbol{a}_{NT}$ and $\boldsymbol{b}_{NT}$ depending on N and T such that $\mu \equiv \lim_{N,T\to\infty} \boldsymbol{a}'_{NT}\boldsymbol{\nu}\boldsymbol{b}_{NT} > 0$.*

Assumption A3.6(a) implies that we restrict our attention to the one-sided local alternative, namely, we only consider explosive common factors under the alternative of local-to-unity. Assumption A3.6(b) implies that $\{\nu_r\}_{r=1}^{R_0}$ under the alternatives cannot be sparse such that the concentration parameter can be positive and moderately large. As documented in related studies on unit root test for time series data, e.g., empirical findings in Phillips and Yu (2011), unit root tests cannot discriminate near unity root process from unit root process very well when the concentration parameter is very small. However, in the context of high dimensional panel data we consider here, we can achieve the enhancement of the power of the proposed test (at least in the ideal case). We will discuss such kind of possible enhancement of power of the proposed test later.

The following theorem states the null distribution and local power properties of $\mathcal{DF}^{(r),f}$ for $r = 1, \ldots, R_0$.

Theorem 3.3.2. *Suppose that Assumptions A3.1–A3.5 hold. Then as $(N, T) \to \infty$, as long as $T/N^2 \to 0$, for $r = 1, \ldots, R_0$*

(a) *Under* $\mathbb{H}_0$,

$$\mathcal{DF}^{(r),f} \quad \Rightarrow \quad \psi_0^{(r),f}, \tag{3.3 .1}$$

where

$$\psi_0^{(r),f} \equiv \left(\boldsymbol{H}'_{(r),\infty}\Sigma_u\boldsymbol{H}_{(r),\infty}\right)^{-1/2}$$
$$\left\{\boldsymbol{H}'_{(r),\infty}\Sigma_u^{1/2}\left[\int_0^1 \mathrm{d}\mathbf{W}(r)'\mathbf{W}(r) - \boldsymbol{W}(1)\int_0^1 \mathbf{W}(r)'\mathrm{dr}\right]\Sigma_u^{1/2}\boldsymbol{H}_{(r),\infty} + \boldsymbol{H}'_{(r),\infty}\Omega_u\boldsymbol{H}_{(r),\infty}\right\}$$
$$\left\{\boldsymbol{H}'_{(r),\infty}\Sigma_u^{1/2}\left[\int_0^1 \mathbf{W}(r)\mathbf{W}(r)'\mathrm{dr} - \left(\int_0^1 \mathbf{W}(r)'\mathrm{dr}\right)\left(\int_0^1 \mathbf{W}(r)\mathrm{dr}\right)'\right]\Sigma_u^{1/2}\boldsymbol{H}_{(r),\infty}\right\}^{-1/2},$$

where $\boldsymbol{H}_{(r),\infty}$ *is the* r*-th column of* $\boldsymbol{H}_\infty$ *defined in Proposition 3.3.1;* $\boldsymbol{W}(r)$ *is the* R*-vector standard Brownian motion on* $\mathcal{C}[0,1]$ *given by the weak limit of the partial sum* $\Sigma_u^{-1/2}T^{-1/2}\sum_{t=1}^{[Tr]}\boldsymbol{u}_t$*, and* $\Omega_u \equiv \sum_{k=1}^{\infty} E\left(\boldsymbol{u}_t\boldsymbol{u}'_{t+k}\right)$.

(b) *Furthermore, if Assumption A3.6 also holds, then under* $\mathbb{H}_{1T}$*, for* $r = 1, \ldots, R_0$,

$$\mathcal{DF}^{(r),f} \quad \Rightarrow \quad \mathcal{DF}_{\nu,\infty}^{(r),f} \equiv \overline{\chi}_\nu^{(r),f} + \psi_\nu^{(r),f}, \tag{3.3 .2}$$

where

$$\overline{\chi}_\nu^{(r),f} \equiv \left(\boldsymbol{H}'_{(r),\infty}\Sigma_u\boldsymbol{H}_{(r),\infty}\right)^{-1/2}$$
$$\left\{\boldsymbol{H}'_{(r),\infty}\boldsymbol{\nu}\Sigma_u^{1/2}\left[\int_0^1 \mathbf{J}_\nu(r)\mathbf{J}_\nu(r)'\mathrm{dr} - \left(\int_0^1 \mathbf{J}_\nu(r)'\mathrm{dr}\right)\left(\int_0^1 \mathbf{J}_\nu(r)\mathrm{dr}\right)'\right]\Sigma_u^{1/2}\boldsymbol{H}_{(r),\infty}\right\}$$
$$\left\{\boldsymbol{H}'_{(r),\infty}\Sigma_u^{1/2}\left[\int_0^1 \mathbf{J}_\nu(r)\mathbf{J}_\nu(r)'\mathrm{dr} - \left(\int_0^1 \mathbf{J}_\nu(r)'\mathrm{dr}\right)\left(\int_0^1 \mathbf{J}_\nu(r)\mathrm{dr}\right)'\right]\Sigma_u^{1/2}\boldsymbol{H}_{(r),\infty}\right\}^{-1/2},$$

and

$$\psi_{\nu}^{(r),f} \equiv \left(\boldsymbol{H}'_{(r),\infty}\Sigma_u \boldsymbol{H}_{(r),\infty}\right)^{-1/2}$$
$$\left\{\boldsymbol{H}'_{(r),\infty}\Sigma_u^{1/2}\left[\int_0^1 \mathrm{d}\mathbf{W}(r)'\mathbf{J}_{\nu}(r) - \boldsymbol{W}(1)\int_0^1 \mathbf{J}_{\nu}(r)'\mathrm{dr}\right]\Sigma_u^{1/2}\boldsymbol{H}_{(r),\infty} + \boldsymbol{H}'_{(r),\infty}\Omega_u \boldsymbol{H}_{(r),\infty}\right\}$$
$$\left\{\boldsymbol{H}'_{(r),\infty}\Sigma_u^{1/2}\left[\int_0^1 \mathbf{J}_{\nu}(r)\mathbf{J}_{\nu}(r)'\mathrm{dr} - \left(\int_0^1 \mathbf{J}_{\nu}(r)'\mathrm{dr}\right)\left(\int_0^1 \mathbf{J}_{\nu}(r)\mathrm{dr}\right)'\right]\Sigma_u^{1/2}\boldsymbol{H}_{(r),\infty}\right\}^{-1/2}.$$

where $\boldsymbol{H}_{(r),\infty}$ *stands for the* r*-th column of* $\boldsymbol{H}_{\infty}$ *defined in Proposition 3.3.1, and* $\mathbf{J}_{\nu}(r) \equiv \left(\mathrm{J}_{\nu_1}(r), \ldots, \mathrm{J}_{\nu_{R_0}}(r)\right)$ *is a Ornstein-Uhlenbeck process that satisfies* $\mathbf{J}_{\nu}(r) = \boldsymbol{W}(r) + \boldsymbol{\nu}\int_0^r e^{(r-s)c}\boldsymbol{W}(s)\mathrm{ds}$*,* $\mathbf{J}_{\nu}(0) = \mathbf{0}$*, and* $\Omega_u \equiv \sum_{k=1}^{\infty} E\left(\boldsymbol{u}_t\boldsymbol{u}'_{t+k}\right)$*.*

Given the result in Theorem 3.3.2(a) and noticing that our test is a one-sided test, we can reject the null when $\mathcal{DF}^{(r),f}$ is sufficiently large, say larger than the associated asymptotic critical value at a given significance level. However, the limiting null distribution of $\mathcal{DF}^{(r),f}$ is generally not pivotal due to the presence of $\boldsymbol{H}_{(r),\infty}$, the element of $\boldsymbol{H}_{\infty}$. In particular, it is worthwhile to point out that the limiting null distribution in Theorem 3.3.2(a) turns out to be pivotal in the Single-Factor model when there is no serial correlations among $\{\boldsymbol{u}_t\}$, which is consistent with the result in Chen et al. (2019).

Notably, let $\boldsymbol{\nu}$ in Assumption A3.6 be the zero matrix in Theorem 3.3.2(b), the result degenerates to the limiting null distribution of $\mathcal{DF}^{(r),f}$ with the replacement of the Ornstein-Uhlenbeck process by the standard Brownian motion in the formula. Besides, Theorem 3.3.2(b) provides two important implications. First, under Assumption A3.6(b), Theorem 3.3.2(b) indicates that the proposed test has nontrivial power to detect the local alternative converging to the null at rate T^{-1}, this is because the concentration parameter $\overline{\chi}_{\nu}^{(r),f}$ involves the rotation matrix depending on both cross-section and time dimensions. In this sense, cross-section information can enhance the power of the proposed test by scaling up the concentration parameter implicitly at least in the ideal case. Second, it is worth mentioning the concentration parameter $\overline{\chi}_{\nu}^{(r),f}$ in Theorem 3.3.2(b) is different from the result of Theorem 2 in Horie and Yamamoto (2016), this is because we allow for different values among $\{\nu_r\}$

while they exclude the heterogeneous effects of the local-to-unity explosiveness from unobserved common factors such that $\nu_1 = \nu_2 \ldots = \nu_{R_0}$.

In the view of above discussions, there is the theoretical possibility that the heterogeneity of the local-to-unity explosive factors and the cross-section information contained in the rotation matrix can contribute to the discriminatory power of the proposed test between the unit root and local-to-unity explosive process in common factors. Thus, we can benefit from employing panel factor models and the PC estimation to detect bubbles in unobserved common factors.

Remark 3.5. It is interesting to point out that, based on our proofs for Theorem 3.3.2 and the related analysis in Feng and Su (2020), when the unobserved common factors are mild explosive with local-to-unity explosive idiosyncratic error terms in panel factor models, our proposed test will diverge at a exponential rate, which is much faster than that in (3.3 .2). So, theoretically, our proposed test also will work with much higher power in this case.

Similar to above results, we now present the asymptotic null and local power properties of the proposed joint tests. The following theorem states the null distribution and local power properties of $\mathcal{DF}^{R_0,f}$.

Theorem 3.3.3. *Suppose that Assumptions A3.1–A3.5 hold. Then as* $(N,T) \to \infty$*, as long as* $T/N^2 \to 0$*, for* $r = 1, \ldots, R_0$

(a) Under $\mathbb{H}_0$*,*

$$\mathcal{DF}^{R_0,f} \quad \Rightarrow \quad \psi_0^{R_0,f}, \tag{3.3 .3}$$

where

$$\psi_0^{R_0,f} \equiv \left(\boldsymbol{H}'_{R_0,\infty}\Sigma_u\boldsymbol{H}_{R_0,\infty}\right)^{-1/2}$$
$$\left\{\boldsymbol{H}'_{R_0,\infty}\Sigma_u^{1/2}\left[\int_0^1 \mathrm{d}\mathbf{W}(r)'\mathbf{W}(r) - \boldsymbol{W}(1)\int_0^1 \mathbf{W}(r)'\mathrm{dr}\right]\Sigma_u^{1/2}\boldsymbol{H}_{R_0,\infty} + \boldsymbol{H}'_{R_0,\infty}\Omega_u\boldsymbol{H}_{R_0,\infty}\right\}$$
$$\left\{\boldsymbol{H}'_{R_0,\infty}\Sigma_u^{1/2}\left[\int_0^1 \mathbf{W}(r)\mathbf{W}(r)'\mathrm{dr} - \left(\int_0^1 \mathbf{W}(r)'\mathrm{dr}\right)\left(\int_0^1 \mathbf{W}(r)\mathrm{dr}\right)'\right]\Sigma_u^{1/2}\boldsymbol{H}_{R_0,\infty}\right\}^{-1/2},$$

where $\boldsymbol{H}_{R_0,\infty} \equiv \sum_{r=1}^{R_0} \boldsymbol{H}_{(r),\infty}$, $\boldsymbol{W}(r)$ *is the* R*-vector standard Brownian motion on* $\mathcal{C}[0,1]$ *that is given by the weak limit of the partial sum* $\Sigma_u^{-1/2} T^{-1/2} \sum_{t=1}^{[Tr]} \boldsymbol{u}_t$*, and* $\Omega_u \equiv \sum_{k=1}^{\infty} E\left(\boldsymbol{u}_t \boldsymbol{u}_{t+k}'\right)$.

(b) *Furthermore, if Assumption A3.6 also holds, then under* $\mathbb{H}_{1T}$*, for* $r = 1, \ldots, R_0$,

$$\mathcal{DF}^{R_0,f} \quad \Rightarrow \quad \mathcal{DF}_{\nu,\infty}^{R_0,f} \equiv \overline{\chi}_{\nu}^{R_0,f} + \psi_{\nu}^{R_0,f}, \tag{3.3.4}$$

where

$$\overline{\chi}_{\nu}^{R_0,f} \equiv \left\{ \boldsymbol{H}_{R_0,\infty}' \boldsymbol{\nu} \Sigma_u^{1/2} \left[\int_0^1 \mathbf{J}_\nu(r)\mathbf{J}_\nu(r)'\mathrm{dr} - \left(\int_0^1 \mathbf{J}_\nu(r)'\mathrm{dr}\right)\left(\int_0^1 \mathbf{J}_\nu(r)\mathrm{dr}\right)' \right] \Sigma_u^{1/2} \boldsymbol{H}_{R_0,\infty} \right\}$$
$$\left\{ \boldsymbol{H}_{R_0,\infty}' \Sigma_u^{1/2} \left[\int_0^1 \mathbf{J}_\nu(r)\mathbf{J}_\nu(r)'\mathrm{dr} - \left(\int_0^1 \mathbf{J}_\nu(r)'\mathrm{dr}\right)\left(\int_0^1 \mathbf{J}_\nu(r)\mathrm{dr}\right)' \right] \Sigma_u^{1/2} \boldsymbol{H}_{R_0,\infty} \right\}^{-1/2}$$
$$\left(\boldsymbol{H}_{R_0,\infty}' \Sigma_u \boldsymbol{H}_{R_0,\infty}\right)^{-1/2},$$

and

$$\psi_{\nu}^{R_0,f} \equiv \left(\boldsymbol{H}_{R_0,\infty}' \Sigma_u \boldsymbol{H}_{R_0,\infty}\right)^{-1/2}$$
$$\left\{ \boldsymbol{H}_{R_0,\infty}' \Sigma_u^{1/2} \left[\int_0^1 \mathrm{d}\mathbf{W}(r)'\mathbf{J}_\nu(r) - \boldsymbol{W}(1)\int_0^1 \mathbf{J}_\nu(r)'\mathrm{dr} \right] \Sigma_u^{1/2} \boldsymbol{H}_{R_0,\infty} + \boldsymbol{H}_{R_0,\infty}' \Omega_u \boldsymbol{H}_{R_0,\infty} \right\}$$
$$\left\{ \boldsymbol{H}_{R_0,\infty}' \Sigma_u^{1/2} \left[\int_0^1 \mathbf{J}_\nu(r)\mathbf{J}_\nu(r)'\mathrm{dr} - \left(\int_0^1 \mathbf{J}_\nu(r)'\mathrm{dr}\right)\left(\int_0^1 \mathbf{J}_\nu(r)\mathrm{dr}\right)' \right] \Sigma_u^{1/2} \boldsymbol{H}_{R_0,\infty} \right\}^{-1/2}.$$

where $\boldsymbol{H}_{R_0,\infty} \equiv \sum_{r=1}^{R_0} \boldsymbol{H}_{(r),\infty}$*, and* $\mathbf{J}_\nu(r) \equiv \left(\mathrm{J}_{\nu_1}(r), \ldots, \mathrm{J}_{\nu_{R_0}}(r)\right)$ *is a Ornstein-Uhlenbeck process such that* $\mathbf{J}_\nu(r) = \boldsymbol{W}(r) + \boldsymbol{\nu} \int_0^r e^{(r-s)c} \boldsymbol{W}(s)\mathrm{ds}$*,* $\mathbf{J}_\nu(0) = \mathbf{0}$*, and* $\Omega_u \equiv \sum_{k=1}^{\infty} E\left(\boldsymbol{u}_t \boldsymbol{u}_{t+k}'\right)$.

3.3.3 A bootstrap version of the test

Although Theorem 3.3.2 shows that the proposed test has non-trivial power against the alternative of local-to-unity, the limiting null distribution of the proposed test is generally not pivotal due to the asymptotic matrix of rotation matrix in the limiting distribution. Therefore, we cannot tabulate the asymptotic critical values as done in Horie and Yamamoto (2016) or Chen et al. (2019). Instead, we employ the dependent wild bootstrap (DWB) method to simulate the critical values, which is shown to be valid in a nonstatioanry setting in Rho and Shao (2019) recently.

Similar to Algorithm 3.1 in Rho and Shao (2019), we apply the DWB method to our proposed test $\mathcal{DF}^{(r),f}$ for $r = 1, \ldots, R_0$ below,

Algorithm 1: The Dependent Wild Bootstrap

1. Calculate the OLS estimate $\widehat{\nu}_r$ by regressing $\widehat{\boldsymbol{B}}_t^{(r),c}$ on $\widehat{\boldsymbol{F}}_{t-1}^{(r),c}$, and ϕ_t is the corresponding regression residual as shown in (3.2 .13) for $t = 3, \ldots, T$. Calculate the proposed test statistic $\mathcal{DF}^{(r),f}$ as defined in (3.2 .15).

2. Randomly generate the l-dependent mean-zero stationary series $\{W_{t,T}\}_{t=3}^{T}$ satisfying following conditions:

   (a) $\{W_{t,T}\}_{t=1}^{n}$ is a realization from a stationary time series with $E\left(W_{t,T}\right) = 0$ and $\operatorname{var}\left(W_{t,T}\right) = 1$. $\{W_{t,T}\}_{t=1}^{n}$ are independent of the data, $\operatorname{cov}\left(W_{t,T}, W_{t',T}\right) = a\left\{\left(t - t'\right)/l\right\}$ where $a(\cdot)$ is a kernel function and $l = l_T$ is a bandwidth parameter that satisfies $l \asymp CT^{\kappa}$ for some $0 < \kappa < 1/3$. Assume that $W_{t,T}$ is l -dependent and $E\left(W_{1,T}^4\right) < \infty$.

   (b) $a : \mathbb{R} \to [0,1]$ is symmetric and has compact support on $[-1,1]$, $a(0) = 1$, $\lim_{x\to 0}\{1 - a(x)\}/|x|^q = k_q \neq 0$ for some $q \in (0,2]$, and $\int_{-\infty}^{\infty} a(u)e^{-iux}du \geq 0$ for $x \in \mathbb{R}$.

   And then generate the residuals $\phi_t^* = \phi_t W_{t,T}$ in the bootstrap world.

3. Generate the bootstrapped sample $\{\widehat{\boldsymbol{B}}_t^{(r)*}\}$ by $\widehat{\boldsymbol{B}}_t^{(r)*} = \widehat{\nu}_r \cdot \widehat{\boldsymbol{F}}_{t-1}^{(r),c} + \phi_t^*$ using ϕ_t^* as if $\widehat{\nu}_r = 0$ is true.

4. Calculate the bootstrapped OLS estimate $\widehat{\nu}_r^*$ by regressing $\boldsymbol{B}_t^{(r)*}$ on $\widehat{\boldsymbol{F}}_{t-1}^{(r),c}$, and calculate the proposed test statistic $\widetilde{\mathcal{DF}}^*_{(r),f}$ by (3.2 .15).

5. Repeat Step 2 to 4 B times and denote the resulting bootstrap test statistics as $\{\widetilde{\mathcal{DF}}^*_{(r),f,b}\}_{b=1}^{B}$. Obtain the bootstrap $p-$value as $p^* = B^{-1}\sum_{b=1}^{B}\mathbb{I}\left\{\widetilde{\mathcal{DF}}^*_{(r),f,b} > \widetilde{\mathcal{DF}}\right\}$, where $\mathbb{I}\{\cdot\}$ is the usual indicator function.

We present the DWB scheme for the joint test as follows.

Algorithm 2: The Dependent Wild Bootstrap for The Joint Test

1. Calculate the proposed joint test statistic $\mathcal{DF}^{R_0,f}$ as defined in (3.2 .16).

2. Execute Step 1 to 3 in Algorithm 1 for each estimated factor separately.

3. Calculate $\widehat{\boldsymbol{B}}_t^{R,*} = \sum_{r=1}^{R_0} \widehat{\boldsymbol{B}}_t^{(r),*}$, where $\widehat{\boldsymbol{B}}_t^{(r),*}$ is obtained by Step 2 just above for each r. Based on $\{\widehat{\boldsymbol{B}}_t^{R,*}, \widehat{\boldsymbol{F}}_t^{R}\}$, calculate the proposed joint test statistic $\mathcal{DF}^{R_0,f}$ as defined in (3.2 .16).

4. Repeat Step 2 to 3 B times and denote the resulting bootstrap test statistics as $\{\widetilde{\mathcal{DF}}^*_{R_0,f,b}\}_{b=1}^{B}$. Obtain the bootstrap $p-$value as $p^* = B^{-1}\sum_{b=1}^{B}\mathbb{I}\left\{\widetilde{\mathcal{DF}}^*_{R_0,f,b} > \widetilde{\mathcal{DF}}\right\}$, where $\mathbb{I}\{\cdot\}$ is the usual indicator function.

The following theorem establishes the asymptotic validity of the above bootstrap test.

Theorem 3.3.4. *(Bootstrap Consistency and Power). Suppose that Assumptions A3.1–A3.6 hold. Then, for any nonzero diagonal matrix $\boldsymbol{\nu} = diag(\nu_1, \ldots, \nu_{R_0})$ with $\nu_r \geq 0$ for all r, as $(N,T) \to \infty$, as long as $T^{1+\eta}/N^2 \to 0$ for some $\eta \in (0, 1/3)$,*

(a)

$$P\left(\mathcal{DF}^{(r),f} \leq \widetilde{\mathcal{DF}}^*_{(r),f}(\alpha) \mid \boldsymbol{\rho}_0^0 = \boldsymbol{I}_R + \boldsymbol{\nu}/T\right) \xrightarrow{D*} P\left(\mathcal{DF}_{\nu}^{(r),f} \leq \psi_0^{(r),f}(\alpha)\right) \quad (3.3\ .5)$$

where $\xrightarrow{D*}$ denotes weak convergence under the bootstrap probability measure conditional on the observed sample $\mathcal{X} = \{Z_{it}\}$, and $\mathcal{DF}_{\nu}^{(r),f}$ is the random variable with the distribution $\mathcal{DF}_{\nu,\infty}^{(r),f}$ shown in (3.3 .2), $\psi_0^{(r),f}(\alpha)$ is the α-quantiles of the limiting null distribution $\psi_0^{(r),f}$ shown in (3.3 .1). $\widetilde{\mathcal{DF}}^*_{(r),f}(\alpha)$ is the α-quantiles of $\{\widetilde{\mathcal{DF}}^*_{(r),f,b}\}_{b=1}^{B}$ conditional on $\mathcal{X}$.

(b) Moreover, for the joint test,

$$P\left(\mathcal{DF}^{R_0,f} \leq \widetilde{\mathcal{DF}}^*_{R_0,f}(\alpha) \mid \boldsymbol{\rho}_0^0 = \boldsymbol{I}_R + \boldsymbol{\nu}/T\right) \xrightarrow{D*} P\left(\mathcal{DF}_{\nu}^{R_0,f} \leq \psi_0^{R_0,f}(\alpha)\right) \quad (3.3\ .6)$$

where $\mathcal{DF}_{\nu}^{(r),f}$ is the random variable with the distribution $\mathcal{DF}_{\nu,\infty}^{R_0,f}$ shown in (3.3 .2), $\psi_0^{R_0,f}(\alpha)$ is the α-quantiles of the limiting null distribution $\psi_0^{R_0,f}$ shown in (3.3 .3). $\widetilde{\mathcal{DF}}^*_{(r),f}(\alpha)$ is the α-quantiles of $\{\widetilde{\mathcal{DF}}^*_{R_0,f,b}\}_{b=1}^{B}$ conditional on $\mathcal{X}$.

We demonstrate that the DWB method is applicable to the alternative of local-to-unity in Theorem 3.3.4 as a direct extension of findings on the alternative of near integration in Rho and Shao (2019). When $\nu_1 = \ldots \nu_{R_0} = 0$ holds under the null, Theorem 3.3.4 shows that the bootstrap provides an asymptotic valid approximation to the limiting null distribution of $\mathcal{DF}^{(r),f}$. This holds because we generate the bootstrap data by imposing the null hypothesis in Algorithms 1 and 2. Since the original test statistic $\mathcal{DF}^{(r),f}$ is generally not pivotal as discussed earlier, the application of the DWB method here may fail to obtain the second order asymptotic refinement from the theoretical perspective. However, the DWB method still can be useful for better finite-sample performances of the proposed test in practice to circumvent finite-sample problems. Similar comments also apply to the asymptotic properties of $\mathcal{DF}^{R_0,f}$.

Remark 3.6. Note that we require $T^{1+\eta}/N^2 \to 0$ for some $\eta \in (0, 1/3)$ as $(N, T) \to \infty$ in Theorem 3.3.4, which is slightly stronger than $T/N^2 \to 0$ in Theorem 3.3.2 and Theorem 3.3.3. The intuition behind such minor difference is simple: when we want to apply the DWB scheme to capture the unknown form of serial dependence among $\{\boldsymbol{u}_t\}$, the additional term T^{η} is the price we pay for this purpose. However, we address that such technical condition is imposed for the purpose of theoretical derivations, and not a stringent restriction in practice for empirical studies.

3.4 Discussions on model selection

In this section, we discuss two issues on the model selection such that practitioners can implement the proposed tests in practice to analyze real data. The first one is to choose modeling data between local-to-unity and mild explosiveness. The second one is to determine the number of common factors used for the PC estimation.

3.4.1 The effect of mild explosiveness in idiosyncratic components

Notice that on the one hand, as pointed out by Onatski and Wang (2020) and Feng and Su (2020), the presence of the mild explosiveness in the idiosyncratic components can lead to the invalidity of the PC estimation because of adverse impacts of spurious factors. Therefore, practitioners cannot test bubbles in common factors for the mildly explosive data by applying the PC estimation because the mild explosiveness in idiosyncratic components can possibly be misidentified as the mild explosiveness in common factors as argued in Feng and Su (2020). On the other hand, in the general setups, to our best knowledge, we have no well-developed method that is robust to the presence of the mild explosiveness in the idiosyncratic components to obtain the consistent estimations of unobserved common factors and idiosyncratic components in panel factor models. We leave this challenging issue for our future research.

Our theoretical arguments for the proposed test indeed can accommodate the case that no mild explosiveness exists in the idiosyncratic components. Therefore, if we want to test the bubbles in common factors, we have two routines to apply the proposed tests in this paper reasonably: first, we have to choose modeling data as local-to-unity explosive data when extracting unobserved common factors by the PC estimation; second, we can make full use the proposed test for idiosyncratic bubbles in Feng and Su (2020) to exclude the existence of explosive bubbles in the idiosyncratic components in panel factor models.

3.4.2 Testing against the mild explosiveness in data based on the largest eigenvalue of sample variance matrix

For the first routine mentioned in the previous subsection, the exclusion of mild explosiveness in data is a preliminary step to implement our proposed test for detecting common bubbles in data.

In our current setups, only data matrix $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N)'$ with $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{iT})'$ is observed. To distinguish between the mild and local-to-unity explosiveness in data, we have to rely upon following quasi-differencing transformation as discussed in Section 3.2,

$$\begin{aligned} X_{it} &= \rho_i^0 X_{it-1} + \boldsymbol{\lambda}_i^{0\prime} \boldsymbol{F}_t^0 - \rho_i^0 \boldsymbol{\lambda}_i^{0\prime} \boldsymbol{F}_{t-1}^0 + \epsilon_{it} \\ &= \rho_i^0 X_{it-1} + \boldsymbol{\lambda}_i^{0\prime} \left(\boldsymbol{\rho}_0^0 - \rho_i^0 \boldsymbol{I}_R \right) \boldsymbol{F}_{t-1}^0 + \boldsymbol{\lambda}_i^{0\prime} \boldsymbol{u}_t + \epsilon_{it}, \end{aligned} \tag{3.4 .1}$$

According to (3.4 .1), one may follow Moon and Weidner (2015) to use the iteration scheme to obtain the consistent estimates of ρ_i^0 and $\boldsymbol{F}_{t-1}^0$. However, the explosiveness in $\boldsymbol{F}_{t-1}^0$ or/and the endogeneity caused by the relation between X_{it-1} and $\boldsymbol{F}_{t-1}^0$ can yield the in consistent estimations. For example, when both $\{F_t\}$ and $\{e_{it}\}$ in (3.2 .7) are mildly explosive, by regressing X_{it} on X_{it-1} series by series, the impact of explosiveness in data on the usual estimation strategies is tricky and unclear. Therefore, the model selection strategy such as information criteria or test statistics based on the estimates of $\{\rho_i^0\}_{i=1}^N$ and $\{\rho_{0,r}^0\}_{r=1}^R$ is not reliable to distinguish between mild explosiveness and local-to-unity explosiveness in data.

However, the largest eigenvalue of sample covariance matrix $(N^{-1}\boldsymbol{X}'\boldsymbol{X})$ is significantly different between mild and local-to-unity explosiveness. To see this point, with out loss of generality, consider a Single-Factor model such that $R_0 = 1$, then, by direct calculations based on the relation in (3.2 .7), there exist $T \times T$ matrix $\boldsymbol{Q}_i$, $\boldsymbol{S}_i$, and the T-dimensional vector $\boldsymbol{W}_i$ $\boldsymbol{V}_i$ such that

$$\boldsymbol{X}_i = \boldsymbol{W}_i F_0^0 + \boldsymbol{V}_i \epsilon_{i0} + \boldsymbol{Q}_i \boldsymbol{U} + \boldsymbol{S}_i \boldsymbol{\epsilon}_i$$

Then, it follows that

$$\Xi = N^{-1} \sum_{i=1}^N \boldsymbol{X}_i \boldsymbol{X}_i' = N^{-1} \sum_{i=1}^N \left(\boldsymbol{W}_i F_0^0 + \boldsymbol{V}_i \epsilon_{i0} + \boldsymbol{Q} \boldsymbol{U} + \boldsymbol{S} \boldsymbol{\epsilon}_i \right) \left(\boldsymbol{W}_i F_0^0 + \boldsymbol{V}_i \epsilon_{i0} + \boldsymbol{Q} \boldsymbol{U} + \boldsymbol{S} \boldsymbol{\epsilon}_i \right)'.$$

By simple and direct calculations, it is readily to obtain that

$$\begin{aligned} \|\Xi\|_{sp} \leq\ & \left\| N^{-1} \sum_{i=1}^N \boldsymbol{W}_i \boldsymbol{W}_i' \right\|_{sp} + \left\| N^{-1} \sum_{i=1}^N \boldsymbol{V}_i \boldsymbol{V}_i' \right\|_{sp} + \left\| N^{-1} \sum_{i=1}^N \boldsymbol{Q}_i \boldsymbol{Q}_i' \right\|_{sp} + \left\| N^{-1} \sum_{i=1}^N \boldsymbol{S}_i \boldsymbol{S}_i' \right\|_{sp} \\ & + \text{dominated terms}, \end{aligned}$$

provided that $E\left|F_0^0\right| < \infty$ and $E|\epsilon_{i0}| < \infty$ for all i.

According to (3.4 .2), if both common factors and idiosyncratic components in (3.2 .7) follow the unit-root process, the right-hand side of (3.4 .2) will be of order $O_p\left(T^2\right)$, so the largest eigenvalue of sample variance matrix $\|\Xi\|_{sp}$ will diverge at the rate T^2 at most; otherwise, if common factors or idiosyncratic components in (3.2 .7) are mildly explosive, $\|\Xi\|_{sp}$ will diverge at a rate much faster than T^2.

To make full use of different divergent rates of the largest eigenvalues under the mild and local-to-unity explosiveness in data, it is possible to distinguish different types of the explosiveness in data. Built upon recent advances in high-dimensional time series, our feasible suggestion is to construct a test statistic based on the largest eigenvalue of sample variance. In particular, Zhang et al. (2018) and Zhang et al. (2020) show that, under fairly mild conditions, their test statistic based on the largest eigenvalue of sample variance will follow the standard normal distribution after being rescaled appropriately for unit root and near-to-unit-root cases in the context of high-dimensional time series. We conjecture that their arguments and conclusions still hold with slight modifications for the local-to-unity explosive case shown in (3.4 .1) under our assumptions. We then reject the null of local-to-unity explosiveness in data when the test statistic is sufficiently large than the right-tailed critical value of the standard normal distribution at a given significance level. For implementation procedures of such feasible test statistic, we refer readers to Zhang et al. (2020). We leave the formal justifications as future research.

3.4.3 Determinant of the numbers of common factors

The determinant of the numbers of common factors is another essential issue for model selections when practitioners applies the PC estimation to data matrix.

When the data is mildly explosive in both common factors and idiosyncratic components, it is very challenging to choose the true number of factors because of the adverse effects of overwhelming dominant factors and spurious factors as remarked by Feng and Su (2020).

Fortunately, when the data is local-to-unity explosive, we can follow Bai (2004) to use the firstdiffereced form of data to estimate the number of common factors by employing PC criterion in Bai and Ng (2002) or other improved criteria proposed in Zeng et al. (2019) and Peng et al. (2020) recently.

3.5 Monte Carlo Simulations

In this section we evaluate the finite sample performance of the proposed test by means of Monte Carlo simulation based on the data generating processes (DGPs) given below, which differ in how we generate the common factors and idiosyncratic error terms.

3.5.1 Data generating processes

In this section, we consider the following data generating process (DGP):

$$\begin{aligned} X_{it} &= \boldsymbol{\lambda}_i^{0\prime}\boldsymbol{F}_t^0 + e_{it}, \\ \boldsymbol{F}_t^0 &= \boldsymbol{\rho}_0^0\boldsymbol{F}_{t-1}^0 + \boldsymbol{u}_t, \\ e_{it} &= \rho_i^0 e_{it-1} + \epsilon_{it} \end{aligned} \tag{3.5 .1}$$

where $\boldsymbol{\rho}_0^0 \equiv diag\left(\rho_{0,1}^0, \ldots, \rho_{0,R_0}^0\right)$ is an $R_0 \times R_0$ diagonal matrix. All data are generated by (3.5 .1) with different settings of parameters.

DGP 1 $\lambda_i \overset{i.i.d}{\sim} \mathcal{N}(0,1)$, $\epsilon_{it} \overset{i.i.d}{\sim} \mathcal{N}(0,1)$, $\boldsymbol{u}_t \overset{i.i.d}{\sim} \mathcal{N}(0,\boldsymbol{I}_{R_0})$, $\rho_{0,r}^0 = 1$ for all r, and $\{\rho_i^0\}_{i=1}^N$ follow the uniform distribution $U[a,b]$ with the parameters a and b to be specified.

DGP 2 $\lambda_i \overset{i.i.d}{\sim} \mathcal{N}(0,1)$, ϵ_{it} and $\boldsymbol{u}_t$ follow the m-dependent process, $\rho_{0,r}^0 = 1$ for all r, and $\{\rho_i^0\}_{i=1}^N$ follow the uniform distribution $U[a,b]$ with the parameters a and b to be specified.

DGP 3 $\lambda_i \overset{i.i.d}{\sim} \mathcal{N}(0,1)$, $\epsilon_{it} \overset{i.i.d}{\sim} \mathcal{N}(0,1)$, $\boldsymbol{u}_t \overset{i.i.d}{\sim} \mathcal{N}(0,\boldsymbol{I}_{R_0})$, $\rho_{0,r}^0 > 1$ for all r, and $\{\rho_i^0\}_{i=1}^N$ follow the uniform distribution $U[a,b]$ with the parameters a and b to be specified.

DGP 4 $\lambda_i \overset{i.i.d}{\sim} \mathcal{N}(0,1)$, , ϵ_{it} and $\boldsymbol{u}_t$ follow the m-dependent process, $\rho_{0,r}^0 > 1$ for all r, and $\{\rho_i^0\}_{i=1}^N$ follow the uniform distribution $U[a,b]$ with the parameters a and b to be specified.

Apparently, DGPs 1 and 2 are used to evaluate the size of the proposed tests, and DGPs 3 and 4 are used to evaluate the power performances of the proposed tests.

To study the empirical size of the proposed test, we set other parameters in DGPs 1 and 2 as follows. First, we consider three types of local-to-unity explosiveness among $\{\rho_i^0\}_{i=1}^N$, namely, (1) all autoregressive coefficients are in the near vicinity of unity such that $\{\rho_i\}_{i=1}^N \overset{i.i.d}{\sim} U(1.001, 1.010)$; (2) autoregressive coefficients lie in a wide spread of the vicinity of unity such that $\{\rho_i\}_{i=1}^N \overset{i.i.d}{\sim} U(1.006, 1.018)$; (3) autoregressive coefficients concentrate on a far vicinity of unity such that $\{\rho_i\}_{i=1}^N \overset{i.i.d}{\sim} U(1.014, 1.016)$. Note that for all cases, we set $\{\rho_i^0\}_{i=1}^N$ to range between 1 and 1.018 because the local-to-unity explosiveness in idiosyncratic components are considered in the current paper. When these autoregressive coefficients are moderately larger than 1 such as 1.02 or 1.04, it can be more appropriate to model the data as mildly explosive model.

We also consider two typical cases of true factors, namely, (1) the Single-Factor Model and (2) the Three-Factors Model. We consider the Single-Factor Model because $\mathcal{DF}^{(r),f}$ is theoretically pivotal udder DGP 1 as pointed out after Theorem 3.3.2, we then can compare the performance of the proposed test based on simulated asymptotic and bootstrapped critical values. At the same time, we also assess the empirical size of the DWB implementation of the proposed test under the Three-Factors Model when the asymptotic critical values cannot be tabulated by simulations.

To study the power of the proposed test under the alternative of local-to-unity, we consider that $\{\rho_{0,r}^0\}_{r=1}^R$ are moderately large with $R_0 = 3$ in DGPs 3 and 4 as follows:

(1) $\boldsymbol{\rho}_0^0 = diag(0.0000, 1.0160, 1.0180)$ (2) $\boldsymbol{\rho}_0^0 = diag(1.0160, 1.0170, 1.0180)$; (3) $\boldsymbol{\rho}_0^0 = diag(1.0176, 1, 0178, 1.0180)$; (4) $\boldsymbol{\rho}_0^0 = diag(1.0175, 1.0180, 1.0185)$.

For the autoregressive coefficients $\{\rho_i^0\}_{i=1}^N$, we focus on the case with a wide spread of the vicinity of unity such that $\{\rho_i^0\}_{i=1}^N \overset{i.i.d}{\sim} U(1.006, 1.018)$, which can lead to the complicated case such that intensity of signals generated from common factors is not significantly dominant than that from idiosyncratic components.

In addition, in related studies on unit root test (e.g., see Müller and Elliott (2003)), the effects of initial conditions matter for the power of related unit root tests in finite samples. Therefore, we consider three different initial conditions of the unobserved common factors under DGPs 3 and 4: (1) initial values are small such that $\boldsymbol{F}_0^{0,(r)} = 5$ for all $r = 1, \ldots, R_0$; (2) initial values are moderately large such that $\boldsymbol{F}_0^{0,(r)} = 20$ for all $r = 1, \ldots, R_0$; and (3) initial values depend on the length of time series, T, such that $\boldsymbol{F}_0^{0,(r)} = c \cdot T^{0.5-\beta}$ with $c = 1.2$ and $\beta = 0.02$ for all $r = 1, \ldots, R_0$, where $\boldsymbol{F}_0^{0,(r)}$ stands for the r-th true common factor at time 0.

Note that ϵ_{it} and $\boldsymbol{u}_t$ in DGPs 2 and 4 follow the m-dependent process. For each i, we generate e_{it} by the m-dependent process as follows: $e_{it} = \sum_{k=0}^{m} q_k \xi_{t-k}$ with $\xi_{t-k} \overset{i.i.d}{\sim} \mathcal{N}(0,1)$, $q_0 = 1$ and $q_k \overset{i.i.d}{\sim} U[0.1, 0.2]$ for $k \geq 1$; and we also generate $\boldsymbol{u}_t$ by the same style. We fix $m = 5$ across all experiments.

Finally, according to Rho and Shao (2019), by employing the DWB method, we generate pseudo-series $\{W_{t,T}\}_{t=1}^{T}$from $i.i.d.N(0, \Sigma_W)$, where Σ_W is an $T \times T$ matrix with its (t, s)th element being $a\left(|t-s|/l_T\right)$ with the Bartlett kernel $a(|\tau|)$ such that $a(|\tau|) = (1 - |\tau|) \cdot \mathbb{I}(|\tau| \leq 1)$. Besides, Rho and Shao (2019) indicates that the moderate bandwidth l_T for the DWB method works in practice. Thus, we set $l_T = \max\left(\lfloor 6[T/100]^{1/4} \rfloor, 6\right)$, where $\lfloor A \rfloor$ refers to the integer that does not exceed A, and $\max(a, b)$ takes the larger value between a and b.

For all simulation experiments, we conduct 1000 replications with $B = 500$ bootstrap resamples in each replication. For brevity, we report simulation results for the proposed test based on the first estimated factor. The nominal level is 0.05 for all cases.

3.5.2 Simulation Results

We display the finite-sample performances of the $\mathcal{DF}^{(1),f}$ when the null hypothesis holds in Tables 3.1 and 3.2. Rows of Table 3.1 exhibit the empirical rejection frequencies based on bootstrapped and simulated asymptotic critical values for each combination of N and T in the Single-Factor model. We simulate the asymptotic critical values with 2000 replications. Similarly, rows of Table 3.2 report the empirical rejection frequencies when $\{\rho_i^0\}_{i=1}^{N}$ lie in the vicinity of unity in different types for each combination of N and T in the Three-Factors model. Overall, the results in Tables 3.1 and 3.2 suggest that, based on the bootstrapped critical values, our proposed test has well-controlled size in

Table 3.1: Finite sample properties of the $\mathcal{DF}^{(1),f}$ under the null for the Single-Factor Model.

DGP 1	Size of the dependent wild bootstrap implementation of the proposed test					
	$\rho^0_{0,r}=1$ for all r and $\{\rho_i\}_{i=1}^N \overset{i.i.d}{\sim} U(1.001, 1.010)$					
(N,T)	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
bootstrapped c.v.	0.051	0.058	0.057	0.060	0.053	0.048
simulated asym. c.v.	0.449	0.484	0.516	0.439	0.484	0.491
DGP 1	Size of the dependent wild bootstrap implementation of the proposed test					
	$\rho^0_{0,r}=1$ for all r and $\{\rho_i\}_{i=1}^N \overset{i.i.d}{\sim} U(1.006, 1.018)$					
(N,T)	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
bootstrapped c.v.	0.057	0.046	0.055	0.064	0.052	0.049
simulated asym. c.v.	0.432	0.481	0.501	0.450	0.464	0.507
DGP 1	Size of the dependent wild bootstrap implementation of the proposed test					
	$\rho^0_{0,r}=1$ for all r and $\{\rho_i\}_{i=1}^N \overset{i.i.d}{\sim} U(1.014, 1.016)$					
(N,T)	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
bootstrapped c.v.	0.061	0.053	0.060	0.057	0.049	0.056
simulated asym. c.v.	0.457	0.491	0.580	0.437	0.460	0.530

all cases. On the contrary, the results in Table 3.1 imply that the null of unit root can be rejected too frequently if practitioners use the simulated asymptotic critical values in practice, which is the worst case to misidentify the bubbles in the idiosyncratic components as ones in common factors. In particular, the bottom panel of Table 3.2 implies that the DWB implementation of our proposed test still works when error terms $\{\boldsymbol{u}_t\}$ are weakly dependent along the time dimension.

Table 3.2: Finite sample properties of the $\mathcal{DF}^{(1),f}$ under the null for the Three-Factor Model.

DGP 1	Size of the DWB implementation of the proposed test					
	$\rho^0_{0,r}=1$ for all r					
(N,T)	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
$\{\rho_i\}_{i=1}^N \overset{i.i.d}{\sim} U(1.001, 1.010)$	0.042	0.058	0.050	0.042	0.050	0.048
$\{\rho_i\}_{i=1}^N \overset{i.i.d}{\sim} U(1.006, 1.018)$	0.050	0.054	0.054	0.052	0.046	0.052
$\{\rho_i\}_{i=1}^N \overset{i.i.d}{\sim} U(1.014, 1.016)$	0.060	0.050	0.056	0.050	0.046	0.058
DGP 2	Size of the DWB implementation of the proposed test					
	$\rho^0_{0,r}=1$ for all r					
(N,T)	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
$\{\rho_i\}_{i=1}^N \overset{i.i.d}{\sim} U(1.001, 1.010)$	0.064	0.056	0.051	0.060	0.054	0.053
$\{\rho_i\}_{i=1}^N \overset{i.i.d}{\sim} U(1.006, 1.018)$	0.064	0.059	0.053	0.053	0.054	0.055
$\{\rho_i\}_{i=1}^N \overset{i.i.d}{\sim} U(1.014, 1.016)$	0.061	0.054	0.058	0.057	0.055	0.051

In Tables 3.3 and 3.4, we explore the power performance of the $\mathcal{DF}^{(1),f}$ under various initial

Table 3.3: Finite sample properties of the $\mathcal{DF}^{(1),f}$ under the alternatives for the Three-Factor Model.

DGP 3	Power of the DWB implementation of the proposed test					
$\rho_i \overset{i.i.d}{\sim} U(1.006, 1.018)$ for any $i \in \{1, \ldots, N\}$ with $\boldsymbol{F}_0^{0,(r)} = 5$ for all r						
$\boldsymbol{\rho}_0^0 \backslash (N, T)$	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
$diag\,(0.0000, 1.0160, 1.0180)$	0.057	0.241	0.981	0.058	0.253	0.983
$diag\,(1.0160, 1.0170, 1.0180)$	0.055	0.240	0.997	0.054	0.254	0.997
$diag\,(1.0176, 1, 0178, 1.0180)$	0.033	0.325	0.999	0.048	0.360	0.999
$diag\,(1.0175, 1.0180, 1.0185)$	0.041	0.384	1.000	0.043	0.388	1.000

DGP 3	Power of the DWB implementation of the proposed test					
$\rho_i \overset{i.i.d}{\sim} U(1.006, 1.018)$ for any $i \in \{1, \ldots, N\}$ with $\boldsymbol{F}_0^{0,(r)} = 20$ for all r						
$\boldsymbol{\rho}_0^0 \backslash (N, T)$	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
$diag\,(0.0000, 1.0160, 1.0180)$	0.048	0.979	1.000	0.056	0.977	1.000
$diag\,(1.0160, 1.0170, 1.0180)$	0.011	0.989	1.000	0.013	0.993	1.000
$diag\,(1.0176, 1, 0178, 1.0180)$	0.013	0.998	1.000	0.010	0.998	1.000
$diag\,(1.0175, 1.0180, 1.0185)$	0.016	0.997	1.000	0.012	1.000	1.000

DGP 3	Power of the DWB implementation of the proposed test					
$\rho_i \overset{i.i.d}{\sim} U(1.006, 1.018)$ for any $i \in \{1, \ldots, N\}$ with $\boldsymbol{F}_0^{0,(r)} = 1.2 \times T^{0.48}$ for all r						
$\boldsymbol{\rho}_0^0 \backslash (N, T)$	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
$diag\,(0.0000, 1.0160, 1.0180)$	0.056	0.766	1.000	0.067	0.763	1.000
$diag\,(1.0160, 1.0170, 1.0180)$	0.039	0.816	1.000	0.029	0.817	1.000
$diag\,(1.0176, 1, 0178, 1.0180)$	0.029	0.916	1.000	0.024	0.921	1.000
$diag\,(1.0175, 1.0180, 1.0185)$	0.026	0.943	1.000	0.032	0.947	1.000

conditions of common factors when the alternative of local-to-unity holds. For each initial value of the common factors, each panel of Tables 3.3 and 3.4 reports the results for each combination of N and T.

In Table 3.3, our proposed test has high power against the alternatives in most cases. When the initial value of the common factors is relatively small as shown in the top panel, the power of the DWB implementation of our proposed test gains gradually as the time dimension T increases from 51 to 201. However, when the initial value of the common factors is moderately large, as shown in the middle panel, the power of our proposed test against the alternative of local-to-unity approaches one quickly. In addition, when the initial value is of order $o\left(T^{1/2}\right)$, the bottom panel also demonstrates a significant step-up in the power of the DWB implementation of our proposed test against the local alternatives.

Table 3.4: Finite sample properties of the $\mathcal{DF}^{(1),f}$ under the alternatives for the Three-Factor Model.

DGP 4	Power of the DWB implementation of the proposed test					
$\rho_i \overset{i.i.d}{\sim} U(1.006, 1.018)$ for any $i \in \{1, \ldots, N\}$ with $\boldsymbol{F}_0^{0,(r)} = 5$ for all r						
$\boldsymbol{\rho}_0^0 \backslash (N, T)$	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
$diag(0.0000, 1.0160, 1.0180)$	0.968	1.000	1.000	0.967	1.000	1.000
$diag(1.0160, 1.0170, 1.0180)$	0.945	1.000	1.000	0.952	1.000	1.000
$diag(1.0176, 1, 0178, 1.0180)$	0.931	1.000	1.000	0.938	1.000	1.000
$diag(1.0175, 1.0180, 1.0185)$	0.830	1.000	1.000	0.852	1.000	1.000
DGP 4	Power of the DWB implementation of the proposed test					
$\rho_i \overset{i.i.d}{\sim} U(1.006, 1.018)$ for any $i \in \{1, \ldots, N\}$ with $\boldsymbol{F}_0^{0,(r)} = 20$ for all r						
$\boldsymbol{\rho}_0^0 \backslash (N, T)$	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
$diag(0.0000, 1.0160, 1.0180)$	0.805	1.000	1.000	0.806	1.000	1.000
$diag(1.0160, 1.0170, 1.0180)$	0.702	1.000	1.000	0.725	1.000	1.000
$diag(1.0176, 1, 0178, 1.0180)$	0.720	1.000	1.000	0.729	1.000	1.000
$diag(1.0175, 1.0180, 1.0185)$	0.702	1.000	1.000	0.731	1.000	1.000
DGP 4	Power of the DWB implementation of the proposed test					
$\rho_i \overset{i.i.d}{\sim} U(1.006, 1.018)$ for any $i \in \{1, \ldots, N\}$ with $\boldsymbol{F}_0^{0,(r)} = 1.2 \times T^{0.48}$ for all r						
$\boldsymbol{\rho}_0^0 \backslash (N, T)$	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
$diag(0.0000, 1.0160, 1.0180)$	0.915	1.000	1.000	0.911	1.000	1.000
$diag(1.0160, 1.0170, 1.0180)$	0.812	1.000	1.000	0.841	1.000	1.000
$diag(1.0176, 1, 0178, 1.0180)$	0.849	1.000	1.000	0.861	1.000	1.000
$diag(1.0175, 1.0180, 1.0185)$	0.843	1.000	1.000	0.844	1.000	1.000

In Table 3.4, the DWB implementation of our proposed test has similar performance as shown in

Table 3.3. However, when error terms are weakly dependent along the time dimension, our proposed test has significantly higher power against the alternatives than those in Table 3.3 when $T = 51$.

Table 3.5: Finite sample properties of the joint test for the Three-Factor Model.

DGP 2	Size of the DWB implementation of the proposed joint test					
$\rho_{0,r}^0 = 1$ for all r						
(N, T)	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
$\{\rho_i\}_{i=1}^N \overset{i.i.d}{\sim} U(1.001, 1.010)$	0.051	0.052	0.058	0.054	0.046	0.052
$\{\rho_i\}_{i=1}^N \overset{i.i.d}{\sim} U(1.006, 1.018)$	0.048	0.041	0.047	0.055	0.048	0.052
$\{\rho_i\}_{i=1}^N \overset{i.i.d}{\sim} U(1.014, 1.016)$	0.058	0.054	0.060	0.049	0.050	0.049
DGP 4	Power of the DWB implementation of the proposed joint test					
$\rho_i \overset{i.i.d}{\sim} U(1.006, 1.018)$ for any $i \in \{1, \ldots, N\}$ with $\boldsymbol{F}_0^{0,(r)} = 5$ for all r						
$\boldsymbol{\rho}_0^0 \backslash (N, T)$	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
$diag(0.0000, 1.0160, 1.0180)$	0.952	0.998	1.000	0.948	0.999	1.000
$diag(1.0160, 1.0170, 1.0180)$	0.956	1.000	1.000	0.955	1.000	1.000
$diag(1.0176, 1, 0178, 1.0180)$	0.957	1.000	1.000	0.966	1.000	1.000
$diag(1.0175, 1.0180, 1.0185)$	0.958	1.000	1.000	0.964	1.000	1.000
DGP 4	Power of the DWB implementation of the proposed joint test					
$\rho_i \overset{i.i.d}{\sim} U(1.006, 1.018)$ for any $i \in \{1, \ldots, N\}$ with $\boldsymbol{F}_0^{0,(r)} = 20$ for all r						
$\boldsymbol{\rho}_0^0 \backslash (N, T)$	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
$diag(0.0000, 1.0160, 1.0180)$	0.979	1.000	1.000	0.988	1.000	1.000
$diag(1.0160, 1.0170, 1.0180)$	0.991	1.000	1.000	0.994	1.000	1.000
$diag(1.0176, 1, 0178, 1.0180)$	0.987	1.000	1.000	0.990	1.000	1.000
$diag(1.0175, 1.0180, 1.0185)$	0.985	1.000	1.000	0.991	1.000	1.000
DGP 4	Power of the DWB implementation of the proposed joint test					
$\rho_i \overset{i.i.d}{\sim} U(1.006, 1.018)$ for any $i \in \{1, \ldots, N\}$ with $\boldsymbol{F}_0^{0,(r)} = 1.2 \times T^{0.48}$ for all r						
$\boldsymbol{\rho}_0^0 \backslash (N, T)$	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
$diag(0.0000, 1.0160, 1.0180)$	0.972	1.000	1.000	0.977	1.000	1.000
$diag(1.0160, 1.0170, 1.0180)$	0.985	1.000	1.000	0.985	1.000	1.000
$diag(1.0176, 1, 0178, 1.0180)$	0.981	1.000	1.000	0.983	1.000	1.000
$diag(1.0175, 1.0180, 1.0185)$	0.983	1.000	1.000	0.984	1.000	1.000

Table 3.5 summarizes the finite-sample performance of the proposed joint test, and shows that the DWB implementation of our proposed joint test also have desirable size and good power against the alternatives of local-to-unity.

Overall, Tables 3.1-3.5 provide evidences that the DWB implementation of our proposed tests has asymptotic correct size and good asymptotic power properties.

3.6 Concluding Remarks

This paper presents an easy-to-implement and effective testing procedure to detect bubbles in unobserved common factors when data displays the local-to-unity persistence as a necessary complement to related studies on detecting bubbles in panel data. We employ the panel factor models to extract the unobserved common factors and idiosyncratic components by the conventional PC estimation. We propose to construct the proposed test statistics based on the estimated common factors. We establish the limiting null distribution, asymptotic power property, and the consistency for our proposed test when the idiosyncratic components can allow for the local-to-unity process. To implement our test, we further use the DWB method to simulate the critical values in practice and justify the validity of the DWB method. The testing procedure allows us to disentangle the sources of local-to-unity explosiveness in data and identify the bubbles in common factors correctly. These theoretical findings are supplemented with Monte Carlo studies under various scenarios, which display that our proposed test has good finite-sample size and power against the alternative of local-to-unity.

Finally, the findings and discussions of this paper also raise some interesting topics for further exploration. First, identifying the phase transition from local-to-unity to mildly explosive bubbles is an interesting but underdeveloped issue to understand the nature of explosiveness in data. We can utilize advances in statistics and random matrix theory to gain more insights into the dynamics of explosiveness in panel data. Second, closely related to the first topic, we only consider a single bubble in the observed period. In the observed period, unobserved common factors and idiosyncratic components exhibit no structural change from the non-explosion to the explosion. Following the recent studies on high-dimensional time series factor models (e.g., Liu and Chen (2020), Gao and Tsay (2021)), there is a possibility to extend our current analysis for explosive data to such kind of more complicated cases. We leave these topics for future research.

Bibliography

Ando, T. and Bai, J. (2015). A simple new test for slope homogeneity in panel data models with interactive effects. *Economics Letters*, 136:112–117.

Athreya, K. B. and Lahiri, S. N. (2006). *Measure theory and probability theory*. Springer Science & Business Media.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171.

Bai, J. (2004). Estimating cross-section common stochastic trends in nonstationary panel data. *Journal of Econometrics*, 122(1):137–183.

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

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

Bai, J. and Ng, S. (2004). A panic attack on unit roots and cointegration. *Econometrica*, 72(4):1127–1177.

Bai, J. and Ng, S. (2010). Panel unit root tests with cross-section dependence: a further investigation. *Econometric Theory*, 26(4):1088–1114.

Bernstein, D. S. (2009). *Matrix mathematics: theory, facts, and formulas*. Princeton university press.

Biagi, F. and Lucifora, C. (2005). Demographic and education effects on unemployment in europe: Economic factors and labour market institutions.

Boneva, L., Linton, O., and Vogt, M. (2015). A semiparametric model for heterogeneous panel data with fixed effects. *Journal of Econometrics*, 188(2):327–345.

Breitung, J. and Das, S. (2008). Testing for unit roots in panels with a factor structure. *Econometric Theory*, pages 88–108.

Bykhovskaya, A. and Phillips, P. C. (2020). Point optimal testing with roots that are functionally local to unity. *Journal of Econometrics*, 219(2):231–259.

Campello, M., Galvao, A. F., and Juhl, T. (2019). Testing for slope heterogeneity bias in panel data models. *Journal of Business & Economic Statistics*, 37(4):749–760.

Castagnetti, C., Rossi, E., and Trapani, L. (2015). Inference on factor structures in heterogeneous panels. *Journal of Econometrics*, 184(1):145–157.

Chen, B. and Huang, L. (2018). Nonparametric testing for smooth structural changes in panel data models. *Journal of Econometrics*, 202(2):245–267.

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, Y. Z., Phillips, P. C., and Shi, S. (2019). Common bubble detection in large dimensional financial systems. *Macquarie Business School Research Paper Forthcoming*.

Chernozhukov, V., Hansen, C., Liao, Y., and Zhu, Y. (2018). Inference for heterogeneous effects using low-rank estimations. *arXiv preprint arXiv:1812.08089*.

Cheung, A., Roca, E., and Su, J.-J. (2015). Crypto-currency bubbles: an application of the phillips–shi–yu (2013) methodology on mt. gox bitcoin prices. *Applied Economics*, 47(23):2348–2358.

Chudik, A., Pesaran, M. H., and Yang, J.-C. (2018). Half-panel jackknife fixed-effects estimation of linear panels with weakly exogenous regressors. *Journal of Applied Econometrics*, 33(6):816–836.

De Mol, C., Giannone, D., and Reichlin, L. (2008). Forecasting using a large number of predictors: Is bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics*, 146(2):318–328.

Dhaene, G. and Jochmans, K. (2015). Split-panel jackknife estimation of fixed-effect models. *The Review of Economic Studies*, 82(3):991–1030.

Dobriban, E. (2017). Sharp detection in pca under correlations: all eigenvalues matter. *The Annals of Statistics*, 45(4):1810–1833.

Dobriban, E. and Owen, A. B. (2019). Deterministic parallel analysis: an improved method for selecting factors and principal components. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 81(1):163–183.

Dou, L. and Müller, U. K. (2019). Generalized local to unity models. *Work*.

Enoksen, F., Landsnes, C. J., Lučivjanská, K., and Molnár, P. (2020). Understanding risk of bubbles in cryptocurrencies. *Journal of Economic Behavior & Organization*, 176:129–144.

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

Fan, J., Liao, Y., and Yao, J. (2015). Power enhancement in high-dimensional cross-sectional tests. *Econometrica*, 83(4):1497–1541.

Feng, G., Gao, J., Peng, B., and Zhang, X. (2017). A varying-coefficient panel data model with fixed effects: Theory and an application to us commercial banks. *Journal of Econometrics*, 196(1):68–82.

Feng, J. and Su, L. (2020). Testing for idiosyncratic bubbles in the presence of nonstationary or mildly explosive factors in panel data models.

Freyaldenhoven, S. (2019). A generalized factor model with local factors.

Gao, J. and Hong, Y. (2008). Central limit theorems for generalized u-statistics with applications in nonparametric specification. *Journal of Nonparametric Statistics*, 20(1):61–76.

Gao, J., Xia, K., and Zhu, H. (2020). Heterogeneous panel data models with cross-sectional dependence. *Journal of Econometrics*, 219(2):329–353.

Gao, Y., Li, C., and Liang, Z. (2015). Binary response correlated random coefficient panel data models. *Journal of Econometrics*, 188(2):421–434.

Gao, Z. and Tsay, R. S. (2021). Modeling high-dimensional time series: a factor model with dynamically dependent factors and diverging eigenvalues. *Journal of the American Statistical Association*, pages 1–17.

Geuder, J., Kinateder, H., and Wagner, N. F. (2019). Cryptocurrencies as financial bubbles: The case of bitcoin. *Finance Research Letters*, 31.

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

Graham, B. S., Hahn, J., Poirier, A., and Powell, J. L. (2018). A quantile correlated random coefficients panel data model. *Journal of Econometrics*, 206(2):305–335.

Hafner, C. M. (2020). Testing for bubbles in cryptocurrencies with time-varying volatility. *Journal of Financial Econometrics*, 18(2):233–249.

Harvey, D. I., Leybourne, S. J., and Sollis, R. (2015). Recursive right-tailed unit root tests for an explosive asset price bubble. *Journal of Financial Econometrics*, 13(1):166–187.

Harvey, D. I., Leybourne, S. J., and Zu, Y. (2020). Sign-based unit root tests for explosive financial bubbles in the presence of deterministically time-varying volatility. *Econometric Theory*, 36(1):122–169.

Hidalgo, J. and Schafgans, M. (2017). Inference and testing breaks in large dynamic panels with strong cross sectional dependence. *Journal of Econometrics*, 196(2):259–274.

Hildreth, C. and Houck, J. P. (1968). Some estimators for a linear model with random coefficients. *Journal of the American Statistical Association*, 63(322):584–595.

Horie, T. and Yamamoto, Y. (2016). Testing for speculative bubbles in large-dimensional financial panel data sets.

Hsiao, C. (1974). Statistical inference for a model with both random cross-sectional and time effects. *International Economic Review*, pages 12–30.

Hsiao, C. (2014). *Analysis of panel data*. Number 54. Cambridge university press.

Huang, W., Jin, S., Phillips, P. C., and Su, L. (2020). Nonstationary panel models with latent group structures and cross-section dependence. *Journal of Econometrics*.

Juhl, T. and Lugovskyy, O. (2014). A test for slope heterogeneity in fixed effects models. *Econometric Reviews*, 33(8):906–935.

Keane, M. and Neal, T. (2020). Climate change and u.s. agriculture: Accounting for multidimensional slope heterogeneity in panel data. *Quantitative Economics*, 11(4):1391–1429.

Kelejian, H. H. and Prucha, I. R. (2007). Hac estimation in a spatial framework. *Journal of Econometrics*, 140(1):131–154.

Kuersteiner, G. M. and Prucha, I. R. (2013). Limit theory for panel data models with cross sectional dependence and sequential exogeneity. *Journal of Econometrics*, 174(2):107–126.

Lahiri, S. N. (2013). *Resampling methods for dependent data*. Springer Science & Business Media.

Li, D., Qian, J., and Su, L. (2016). Panel data models with interactive fixed effects and multiple structural breaks. *Journal of the American Statistical Association*, 111(516):1804–1819.

Liu, X. and Chen, R. (2020). Threshold factor models for high-dimensional time series. *Journal of Econometrics*, 216(1):53–70.

Lu, X. and Su, L. (2016). Shrinkage estimation of dynamic panel data models with interactive fixed effects. *Journal of Econometrics*, 190(1):148–175.

Lu, X. and Su, L. (2021). Uniform inference in linear panel data models with two-dimensional heterogeneity.

Merlevède, F., Peligrad, M., and Rio, E. (2011). A bernstein type inequality and moderate deviations for weakly dependent sequences. *Probability Theory and Related Fields*, 151(3-4):435–474.

Miao, K., Li, K., and Su, L. (2020). Panel threshold models with interactive fixed effects. *Journal of Econometrics*.

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.

Müller, U. K. and Elliott, G. (2003). Tests for unit roots and the initial condition. *Econometrica*, 71(4):1269–1286.

Onatski, A. (2009). Asymptotics of the principal components estimator of large factor models with weak factors. *manuscript, Columbia University*.

Onatski, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics*, 168(2):244–258.

Onatski, A. and Wang, C. (2020). Spurious factor analysis. Technical report, Faculty of Economics, University of Cambridge.

Peng, B., Su, L., Westerlund, J., and Yang, Y. (2020). Interactive effects panel data models with general factors and regressors. *Available at SSRN 3617477*.

Pesaran, M. H. and Yamagata, T. (2008). Testing slope homogeneity in large panels. *Journal of econometrics*, 142(1):50–93.

Phillips, P. C. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika*, 74(3):535–547.

Phillips, P. C. (1988). Regression theory for near-integrated time series. *Econometrica: Journal of the Econometric Society*, pages 1021–1043.

Phillips, P. C., Shi, S., and Yu, J. (2015). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the s&p 500. *International economic review*, 56(4):1043–1078.

Phillips, P. C., Wu, Y., and Yu, J. (2011). Explosive behavior in the 1990s nasdaq: When did exuberance escalate asset values? *International economic review*, 52(1):201–226.

Phillips, P. C. and Yu, J. (2011). Dating the timeline of financial bubbles during the subprime crisis. *Quantitative Economics*, 2(3):455–491.

Rho, Y. and Shao, X. (2019). Bootstrap-assisted unit root testing with piecewise locally stationary errors. *Econometric Theory*, 35(1):142–166.

Robinson, P. M. (1989). Nonparametric estimation of time-varying parameters. In *Statistical Analysis and Forecasting of Economic Structural Change*, pages 253–264. Springer.

Shao, Q.-M., Yu, H., et al. (1996). Weak convergence for weighted empirical processes of dependent sequences. *The Annals of Probability*, 24(4):2098–2127.

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

Shimer, R. (2001). The impact of young workers on the aggregate labor market. *The Quarterly Journal of Economics*, 116(3):969–1007.

Skans, O. N. (2005). Age effects in swedish local labor markets. *Economics Letters*, 86(3):419–426.

Stock, J. H. and Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, 97(460):1167–1179.

Stock, J. H. and Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162.

Su, L. and Chen, Q. (2013). Testing homogeneity in panel data models with interactive fixed effects. *Econometric Theory*, pages 1079–1135.

Su, L. and Wang, X. (2017). On time-varying factor models: Estimation and testing. *Journal of Econometrics*, 198(1):84–101.

Su, L., Wang, X., and Jin, S. (2017). Sieve estimation of time-varying panel data models with latent structures. *Journal of Business & Economic Statistics*, pages 1–16.

Swamy, P. A. (1970). Efficient inference in a random coefficient regression model. *Econometrica: Journal of the Econometric Society*, pages 311–323.

Wang, W., Phillips, P. C., and Su, L. (2018). Homogeneity pursuit in panel data models: Theory and application. *Journal of Applied Econometrics*, 33(6):797–815.

Westerlund, J. (2015). The power of panic. *Journal of econometrics*, 185(2):495–509.

Whitehouse, E. J. (2019). Explosive asset price bubble detection with unknown bubble length and initial condition. *Oxford Bulletin of Economics and Statistics*, 81(1):20–41.

Xia, Q., Liang, R., and Wu, J. (2017). Transformed contribution ratio test for the number of factors in static approximate factor models. *Computational Statistics & Data Analysis*, 112:235–241.

Zeng, X., Xia, Y., and Zhang, L. (2019). Double cross validation for the number of factors in approximate factor models. *arXiv preprint arXiv:1907.01670*.

Zhang, B., Gao, J., and Pan, G. (2020). Estimation and testing for high-dimensional near unit root time series. *Available at SSRN 3579168*.

Zhang, B., Pan, G., Gao, J., et al. (2018). Clt for largest eigenvalues and unit root testing for high-dimensional nonstationary time series. *Annals of Statistics*, 46(5):2186–2215.

Zhu, X., Guo, X., Wang, T., and Zhu, L. (2020). Dimensionality determination: A thresholding double ridge ratio approach. *Computational Statistics & Data Analysis*, 146:106910.

Zhu, Y. (2017). High-dimensional panel data with time heterogeneity: estimation and inference. *Available at SSRN 2665374*.

Appendix A

Appendix to Chapter 1

A.1 Proofs of the Main Results in Section 1.3

Recall that $\delta_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$.

A.1.1 Proofs of Theorems 1.3.1 and Corollary 1.3.2

In this appendix, we prove Theorem 1.3.1 and Corollary 1.3.2 in the paper. We call upon Proposition A.1.1 below whose proof can be found in the online supplement.

Proofs of Theorem 1.3.1

(i) Note that

$$\widehat{\beta}-\beta = \left(\frac{1}{NT}\sum_i\sum_t \ddot{x}_{it}\ddot{x}_{it}'\right)^{-1}\left[\frac{1}{NT}\sum_i\sum_t \ddot{x}_{it}\left(\ddot{u}_{it}+\ddot{\theta}_{it}+\ddot{\varphi}_{it}\right)\right] \equiv A_{1NT}^{-1}\left(B_{1NT}+B_{2NT}+B_{3NT}\right),$$

where

$$\begin{aligned} A_{1NT} &= \frac{1}{NT}\sum_i\sum_t \ddot{x}_{it}\ddot{x}_{it}',\ B_{1NT} = \frac{1}{NT}\sum_i\sum_t \ddot{x}_{it}\ddot{u}_{it}, \\ B_{2NT} &= \frac{1}{NT}\sum_i\sum_t \ddot{x}_{it}\tilde{x}_{it}'\lambda_i,\ \text{and } B_{3NT} = \frac{1}{NT}\sum_i\sum_t \ddot{x}_{it}\breve{x}_{it}'\gamma_t. \end{aligned}$$

By Proposition A.1.1 below,

$$\begin{aligned}\left\|\widehat{\beta}-\beta\right\| &\leq \left\|A_{1NT}^{-1}\right\|\left(\|B_{1NT}\|+\|B_{2NT}\|+\|B_{3NT}\|\right)\\ &= O_p(1)\,O_p((NT)^{-1/2}+N^{-1/2}+T^{-1/2})=O_p(\delta_{NT}^{-1}).\end{aligned}$$

(ii) By (i), we can rewrite $\widehat{\beta}-\beta=A_{1NT}^{-1}\left(B_{1NT}+B_{2NT}+B_{3NT}\right)$. By neglecting higher-order terms, we can show that

$$\begin{aligned}\widehat{\beta}-\beta &= \left(\frac{1}{NT}\sum_i\sum_t x_{it}^{*}x_{it}^{*\prime}\right)^{-1}\left[\frac{1}{NT}\sum_i\sum_t x_{it}^{*}x_{it}^{*\prime}\lambda_i+\frac{1}{NT}\sum_i\sum_t x_{it}^{*}x_{it}^{*\prime}\gamma_t+\frac{1}{NT}\sum_i\sum_t x_{it}^{*}u_{it}\right]\\ &\quad +o_p\left(\delta_{NT}^{-1}\right)\equiv A_{1NT}^{-1}\left(N^{-1/2}\xi_{NT}+T^{-1/2}\zeta_{NT}+N^{-1/2}T^{-1/2}\phi_{NT}\right)+o_p\left(\delta_{NT}^{-1}\right),\end{aligned}$$

where $\xi_{NT}\equiv\frac{1}{\sqrt{N}T}\sum_i\sum_t x_{it}^{*}x_{it}^{*\prime}\lambda_i$, $\zeta_{NT}\equiv\frac{1}{N\sqrt{T}}\sum_i\sum_t x_{it}^{*}x_{it}^{*\prime}\gamma_t$, and $\phi_{NT}\equiv\frac{1}{\sqrt{NT}}\sum_i\sum_t x_{it}^{*}u_{it}$. By Proposition A.1.1(i) below and the fact that $\Sigma_x>0$, $A_{1NT}^{-1}=\Sigma_x^{-1}+o_p(1)$.

Note that $\mathcal{V}_\beta\equiv\frac{1}{N}\Sigma_x^{-1}V_\lambda\Sigma_x^{-1}+\frac{1}{T}\Sigma_x^{-1}V_\gamma\Sigma_x^{-1}+\frac{1}{NT}\Sigma_x^{-1}V_u\Sigma_x^{-1}=O(\delta_{NT}^{-2})$. Under the stated conditions in the theorem, we will show that $\xi_{NT}\xrightarrow{d}\mathcal{N}(0,V_\lambda)$, $\zeta_{NT}\xrightarrow{d}\mathcal{N}(0,V_\gamma)$, and $\phi_{NT}\xrightarrow{d}\mathcal{N}(0,V_u)$. Next,

$$\begin{aligned}\mathrm{Cov}\left(\xi_{NT},\zeta_{NT}\right) &= E\left[\frac{1}{N^{3/2}T^{3/2}}\sum_{i,j}\sum_{t,s}x_{it}^{*}x_{it}^{*\prime}\lambda_i\gamma_s^{\prime}x_{js}^{*}x_{js}^{*\prime}\right]\\ &= E\left[\frac{1}{N^{3/2}T^{3/2}}\sum_{i,j}\sum_{t,s}x_{it}^{*}x_{it}^{*\prime}\mathrm{Cov}\left(\lambda_i,\gamma_s\mid X\right)x_{js}^{*}x_{js}^{*\prime}\right]=o(1),\end{aligned}$$

by Assumption A1.1(v). Similarly, we can show that $\mathrm{Cov}(\phi_{NT},\zeta_{NT})=o(1)$ and $\mathrm{Cov}(\xi_{NT},\phi_{NT})=o(1)$ by the law of iterated expectations under Assumption A1.1(iii)-(iv) and Assumption A1.2(ii).

So $(\xi_{NT},\zeta_{NT},\phi_{NT})\xrightarrow{d}(\xi,\zeta,\phi)$, where (ξ,ζ,ϕ) is a triple-variate Gaussian random vector with mean zero, and covariance $\mathrm{diag}\{\Sigma_x^{-1}V_\lambda\Sigma_x^{-1},\Sigma_x^{-1}V_\gamma\Sigma_x^{-1},\Sigma_x^{-1}V_u\Sigma_x^{-1}\}$, where $\mathrm{diag}\{\cdot\}$ represents the (block) diagonal matrix. Then, by the almost sure representation, there exists $(\xi_{NT}^{*},\zeta_{NT}^{*},\phi_{NT}^{*})$ and $(\xi^{*},\zeta^{*},\phi^{*})$ such that $(\xi_{NT}^{*},\zeta_{NT}^{*},\phi_{NT}^{*})$ has the same distribution as $(\xi_{NT},\zeta_{NT},\phi_{NT})$, $(\xi^{*},\zeta^{*},\phi^{*})$ has the same distributions as (ξ,ζ,ϕ), then $(\xi^*_{NT},\zeta^*_{NT},\phi^*_{NT})\xrightarrow{d}(\xi^*,\zeta^*,\phi^*)$ almost surely. It follows that

$$\begin{aligned}
\frac{\xi_{NT}/\sqrt{N}+\zeta_{NT}/\sqrt{T}}{\mathcal{V}_\beta^{1/2}} &\stackrel{d}{=} \frac{\xi^*_{NT}/\sqrt{N}+\zeta^*_{NT}/\sqrt{T}}{\mathcal{V}_\beta^{1/2}} \\
&= \underbrace{\frac{\xi^*/\sqrt{N}+\zeta^*/\sqrt{T}}{\mathcal{V}_\beta^{1/2}}}_{\mathcal{N}(0_K,I_K)} + \underbrace{\frac{\xi^*_{NT}-\xi^*}{\left(\Sigma_x^{-1}V_\lambda\Sigma_x^{-1}+\frac{T}{N}\Sigma_x^{-1}V_\gamma\Sigma_x^{-1}+\frac{1}{N}\Sigma_x^{-1}V_u\Sigma_x^{-1}\right)^{1/2}}}_{a_1} \\
&\quad + \underbrace{\frac{\zeta^*_{NT}-\zeta^*}{\left(\frac{N}{T}\Sigma_x^{-1}V_\lambda\Sigma_x^{-1}+\Sigma_x^{-1}V_\gamma\Sigma_x^{-1}+\frac{1}{T}\Sigma_x^{-1}V_u\Sigma_x^{-1}\right)^{1/2}}}_{a_2} \\
&\stackrel{d}{=} \mathcal{N}\left(0_K,I_K\right)+o_{a.s.}(1).
\end{aligned}$$

where $a_1\to 0$ and $a_2\to 0$ almost surely. Therefore,

$$\mathcal{V}_\beta^{-1/2}\left(\widehat{\beta}-\beta\right)\xrightarrow{d}\mathcal{N}(0_K,I_K)$$

Now, we are in the position to show the asymptotic normality of ξ_{NT}, ζ_{NT} and ϕ_{NT} defined above. Then, together with continuous mapping theorem, the desired result holds immediately.

Note that $\xi_{NT}\xrightarrow{d}\mathcal{N}(0,V_\lambda)$ under Assumption A1.5. Next, we apply the CLT for mixing sequences as given in Corollary 16.3.6 in Athreya and Lahiri (2006) to $\zeta_{NT}=\frac{1}{N\sqrt{T}}\sum_i\sum_s x^*_{is}x^{*\prime}_{is}\gamma_t$ and $\phi_{NT}=\frac{1}{\sqrt{NT}}\sum_i\sum_t x^*_{it}u_{it}$ below. Let $\mathbb{Z}_{Ns}\equiv\frac{1}{N}\sum_i x^*_{is}x^{*\prime}_{is}\gamma_s$. Then $\zeta_{NT}=T^{-1/2}\sum_s\mathbb{Z}_{Ns}$ and $\{\mathbb{Z}_{Ns}\}$ is strong mixing sequence with some mixing coefficeint $\alpha_Z(\cdot)$. It suffices to show that (1) $E(\mathbb{Z}_{Ns})=0$; (2) $E\left\|\mathbb{Z}_{Ns}\right\|^4<\infty$; (3) $\sum_\tau^{+\infty}\alpha_Z(\tau)^{1/2}<\infty$; (4) $\Sigma_{\mathbb{Z}}\equiv\lim_{(N,T)\to\infty}\text{Var}(\frac{1}{\sqrt{T}}\sum_s\mathbb{Z}_{Ns})$ exists and is positive definite.

By Assumption A1.1, (1) holds. Next, $E\left\|\mathbb{Z}_{Ns}\right\|^4<\infty$ under Assumption A1.1 by direct moments calculations for each s. Under Assumption A1.1, $\{x^*_{it}\gamma_t\}$ and $\{x^*_{it}\}$ are α-mixing sequences with the α-mixing coefficients that satisfy $\alpha(\tau)\leq M\tau^{-\kappa}$ for some $\kappa>2$. Therefore, $\mathbb{Z}_{Ns}=\frac{1}{N}\sum_i x^*_{is}x^{*\prime}_{is}\gamma_s$ is still the α-mixing sequence with the α-mixing coefficients $\alpha_Z(\tau)$ satisfying

$\alpha_Z(\tau) \leq M\tau^{-\kappa}$ for some $\kappa > 2$. Then $\sum_{\tau=1}^{+\infty} \alpha_Z^{1/2}(\tau) \leq M \sum_{\tau=1}^{+\infty} \tau^{-\frac{\kappa}{2}} < \infty$ and (3) follows. Lastly,

$$\Sigma_{\mathbb{Z}} = \lim_{(N,T)\to\infty} \text{Var}(\frac{1}{\sqrt{T}} \sum_s \mathbb{Z}_{Ns}) = \lim_{(N,T)\to\infty} \frac{1}{N^2 T} \sum_{i,j} \sum_{t,s} E\left[x_{it}^* x_{it}^{*\prime} \gamma_t \gamma_s' x_{js}^* x_{js}^{*\prime}\right] \equiv V_\gamma$$

exists and is positive definite under Assumption A1.4. Then by Corollary 16.3.6 of Athreya and Lahiri (2006), $\xi_{NT} = \frac{1}{\sqrt{T}} \sum_t \mathbb{Z}_t \xrightarrow{d} \mathcal{N}(0, V_\gamma)$. Similarly, we can show that $\phi_{NT} \xrightarrow{d} \mathcal{N}(0, V_u)$. ■

Proposition A.1.1. *Suppose that Assumptions A1.1 to A1.4 hold. Then*

(i) $A_{1NT} \xrightarrow{p} \Sigma_x$,

(ii) $\|B_{1NT}\| = O_p(\delta_{NT}^{-2})$,

(iii) $\|B_{2NT}\| = O_p(\frac{1}{\sqrt{N}})$,

(iv) $\|B_{3NT}\| = O(\frac{1}{\sqrt{T}})$.

Proofs of Corollary 1.3.2

The corollary follows from Theorem 1.3.1 and Slutsky theorem provided that $\widehat{\mathcal{V}}_\beta$ is the consistent estimator of $\mathcal{V}_\beta$. The latter can be shown by direct calculations. ■

A.1.2 Proofs of Theorem 1.3.3 and Corollary 1.3.5

In this section, we prove Theorem 1.3.3 and Corollary 1.3.5.

Proof of Theorem 1.3.3

Recall that $e_{it} = \ddot{x}_{it}'(\beta - \widehat{\beta}) + \ddot{\varphi}_{it} + \ddot{u}_{it} - \frac{1}{N} \sum_{l=1, l\neq i}^{N} \tilde{x}_{lt}' \lambda_l$. It is easy to see that

$$\begin{aligned} \widehat{\lambda}_i - \lambda_i &= \frac{N-1}{N} \left(\frac{1}{T} \sum_t \tilde{x}_{it} \tilde{x}_{it}' \right)^{-1} \frac{1}{T} \sum_t \tilde{x}_{it} e_{it} \\ &= \frac{N-1}{N} A_{2iT}^{-1} \left(-C_{1iT} + C_{2iT} + C_{3iT} + C_{4iT} \right), \end{aligned}$$

where

$$A_{2iT} = \frac{1}{T}\sum_{t}\tilde{x}_{it}\tilde{x}'_{it},$$

$$C_{1iT} = \frac{1}{T}\sum_{t}\tilde{x}_{it}\tilde{x}'_{it}\left(\widehat{\beta}-\beta\right) - \frac{1}{NT}\sum_{l}\sum_{t}\tilde{x}_{it}\tilde{x}'_{lt}\left(\widehat{\beta}-\beta\right),$$

$$C_{2iT} = \frac{1}{T}\sum_{t}\tilde{x}_{it}\breve{x}'_{it}\gamma_t - \frac{1}{T^2}\sum_{t}\sum_{s}\tilde{x}_{it}\breve{x}'_{is}\gamma_s,$$

$$C_{3iT} = \frac{1}{T}\sum_{t}\tilde{x}_{it}\left(\tilde{u}_{it} - \frac{1}{N}\sum_{l}\tilde{u}_{lt}\right),$$

$$C_{4iT} = \frac{1}{NT}\sum_{l\neq i}\sum_{t}\tilde{x}_{it}\tilde{x}'_{lt}\lambda_l.$$

Furthermore, according to those immediate results for $\widehat{\lambda}_i - \lambda_i$ in Section S1, for each i, we then have the following decomposition,

$$\begin{aligned}
\widehat{\lambda}_i - \lambda_i &= -\Bigg[\left(\frac{1}{NT}\sum_{j}\sum_{t}x^*_{jt}x^{*\prime}_{jt}\right)^{-1}\frac{1}{NT}\sum_{j}\sum_{t}x^*_{jt}x^{*\prime}_{jt}\lambda_j + \left(\frac{1}{NT}\sum_{j}\sum_{t}x^*_{jt}x^{*\prime}_{jt}\right)^{-1} \\
&\qquad \frac{1}{NT}\sum_{j}\sum_{t}x^*_{jt}x^{*\prime}_{jt}\gamma_t\Bigg] \\
&\quad + \left(\frac{1}{T}\sum_{t}x^*_{it}x^{*\prime}_{it}\right)^{-1}\frac{1}{T}\sum_{t}x^*_{it}x^{*\prime}_{it}\gamma_t + \left(\frac{1}{T}\sum_{t}x^*_{it}x^{*\prime}_{it}\right)^{-1}\frac{1}{T}\sum_{t}x^*_{it}u_{it} + o_p\left(\delta_{NT}^{-1}\right) \\
&= -\Sigma_x^{-1}\frac{1}{NT}\sum_{j\neq i}\sum_{t}x^*_{jt}x^{*\prime}_{jt}\lambda_j + \Sigma_x^{-1}\frac{1}{T}\sum_{t}\left(x^*_{it}x^{*\prime}_{it} - \Sigma_x\right)\gamma_t + \Sigma_x^{-1}\frac{1}{T}\sum_{t}x^*_{it}u_{it} + o_p\left(\delta_{NT}^{-1}\right) \\
&= -\Sigma_x^{-1}(\xi^\lambda_{NT}/\sqrt{N}) + \Sigma_x^{-1}(\zeta^\lambda_{iT}/\sqrt{T}) + \Sigma_x^{-1}(\Psi^\lambda_{iT}/\sqrt{T}) + o_p\left(\delta_{NT}^{-1}\right),
\end{aligned}$$

where $\xi^\lambda_{NT} = \frac{1}{\sqrt{NT}}\sum_{j\neq i}\sum_t x^*_{jt}x^{*\prime}_{jt}\lambda_j$, $\zeta^\lambda_{iT} = \frac{1}{\sqrt{T}}\sum_t\left(x^*_{it}x^{*\prime}_{it} - \Sigma_x\right)\gamma_t$, and $\Psi^\lambda_{iT} = \frac{1}{\sqrt{T}}\sum_t x^*_{it}u_{it}$. By applying suitable CLT's for ξ^λ_{NT}, ζ^λ_{iT} and Ψ^λ_{iT} in the next subsection, we can show the asymptotic normality of these partial sums of random variables under mild conditions.

Our proof strategy is outlined here: Step (i), we show that ξ^λ_{NT}, ζ^λ_{iT} and Ψ^λ_{iT} follow the normal distribution after appropriate rescaling in turn, Step (ii), all of them are asymptotically pairwise-independent, Step (iii), we use almost sure representation theorem again as done for proofs of Theorem 1.3.1 in the previous section.

Recall that

$$\widehat{\lambda}_i - \lambda_i \;=\; -\Sigma_x^{-1}(\xi_{NT}^{\lambda}/\sqrt{N}) + \Sigma_x^{-1}(\zeta_{iT}^{\lambda}/\sqrt{T}) + \Sigma_x^{-1}(\Psi_{iT}^{\lambda}/\sqrt{T}) + o_p\left(\delta_{NT}^{-1}\right),$$

where $\xi_{NT}^{\lambda} = \frac{1}{\sqrt{NT}}\sum_{j\neq i}\sum_t x_{jt}^* x_{jt}^{*\prime}\lambda_j$, $\zeta_{iT}^{\lambda} = \frac{1}{\sqrt{T}}\sum_t \left(x_{it}^* x_{it}^{*\prime} - \Sigma_x\right)\gamma_t$, and $\Psi_{iT}^{\lambda} = \frac{1}{\sqrt{T}}\sum_t x_{it}^* u_{it}$.

Step (i) We will show the asymptotic normality of ξ_{NT}^{λ}, ζ_{iT}^{λ} and Ψ_{iT}^{λ} as below.

(ia) Under Assumptions A1.4 and A1.5, $\xi_{NT}^{\lambda} \xrightarrow{d} \mathcal{N}\left(0, V_\lambda\right)$, where V_λ exists and is nonsingular.

(ib) Note that for each i, ξ_{iT}^{λ} and Ψ_{iT}^{λ} can be regarded as the partial sums of mixing sequences, the asymptotic normality is proved under Assumptions A1.1 to A1.4 by employing Corollary 16.3.6 in Athreya and Lahiri (2006) here.

To this end, define $\mathbb{Z}_{it} = \{\left(x_{it}^* x_{it}^{*\prime} - \Sigma_x\right)\gamma_t, x_{it}^* u_{it}\}$, then $\{\mathbb{Z}_{it}\}$ are stationary and mixing sequences under Assumptions A1.1 and A1.2.

In order to show the asymptotic normality for $\frac{1}{\sqrt{T}}\sum_t \mathbb{Z}_{it}$, it is sufficient to show following four conditions are satisfied as (N, T) go to infinity simultaneously. In particular, these conditions are: (1) $E\mathbb{Z}_{it} = 0$; (2) $E\left\|\mathbb{Z}_{it}\right\|_2^4 < \infty$ for some $\delta \in (0, +\infty)$; (3) $\sum_\tau^{+\infty}\alpha_Z(\tau)^{1/2} < \infty$; (4) $\Sigma_{\mathbb{Z}} = \lim_{T\to\infty} Var(\frac{1}{\sqrt{T}}\sum_t \mathbb{Z}_{it})$ exists and is positive definite matrix. Now, we are in the position to verify these conditions under Assumptions A1.1 to A1.4.

(1) Obviously, $E\mathbb{Z}_{it} = 0$ under Assumptions A1.1 and A1.2.

(2) $E\left\|\mathbb{Z}_{it}\right\|_2^4 < \infty$ holds for $\mathbb{Z}_{it} = \{\left(x_{it}^* x_{it}^{*\prime} - \Sigma_x\right)\gamma_t, x_{it}^* u_{it}\}$ under Assumption A1.1 and A1.2 by straightforward moments calculations.

(3) Under Assumption A1.1 and A1.2, x_{it} and u_{it} are α-mixing sequences with the α-mixing coefficients that satisfy $\alpha(\tau) \leq M\tau^{-\kappa}$ for some $\kappa > 0$. Under Assumption A1.1, γ_t is independent of x_{is} for any (i, t, s). Therefore, $\mathbb{Z}_{it} = \{\left(x_{it}^* x_{it}^{*\prime} - \Sigma_x\right)\gamma_t, x_{it}^* u_{it}\}$ is still the α-mixing sequence with the α-mixing coefficients $\alpha_Z(\tau)$ that satisfy $\alpha_Z(\tau) \leq M\tau^{-\kappa}$ for some $\kappa > 0$. It follows that

$$\sum_{\tau=1}^{+\infty} \alpha_Z^{1/2} \leq M \sum_{\tau=1}^{+\infty} \tau^{-\frac{\kappa}{2}} < \infty,$$

holds as long as $\kappa > 2$, which is satisfied under Assumptions imposed for Theorem 1.3.3.

(4) We verify the condition by treating $\mathbb{Z}_{it} = (x_{it}^* x_{it}^{*\prime} - \Sigma_x)\,\gamma_t$ and $\mathbb{Z}_{it} = x_{it}^* u_{it}$ separately below.

(4a) By direct moments calculations, we have

$$\lim_{T\to\infty} Var(\frac{1}{\sqrt{T}} \sum_t \mathbb{Z}_{it}) = \lim_{T\to\infty} \frac{1}{T} \sum_{t,s=1}^{T} E\left[(x_{it}^* x_{it}^{*\prime} - \Sigma_x)\,\gamma_t \gamma_s' \,(x_{is}^* x_{is}^{*\prime} - \Sigma_x)\right] \equiv \Sigma_{i,x\gamma}$$

exists and is nonsingular uniformly over t by Assumption A1.1 and A1.4.

(4b) By direct moments calculations, we have

$$\lim_{T\to\infty} Var(\frac{1}{\sqrt{T}} \sum_t \mathbb{Z}_{it}) = \lim_{T\to\infty} \frac{1}{T} \sum_{t,s} E\left(x_{it}^* u_{it} u_{is} x_{is}^{*\prime}\right) = \tilde{\Sigma}_{i,xu}$$

exists and is nonsingular by Assumptions A1.1, A1.2 and A1.4.

Hence, according to the Corollary 16.3.6 of Athreya and Lahiri (2006), we have $\frac{1}{\sqrt{T}} \sum_t \mathbb{Z}_{it} \xrightarrow{d} N(0, \Sigma_{\mathbb{Z}})$, which implies

$$\xi_{iT}^{\lambda} \xrightarrow{d} N(0, \Sigma_{i,x\gamma}), \quad \Psi_{iT}^{\lambda} \xrightarrow{d} N(0, \tilde{\Sigma}_{i,xu})$$

Step (ii) In this step, we want to show, as $(N,T) \to \infty$, $\{\xi_{NT}^{\lambda}, \zeta_{iT}^{\lambda}, \Psi_{iT}^{\lambda}\}$ are asymptotically uncorrelated such that they are asymptotically independent by the virtue of their asymptotic normality. To see this, we recall that $\xi_{NT}^{\lambda} = \frac{1}{\sqrt{NT}} \sum_{j\neq i} \sum_t x_{jt}^* x_{jt}^{*\prime} \lambda_j$, $\zeta_{iT}^{\lambda} = \frac{1}{\sqrt{T}} \sum_t (x_{it}^* x_{it}^{*\prime} - \Sigma_x)\,\gamma_t$, and $\Psi_{iT}^{\lambda} = \frac{1}{\sqrt{T}} \sum_t x_{it}^* u_{it}$.

By straightforward moments calculations and the law of iterated expectations, we can readily show that $Cov(\xi_{NT}^{\lambda}, \zeta_{iT}^{\lambda}) = 0$, $Cov(\xi_{NT}^{\lambda}, \Psi_{iT}^{\lambda}) = 0$, and $Cov(\Psi_{iT}^{\lambda}, \zeta_{iT}^{\lambda}) = 0$ hold as (N,T) go to infinity simultaneously under Assumptions A1.1 and A1.2, which is similar to corresponding arguments of proofs for Theorem 1.3.1.

Step (iii) Recall that

$$\widehat{\lambda}_i - \lambda_i = \xi_{NT}^{\lambda}/\sqrt{N} + \zeta_{iT}^{\lambda}/\sqrt{T} + \Psi_{iT}^{\lambda}/\sqrt{T} + o_p\left(\delta_{NT}^{-1}\right).$$

Given that $\xi_{NT}^{\lambda}, \zeta_{iT}^{\lambda}$ and Ψ_{iT}^{λ} follow the normal distribution, and are asymptotically pairwise-independent as $(N, T) \to \infty$ jointly. It follows that as $(N, T) \to \infty$, $\{\xi_{NT}^{\lambda}, \zeta_{iT}^{\lambda}, \Psi_{iT}^{\lambda}\} \xrightarrow{d} (\xi^{\lambda}, \zeta_i^{\lambda}, \Psi_i^{\lambda})$, where $\left(\xi^{\lambda}, \xi_i^{\xi}, \zeta_i^{\lambda}\right)$ is multivariate Gaussian random vector with mean zero and the covariance matrix defined as $diag\left(V_{\lambda}, \Sigma_{i,x\gamma}, \tilde{\Sigma}_{i,xu}\right)$, where $diag(.)$ represents the (block) diagonal matrix. Therefore, we can follow corresponding arguments of proofs for Theorem 1.3.1 to employ almost sure representation theorem again. Finally, according to continuous mapping theorem, the desired result follows immediately based on notations defined in Theorem 1.3.3. ■

Proposition A.1.2. *Under Assumptions A1.1 to A1.4, as $(N, T) \to \infty$ simultaneously, the followings hold,*

(i) $A_{2iT} \xrightarrow{p} \Sigma_x$,

(ii) $\|C_{1iT}\| = O_p\left(\left\|\widehat{\beta} - \beta\right\|\right)$,

(iii) $\|C_{2iT}\| = O_p(\frac{1}{\sqrt{T}})$,

(iv) $\|C_{3iT}\| = O_p(\frac{1}{\sqrt{T}})$,

(v) $\|C_{4iT}\| = O_p(\frac{1}{\sqrt{NT}})$.

Proof of Corollary 1.3.5.

Provided that $\widehat{\mathcal{V}}_{\lambda i}$ is constructed as the plug-in estimator of $\mathcal{V}_{\lambda i}$, its consistency can be verified by direct calculations and the applications of Law of Large Numbers under imposed Assumptions we impose, Theorem 1.3.4 and Proposition 1.3.8. Finally it suffices to use the Slutsky's lemma to obtain the desired result.

In the proofs of Theorem 1.4.1 in the later section, we give a sketch of the main arguments for the uniform consistency of $\widehat{\mathcal{V}}_{\lambda i}$. The details of proofs for Corollary 1.3.5 here are very similar to those justifications. We refer readers to the corresponding arguments of proofs for Theorem 1.4.1. ■

A.1.3 Proofs of Theorem 1.3.4 and Corollary 1.3.6

In this section, we prove Theorem 1.3.4 and Corollary 1.3.6.

Proof of Theorem 1.3.4

Recall that $\nu_{it} = \ddot{x}_{it}'(\beta - \widehat{\beta}) + \ddot{\theta}_{it} + \ddot{u}_{it} - \frac{1}{T}\sum_{r\neq t}\breve{x}_{ir}'\gamma_r$. Note that

$$\begin{aligned}
\widehat{\gamma}_t - \gamma_t &= \frac{T}{T-1}\left(\frac{1}{N}\sum_i \breve{x}_{it}\breve{x}_{it}'\right)^{-1}\left[\frac{1}{N}\sum_i \breve{x}_{it}\nu_{it}\right] \\
&= \frac{T}{T-1}A_{3Nt}^{-1}\left(-D_{1Nt} + D_{2Nt} + D_{3Nt} + D_{4Nt}\right),
\end{aligned}$$

where

$$\begin{aligned}
A_{3Nt} &= \frac{1}{N}\sum_i \breve{x}_{it}\breve{x}_{it}', \\
D_{1Nt} &= \frac{1}{N}\sum_i \breve{x}_{it}\tilde{x}_{it}'\left(\widehat{\beta}_{OLS} - \beta\right) - \frac{1}{N^2}\sum_i\sum_l \breve{x}_{it}\tilde{x}_{lt}'\left(\widehat{\beta}_{OLS} - \beta\right), \\
D_{2Nt} &= \frac{1}{N}\sum_i \breve{x}_{it}\tilde{x}_{it}'\lambda_i - \frac{1}{N^2}\sum_i\sum_l \breve{x}_{it}\tilde{x}_{ls}'\lambda_l, \\
D_{3Nt} &= \frac{1}{N}\sum_i \breve{x}_{it}\left(\tilde{u}_{it} - \frac{1}{N}\sum_l \tilde{u}_{lt}\right) = \frac{1}{N}\sum_i \breve{x}_{it}\tilde{u}_{it} - \frac{1}{N^2}\sum_l\sum_i \breve{x}_{it}\tilde{u}_{lt}, \\
D_{4Nt} &= \frac{1}{NT}\sum_i\sum_{t\neq s}\breve{x}_{it}\breve{x}_{is}'\gamma_s.
\end{aligned}$$

Furthermore, according to those immediate results for $\widehat{\gamma}_t - \gamma_t$ in Section S1, for each t, we then have the following decomposition,

$$\begin{aligned}
&\widehat{\gamma}_t - \gamma_t \\
= \;& -\left[\left(\frac{1}{NT} \sum_i \sum_s x_{is}^* x_{is}^{*\prime} \right)^{-1} \frac{1}{NT} \sum_i \sum_s x_{is}^* x_{is}^{*\prime} \lambda_i + \left(\frac{1}{NT} \sum_i \sum_s x_{is}^* x_{is}^{*\prime} \right)^{-1} \frac{1}{NT} \sum_i \sum_s x_{is}^* x_{is}^{*\prime} \gamma_s \right] \\
& + \left(\frac{1}{N} \sum_i x_{it}^* x_{it}^{*\prime} \right)^{-1} \frac{1}{N} \sum_i x_{it}^* x_{it}^{*\prime} \lambda_i + \left(\frac{1}{N} \sum_i x_{it}^* x_{it}^{*\prime} \right)^{-1} \frac{1}{N} \sum_i x_{it}^* u_{it} + o_p\left(\delta_{NT}^{-1}\right) \\
= \;& -\Sigma_x^{-1} \frac{1}{NT} \sum_i \sum_{s \neq t} x_{is}^* x_{is}^{*\prime} \gamma_s + \Sigma_x^{-1} \frac{1}{N} \sum_i \left(x_{it}^* x_{it}^{*\prime} - \Sigma_x \right) \lambda_i + \Sigma_x^{-1} \frac{1}{N} \sum_i x_{it}^* u_{it} + o_p\left(\delta_{NT}^{-1}\right) \\
= \;& -\Sigma_x^{-1} (\xi_{NT}^{\gamma} / \sqrt{T}) + \Sigma_x^{-1} (\zeta_{Nt}^{\gamma} / \sqrt{N}) + \Sigma_x^{-1} (\Psi_{Nt}^{\gamma} / \sqrt{N}) + o_p\left(\delta_{NT}^{-1}\right),
\end{aligned}$$

where $\xi_{NT}^{\gamma} = \frac{1}{N\sqrt{T}} \sum_i \sum_{s \neq t} x_{is}^* x_{is}^{*\prime} \gamma_s$, $\zeta_{Nt}^{\gamma} = \frac{1}{\sqrt{N}} \sum_i \left(x_{it}^* x_{it}^{*\prime} - \Sigma_x \right) \lambda_i$, and $\Psi_{Nt}^{\gamma} = \frac{1}{\sqrt{N}} \sum_i x_{it}^* u_{it}$. By applying suitable CLT's for ξ_{NT}^{γ}, ζ_{Nt}^{γ} and Ψ_{Nt}^{γ} in the next subsection, we can show the asymptotic normality of these partial sums of random variables under mild conditions.

As discussed in previous subsection, the asymptotic normality of $\widehat{\gamma}_t$ is related to the asymptotic normality of ξ_{NT}^{γ}, ζ_{Nt}^{γ} and Ψ_{Nt}^{γ}. Our proof strategy is outlined here: Step (i), we show that ξ_{NT}^{γ}, ζ_{Nt}^{γ} and Ψ_{Nt}^{γ} follow the normal distribution after appropriate rescaling, Step (ii), all of them are asymptotically pairwise-independent, Step (iii), we use almost sure representation as we did for proofs of Theorem 1.3.1 and Theorem 1.3.3 in previous sections.

In short, the details of proofs below are very similar to those for the asymptotic normality of $\widehat{\lambda}_i - \lambda_i$ in the previous section. Thus, we give proofs briefly below.

Recall that

$$\widehat{\gamma}_t - \gamma_t = -\Sigma_x^{-1} (\xi_{NT}^{\gamma} / \sqrt{T}) + \Sigma_x^{-1} (\zeta_{Nt}^{\gamma} / \sqrt{N}) + \Sigma_x^{-1} (\Psi_{Nt}^{\gamma} / \sqrt{N}) + o_p\left(\delta_{NT}^{-1}\right),$$

where $\xi_{NT}^{\gamma} = \frac{1}{N\sqrt{T}} \sum_i \sum_{s \neq t} x_{is}^* x_{is}^{*\prime} \gamma_s$, $\zeta_{Nt}^{\gamma} = \frac{1}{\sqrt{N}} \sum_i \left(x_{it}^* x_{it}^{*\prime} - \Sigma_x \right) \lambda_i$, and $\Psi_{Nt}^{\gamma} = \frac{1}{\sqrt{N}} \sum_i x_{it}^* u_{it}$.

Step(i) We now establish the asymptotic distribution of ξ_{NT}^{γ}, ζ_{Nt}^{λ} and Ψ_{Nt}^{γ} as below.

(ia) Note that ξ_{NT}^{γ} can be regarded as the partial sums of mixing sequences that CLT for mixing sequences such as Corollary 16.3.6 in Athreya and Lahiri (2006) can be applied.

Define $T^{-1/2}\sum_{s\neq t}\mathbb{Z}_{Ns,-t} = T^{-1/2}\sum_{s\neq t}\left(\frac{1}{N}\sum_i x_{is}^{*}x_{is}^{*\prime}\right)\gamma_s$, then it is straightforward to see $\{\mathbb{Z}_{Ns,-t}\}$ are mixing sequences over s by construction. In order to show the center limit theorem for $T^{-1/2}\sum_s \mathbb{Z}_{Ns,-t}$, we just need to show that, for each t, four conditions are satisfied, namely, (1) $E\mathbb{Z}_{Ns,-t} = 0$; (2) $E\left\|\mathbb{Z}_{Ns,-t}\right\|_2^4 < \infty$ for some $\delta \in (0,+\infty)$; (3) $\sum_{\tau}^{+\infty}\alpha_Z(\tau)^{1/2} < \infty$; (4) $\Sigma_{\mathbb{Z}} = \lim_{T\to\infty} Var(\frac{1}{\sqrt{T}}\sum_s \mathbb{Z}_{Ns,-t})$ exists and is positive definite matrix. Following corresponding arguments of proofs for Theorem 1.3.1, we can readily verify these conditions under Assumptions A1.1 to A1.4.

Then, under Assumptions A1.1 and A1.4, it follows that $\xi_{NT}^{\gamma} \xrightarrow{d} \mathcal{N}\left(0, V_{\gamma}\right)$

(ib) Under Assumption A1.5, $\zeta_{Nt}^{\gamma} \xrightarrow{d} \mathcal{N}\left(0, \Sigma_{t,x\lambda}\right)$ and $\Psi_{Nt}^{\gamma} \xrightarrow{d} \mathcal{N}\left(0, \breve{\Sigma}_{t,xu}\right)$.

Step (ii) According to the definition of covariance, we want to show, as $(N,T)\to\infty$, $\{\xi_{NT}^{\gamma}, \zeta_{Nt}^{\lambda}, \Psi_{Nt}^{\lambda}\}$ are asymptotically uncorrelated such that they are asymptotically independent. To see this, we recall $\xi_{NT}^{\gamma} = \frac{1}{N\sqrt{T}}\sum_i\sum_{s\neq t} x_{is}^{*}x_{is}^{*\prime}\gamma_s$, $\zeta_{Nt}^{\lambda} = \frac{1}{\sqrt{N}}\sum_i\left(x_{it}^{*}x_{it}^{*\prime} - \Sigma_x\right)\lambda_i$, and $\Psi_{Nt}^{\gamma} = \frac{1}{\sqrt{i}}\sum_i x_{it}^{*}u_{it}$.

By direct moments calculations, we can readily obtain that $Cov(\xi_{NT}^{\gamma}, \zeta_{Nt}^{\lambda}) = 0$, $Cov(\xi_{NT}^{\gamma}, \Psi_{Nt}^{\lambda}) = 0$, $Cov(\Psi_{Nt}^{\lambda}, \zeta_{Nt}^{\lambda}) = 0$ holds under Assumptions A1.1 and A1.2 as (N,T) go to infinity simultaneously, which is similar to corresponding arguments of proofs for Theorem 1.3.1.

Step (iii) Recall that

$$\widehat{\gamma}_t - \gamma_t = -\Sigma_x^{-1}(\xi_{NT}^{\gamma}/\sqrt{T}) + \Sigma_x^{-1}(\zeta_{Nt}^{\gamma}/\sqrt{N}) + \Sigma_x^{-1}(\Psi_{Nt}^{\gamma}/\sqrt{N}) + o_p\left(\delta_{NT}^{-1}\right).$$

Since ξ_{NT}^{γ}, ζ_{Nt}^{λ}, and Ψ_{Nt}^{λ} follow the normal distribution and are asymptotically pairwise-independent as (N,T) go to infinity jointly, it follows that $\left(\xi_{NT}^{\gamma}, \zeta_{Nt}^{\lambda}, \Psi_{Nt}^{\lambda}\right) \xrightarrow{d} (\xi^{\gamma}, \zeta_t^{\gamma}, \Psi_t^{\gamma})$, where $(\xi^{\gamma}, \zeta_t^{\gamma}, \Psi_t^{\gamma})$ is multivariate Gaussian random vector that has mean zero and the covariance matrix $diag\left(V_{\gamma}, \Sigma_{t,x\lambda}, \breve{\Sigma}_{t,xu}\right)$. We then can employ almost sure representation theorem again, which is similar to the corresponding arguments of proofs for Theorem 1.3.1. Finally, together with continuous mapping theorem, based on notations defined in Theorem 1.3.4, the desired result follows immediately. ■

Proposition A.1.3. *Under Assumptions A1.1 to A1.4, as* $(N, T) \to \infty$ *simultaneously, the followings hold:*

(i) $A_{3Nt} \xrightarrow{p} \Sigma_x$,

(ii) $\|D_{1Nt}\| = O_p\left(\left\|\widehat{\beta} - \beta\right\|\right)$,

(iii) $\|D_{2Nt}\| = O_p\left(\frac{1}{\sqrt{N}}\right)$,

(iv) $\|D_{3Nt}\| = O_p\left(\frac{1}{\sqrt{N}}\right)$,

(v) $\|D_{4Nt}\| = O(\frac{1}{\sqrt{N^2T}}) + O(\frac{1}{\sqrt{NT^2}})$.

Proof of Corollary 1.3.6

Provided that $\widehat{\mathcal{V}}_{\gamma t}$ is the plug-in estimator of $\mathcal{V}_{\gamma t}$ by construction, its consistency can be verified by direct calculations and the applications of Law of Large Numbers under imposed Assumptions we impose, Theorem 1.3.3, and Proposition 1.3.8. Finally, the result holds directly by the Slutsky's lemma.

The details of proofs for Corollary 1.3.5 here are very similar to those justifications of uniform consistency of $\widehat{\mathcal{V}}_{\lambda i}$. We refer readers to the corresponding arguments of proofs for Theorem 1.4.1 in the later section. ■

A.1.4 Proofs of Theorem 1.3.7

Note that, for $\widehat{\upsilon}_t$, according to the decomposition for $\left(\widehat{\lambda}_i - \lambda_i\right)$ in the proofs of Theorem 1.3.3 and those immediate results in the online supplementary material, we have

$$\begin{aligned}
&\frac{1}{N}\widehat{\upsilon}_t = \frac{1}{N}\sum_i \breve{x}_{it}\tilde{x}_{it}'\widehat{\lambda}_i = \frac{1}{N}\sum_i \breve{x}_{it}^*\tilde{x}_{it}^{*\prime}\lambda_i + \frac{1}{N}\sum_i \breve{x}_{it}^*\tilde{x}_{it}^{*\prime}\left(\widehat{\lambda}_i - \lambda_i\right) \\
= \;& \frac{1}{N}\sum_i \breve{x}_{it}^*\tilde{x}_{it}^{*\prime}\lambda_i - \left(\frac{1}{N}\sum_i \breve{x}_{it}^*\tilde{x}_{it}^{*\prime}\right)\left(\frac{1}{NT}\sum_j\sum_s x_{js}^*x_{js}^{*\prime}\right)^{-1}\left(\frac{1}{NT}\sum_j\sum_s x_{js}^*x_{js}^{*\prime}\lambda_j\right) \\
&+\frac{1}{N}\sum_i \breve{x}_{it}^*\tilde{x}_{it}^{*\prime}\left[\left(\frac{1}{T}\sum_t x_{it}^*x_{it}^{*\prime}\right)^{-1}\frac{1}{T}\sum_s x_{is}^*x_{is}^{*\prime}\gamma_s\right] \\
&-\frac{1}{N}\sum_i \breve{x}_{it}^*\tilde{x}_{it}^{*\prime}\left[\left(\frac{1}{NT}\sum_j\sum_t x_{jt}^*x_{jt}^{*\prime}\right)^{-1}\frac{1}{NT}\sum_j\sum_t x_{jt}^*x_{jt}^{*\prime}\gamma_t\right] \\
&+\frac{1}{N}\sum_i \breve{x}_{it}^*\tilde{x}_{it}^{*\prime}\left[\left(\frac{1}{T}\sum_t x_{it}^*x_{it}^{*\prime}\right)^{-1}\frac{1}{T}\sum_s x_{is}^*u_{is}\right] + O_p\left(N^{-1}+T^{-1}\right) \\
= \;& \frac{1}{N}\sum_i x_{it}^*x_{it}^{*\prime}\lambda_i - \frac{1}{T}\sum_j\sum_s x_{js}^*x_{js}^{*\prime}\lambda_j + \frac{1}{N}\sum_i x_{it}^*x_{it}^{*\prime}\left[\Sigma_x^{-1}\frac{1}{T}\sum_s\left(x_{is}^*x_{is}^{*\prime} - \Sigma_x\right)\gamma_s\right] \\
&+\frac{1}{N}\sum_i x_{it}^*x_{it}^{*\prime}\left[\Sigma_x^{-1}\frac{1}{T}\sum_s x_{is}^*u_{is}\right] + O_p\left(N^{-1}+T^{-1}\right) \\
= \;& \frac{1}{N}\sum_i\left(x_{it}^*x_{it}^{*\prime} - \frac{1}{T}\sum_s x_{is}^*x_{is}^{*\prime}\right)\lambda_i + \frac{1}{T}\sum_s\left(\frac{1}{N}\sum_i x_{it}^*x_{it}^{*\prime}\Sigma_x^{-1}\left(x_{is}^*x_{is}^{*\prime} - \Sigma_x\right)\right)\gamma_s \\
&+\frac{1}{T}\sum_s\left(\frac{1}{N}\sum_i x_{it}^*x_{it}^{*\prime}\Sigma_x^{-1}x_{is}^*u_{is}\right) + O_p\left(N^{-1}+T^{-1}\right) \\
\equiv \;& E_{1t} + E_{2t} + E_{3t} + O_p\left(N^{-1}+T^{-1}\right)
\end{aligned}$$

Following lines in the online supplementary materials, we can readily show that $E_{1t} = O_p\left(1/\sqrt{N}\right)$, $E_{1t} = O_p\left(1/\sqrt{NT}\right)$ and $E_{1t} = O_p\left(1/\sqrt{NT}\right)$, then, it follows that,

$$
\begin{aligned}
\frac{1}{\sqrt{N}}\widehat{\upsilon}_t &= \frac{1}{\sqrt{N}}\sum_i \left(x_{it}^* x_{it}^{*\prime} - \frac{1}{T}\sum_s x_{is}^* x_{is}^{*\prime}\right)\lambda_i + O_p\left(N^{-1/2} + N^{1/2}T^{-1}\right) \\
&= \frac{1}{\sqrt{N}}\sum_i \left(x_{it}^* x_{it}^{*\prime} - \Sigma_x\right)\lambda_i + O_p\left(N^{-1/2} + N^{1/2}T^{-1}\right)
\end{aligned}
$$

It follows that

$$
\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_t \left(\frac{1}{N}\widehat{\upsilon}_t'\widehat{\upsilon}_t - \widehat{\mathcal{B}}_\upsilon\right) \\
=& \frac{1}{N\sqrt{T}}\sum_i\sum_j\sum_t \lambda_i'\left(x_{jt}^* x_{jt}^{*\prime} - \Sigma_x\right)\left(x_{jt}^* x_{it}^{*\prime} - \Sigma_x\right)\lambda_j - \frac{1}{\sqrt{T}}\sum_t \widehat{\mathcal{B}}_\upsilon + O_p\left(T^{1/2}N^{-1} + NT^{-3/2}\right) \\
=& \frac{1}{\sqrt{T}}\sum_t \left(\mathcal{Z}_{t,\upsilon} - E\mathcal{Z}_{t,\upsilon}\right) + \frac{1}{\sqrt{T}}\sum_t \left(E\mathcal{Z}_{t,\upsilon} - \widehat{\mathcal{B}}_\upsilon\right) + O_p\left(T^{1/2}N^{-1} + NT^{-3/2}\right) \\
\equiv& \frac{1}{\sqrt{T}}\sum_t \zeta_{t,\upsilon} + \frac{1}{\sqrt{T}}\sum_t \left(E\mathcal{Z}_{t,\upsilon} - \widehat{\mathcal{B}}_\upsilon\right) + O_p\left(T^{1/2}N^{-1} + NT^{-3/2}\right)
\end{aligned}
$$

where $\mathcal{Z}_{t,\upsilon} \equiv \frac{1}{N}\sum_i\sum_j \lambda_i'\left(x_{jt}^* x_{jt}^{*\prime} - \Sigma_x\right)\left(x_{jt}^* x_{it}^{*\prime} - \Sigma_x\right)\lambda_j$.

First, by construction, $\{\zeta_{t,\upsilon}\}$ are mixing sequence with zero mean, following lines in the proofs of Theorem 1.3.1 with respect to the CLT for the mixing sequences, the asymptotic normality of $\frac{1}{\sqrt{T}}\sum_t \zeta_{t,\upsilon}$ follows directly. Namely,

$$
\frac{1}{\sqrt{T}}\sum_t \zeta_{t,\upsilon} \xrightarrow{d} \mathcal{N}\left(0, \mathcal{V}_\upsilon\right),
$$

where $\mathcal{V}_\upsilon = \frac{1}{T}\sum_{t,s}\zeta_{t,\upsilon}'\zeta_{s,\upsilon}$.

Then, it suffices to show that $\frac{1}{\sqrt{T}}\sum_t \widehat{\mathcal{B}}_\upsilon = \frac{1}{\sqrt{T}}\sum_t E\mathcal{Z}_{t,\upsilon} + o_p(1)$ as (N, T) go to infinity jointly. Note that, $\widehat{\mathcal{B}}_\upsilon$ is the plug-in estimator by construction, then, by direct calculations and the law of large numbers, we have

$$\begin{aligned}
\frac{1}{\sqrt{T}}\sum_t \widehat{\mathcal{B}}_v &= \frac{1}{\sqrt{T}}\sum_t E\mathcal{Z}_{t,v} + \frac{1}{N\sqrt{T}}\sum_{i,j}\sum_t \left(\widehat{\lambda}_j - \lambda_j\right)'\left(x_{jt}^* x_{jt}^{*\prime} - \Sigma_x\right)\left(x_{it}^* x_{it}^{*\prime} - \Sigma_x\right)\lambda_i \\
&+ \frac{1}{N\sqrt{T}}\sum_{i,j}\sum_t \lambda_j'\left(x_{jt}^* x_{jt}^{*\prime} - \Sigma_x\right)\left(x_{it}^* x_{it}^{*\prime} - \Sigma_x\right)\left(\widehat{\lambda}_i - \lambda_i\right) \\
&+ \frac{1}{N\sqrt{T}}\sum_{i,j}\sum_t \left(\widehat{\lambda}_j - \lambda_j\right)'\left(x_{jt}^* x_{jt}^{*\prime} - \Sigma_x\right)\left(x_{it}^* x_{it}^{*\prime} - \Sigma_x\right)\left(\widehat{\lambda}_i - \lambda_i\right) + \text{dominated terms} \\
&\equiv \frac{1}{\sqrt{T}}\sum_t E\mathcal{Z}_{t,v} + \mathcal{H}_1 + \mathcal{H}_2 + \mathcal{H}_3 + \text{dominated terms.}
\end{aligned}$$

Thus, it suffices to bound $\mathcal{H}_1$, $\mathcal{H}_2$ and $\mathcal{H}_3$. For $\mathcal{H}_3$, according to the decomposition for $\widehat{\lambda}_i - \lambda_i$ in the proofs of Theorem 1.3.3 and those immediate results in the online supplementary material, we have

$$\begin{aligned}
\|\mathcal{H}_3\|_2 &\leq \frac{N}{\sqrt{T}}\sum_t \left\|\frac{1}{N}\sum_i \left(x_{it}^* x_{it}^{*\prime} - \Sigma_x\right)\left(\widehat{\lambda}_i - \lambda_i\right)\right\|_2^2 \\
&\leq \frac{N}{\sqrt{T}}\sum_t \left\|\left[\frac{1}{N}\sum_i \left(x_{it}^* x_{it}^{*\prime} - \Sigma_x\right)\right]\Sigma_x^{-1}\left[\frac{1}{NT}\sum_j\sum_s x_{js}^* x_{js}^{*\prime}\lambda_j\right]\right\|_2^2 \\
&+ \frac{N}{\sqrt{T}}\sum_t \left\|\frac{1}{T}\sum_s\left[\frac{1}{N}\sum_i \left(x_{it}^* x_{it}^{*\prime} - \Sigma_x\right)\Sigma_x^{-1}\left(x_{is}^* x_{is}^{*\prime} - \Sigma_x\right)\right]\gamma_s\right\|_2^2 \\
&+ \frac{N}{\sqrt{T}}\sum_t \left\|\frac{1}{T}\sum_s\left(\frac{1}{N}\sum_i \left(x_{it}^* x_{it}^{*\prime} - \Sigma_x\right)\Sigma_x^{-1} x_{is}^* u_{is}\right)\right\|_2^2 + O_p\left(T^{1/2}N^{-1} + NT^{-3/2}\right) \\
&\equiv \frac{N}{\sqrt{T}}\sum_t \left(\|\mathcal{H}_{31t}\|_2^2\right) + \frac{N}{\sqrt{T}}\sum_t \left(\|\mathcal{H}_{32t}\|_2^2\right) + \frac{N}{\sqrt{T}}\sum_t \left(\|\mathcal{H}_{33t}\|_2^2\right) + O_p\left(T^{1/2}N^{-1} + NT^{-3/2}\right).
\end{aligned}$$

For each t, we can follow the lines for those lemmas in the online supplementary to show that $\|\mathcal{H}_{31t}\|_2 = O_p\left(N^{-1}\right)$, $\|\mathcal{H}_{32t}\|_2 = O_p\left(N^{-1/2}T^{-1/2}\right)$ and $\|\mathcal{H}_{33t}\|_2 = O_p\left(N^{-1/2}T^{-1/2}\right)$. Thus, it follows that $\|\mathcal{H}_3\|_2 = O_p\left(T^{1/2}N^{-1} + T^{-1/2} + NT^{-3/2}\right) = o_p(1)$.

For $\mathcal{H}_2$, according to the decomposition for $\left(\widehat{\lambda}_i - \lambda_i\right)$ in the proofs of Theorem 1.3.3 and following lines for those lemmas in the online supplementary materials, by lengthy arguments, we can show that $\|\mathcal{H}_1\|_2 = O_p\left(N^{-1/2} + T^{-1/2} + N^{1/2}T^{-1}\right) = o_p(1)$ and $\|\mathcal{H}_2\|_2 = O_p\left(N^{-1/2} + T^{-1/2} + N^{1/2}T^{-1}\right) = o_p(1)$ as done for $\|\mathcal{H}_3\|_2$ just above.

Because $\widehat{\mathcal{V}}_v = \frac{1}{T}\sum_{t,s}\widehat{\zeta}'_{t,v}\widehat{\zeta}_{s,v}$, $\widehat{\zeta}_{t,v} = \widehat{\mathcal{Z}}_{t,v} - \widehat{\mathcal{B}}_v$, and $\widehat{\mathcal{Z}}_{t,v} \equiv \frac{1}{N}\sum_{i,j}\left(\breve{x}_{it}\breve{x}'_{it} - \widehat{\breve{\Sigma}}_X\right)\widehat{\lambda}_i\widehat{\lambda}'_j\left(\breve{x}_{jt}\breve{x}'_{jt} - \widehat{\breve{\Sigma}}_X\right)'$, $\widehat{\mathcal{V}}_v$ is also the plug-in estimator for $\mathcal{V}_v$. By direct and lengthy calculations, similar to those arguments for $\widehat{\mathcal{B}}_v$ above, we also can show that $\widehat{\mathcal{V}}_v - \mathcal{V}_v = o_p(1)$ and omit the details here.

Finally, by the continuous mapping theorem and the CLT for the mixing sequences, the desired result follows.

A.1.5 Proofs of Proposition 1.3.8

Proposition 3.8 states the maximal deviations of the estimates of $\{\lambda_i\}_{i=1}^N$ and $\{\gamma_t\}_{t=1}^T$ from the corresponding true values as below,

(1) $\|\widehat{\lambda} - \lambda\|_\infty = O_p\left(\sqrt{\frac{\ln N}{T}}\right)$,

(2) $\|\widehat{\gamma} - \gamma\|_\infty = O_p\left(\sqrt{\frac{\ln T}{N}}\right)$, where $\widehat{\lambda} \equiv \left(\widehat{\lambda}'_1, .., \widehat{\lambda}'_N\right)' \in \mathbb{R}^{NK}$, $\widehat{\gamma} \equiv (\widehat{\gamma}'_1, .., \widehat{\gamma}'_T)' \in \mathbb{R}^{TK}$.

Uniform Consistency of $\widehat{\gamma}_t$

Note that $\|\widehat{\gamma} - \gamma\|_\infty = \max_t \|\widehat{\gamma}_t - \gamma_t\|_\infty \leq \max_t \|\widehat{\gamma}_t - \gamma_t\|_2$, then it suffices to bound $\max_t \|\widehat{\gamma}_t - \gamma_t\|_2$ The proofs are very similar to those arguments for the uniform consistency of $\widehat{\gamma}_i$ in the above subsection, and highly repeated, so here we only give proofs briefly for the ease of exposition.

Neglecting those terms that are of smaller order, we have

$$\widehat{\gamma}_t - \gamma_t \quad = \quad -\left(\frac{1}{NT}\sum_i\sum_s x^*_{is}x^{*\prime}_{is}\right)^{-1}\frac{\xi^\gamma_{NT}}{\sqrt{T}} + \left(\frac{1}{N}\sum_i x^*_{it}x^{*\prime}_{it}\right)^{-1}\left(\frac{\Psi^\gamma_{Nt}}{\sqrt{N}} + \frac{\zeta^\gamma_{Nt}}{\sqrt{N}}\right) + o_p\left(\delta^{-1}_{NT}\right),$$

where $\xi^\gamma_{NT} = \frac{1}{N\sqrt{T}}\sum_i\sum_{s\neq t} x^*_{is}x^{*\prime}_{is}\gamma_s$, $\zeta^\gamma_{Nt} = \frac{1}{\sqrt{N}}\sum_i\left(x^*_{it}x^{*\prime}_{it} - \Sigma_x\right)\lambda_i$, and $\Psi^\gamma_{Nt} = \frac{1}{\sqrt{N}}\sum_i x^*_{it}u_{it}$

We then focus on the dominant terms below. Obviously, the first term at the right-hand side of the above equality does not involve taking the maximum over t, and based on Theorem 1.3.1, it is of order $O_p\left(T^{-1/2}\right)$, which will be the dominated term when N and T are comparable and diverge to infinity simultaneously. For second term at the right-hand side of the above equality, we have

$$\begin{aligned}\left(\frac{1}{N}\sum_i x_{it}^* x_{it}^{*\prime}\right)^{-1}\left(\frac{\Psi_{Nt}^{\gamma}}{\sqrt{N}}+\frac{\zeta_{Nt}^{\gamma}}{\sqrt{N}}\right) &= \left(\frac{1}{N}\sum_i x_{it}^* x_{it}^{*\prime}\right)^{-1}\left[\frac{1}{N}\sum_i \left(x_{it}^* x_{it}^{*\prime}-\Sigma_x\right)\lambda_i+\frac{1}{N}\sum_i x_{it}^* u_{it}\right] \\ &\equiv Q_{Nt}^{-1}\left(G_{1Nt}+G_{2Nt}\right).\end{aligned}$$

Then, it follows that

$$\begin{aligned}\max_t \|\widehat{\gamma}_t-\gamma_t\|_2 &= \max_i \left\|Q_{Nt}^{-1}\left(G_{1Nt}+G_{2Nt}\right)\right\|_2 \\ &\le \left(\max_i \left\|Q_{Nt}^{-1}\right\|_{op}\right)\left(\max_t \|G_{1Nt}\|_2+\max_t \|G_{2Nt}\|_2\right) \\ &\le \left(\left\|\Sigma_x^{-1}\right\|_{op}+\max_t \left\|Q_{Nt}^{-1}-\Sigma_x^{-1}\right\|_{op}\right)\left(\max_t \|G_{1Nt}\|_2+\max_t \|G_{2Nt}\|_2\right).\end{aligned}$$

Below, we are going to show: (i) $\max_t \left\|Q_{Nt}^{-1}-\Sigma_x^{-1}\right\|_{op}=o_p(1)$; (ii) $\max_t \left\|\frac{1}{N}\sum_i \left(x_{it}^* x_{it}^{*\prime}-\Sigma_x\right)\lambda_i\right\|_2 = O_p\left(\sqrt{\frac{\ln T}{N}}\right)$; (iii) $\max_t \left\|\frac{1}{N}\sum_i x_{it}^* u_{it}\right\|_2=O_p\left(\sqrt{\frac{\ln T}{N}}\right)$.

(i) For $Q_{Nt}=\frac{1}{N}\sum_i x_{it}^* x_{it}^{*\prime}$, because

$$\begin{aligned}\left\|Q_{Nt}^{-1}-\Sigma_x^{-1}\right\|_{op} &= \left\|Q_{Nt}^{-1}(\Sigma_x-Q_{Nt})\Sigma_x^{-1}\right\|_{op} \\ &\le \left\|Q_{Nt}^{-1}\right\|_{op}\|\Sigma_x-Q_{Nt}\|_{op}\left\|\Sigma_x^{-1}\right\|_{op}=\frac{\|\Sigma_x-Q_{Nt}\|_{op}}{s_{min}\left(Q_{Nt}\right)s_{min}\left(\Sigma_x\right)}.\end{aligned}$$

We then have

$$\begin{aligned}\max_t \left\|Q_{Nt}^{-1}-\Sigma_x^{-1}\right\|_{op} &\le \frac{\max_t \|\Sigma_x-Q_{Nt}\|_{op}}{\min_t s_{min}\left(Q_{Nt}\right)s_{min}\left(\Sigma_x\right)} \\ &\le \frac{\max_t \|\Sigma_x-Q_{Nt}\|_{op}}{s_{min}\left(\Sigma_x\right)\left(s_{\min}\left(\Sigma_x\right)-\max_t \|\Sigma_x-Q_{Nt}\|_{op}\right)}=O_p(1)\max_t \|\Sigma_x-Q_{Nt}\|_{op},\end{aligned}$$

where the last equality follows under Assumption A1.4(i) if $\max_t \|\Sigma_x-Q_{Nt}\|_F=o_p(1)$ also holds, which will be shown as below. Given the fact that

$$\max_t \|\Sigma_x - Q_{Nt}\|_{op} = \max_t \left\| \Sigma_x - \frac{1}{N}\sum_i x^*_{it}x^{*\prime}_{it} \right\|_{op}.$$

We now use the equivalent definition of operator norm for a symmetric $K \times K$ matrix A, $\|A\|_{op} = \max_{v\in S^{K-1},\|v\|=1} \|v'Av\|_2$. Denote $\widehat{\Sigma}_{t,x} = \frac{1}{N}\sum_i x^*_{it}x^{*\prime}_{it}$, then, $\left\|\widehat{\Sigma}_{t,x} - \Sigma_x\right\|_{op} \equiv \|\mathcal{Q}_t\|_{op} = \max_{v\in S^{k-1},\|v\|=1} |v'\mathcal{Q}_t v|$. By constructing a ϵ-net covering of S^{K-1}, we have $\left\|\widehat{\Sigma}_{t,x} - \Sigma_x\right\|_{op} \leq \frac{1}{1-2\epsilon}\max_{z\in N_\epsilon} |z'\mathcal{Q}_t z|$, where N_ϵ denote the set of points within ϵ-net covering. Particularly, let $\epsilon = \frac{1}{4}$, it follows that $|N_\epsilon| \leq 9^K = C_K < \infty$ and,

$$\begin{aligned} \left\|\widehat{\Sigma}_{t,x} - \Sigma_x\right\|_{op} &\leq 2\max_{z\in N_\epsilon} |z'\mathcal{Q}_t z| \\ &= 2\max_{s\in\{1,2,3,\ldots,C_K\}} |z'_s\mathcal{Q}_t z_s|, \end{aligned}$$

where $\{z_1, ..., z_{C_K}\}$ is a ϵ-net covering of S_{K-1}.Then, it suffices to bound the following quantity,

$$\begin{aligned} \mathbb{P}\left(\max_t \left\|\widehat{\Sigma}_{t,x} - \Sigma_x\right\|_{op} \geq \frac{2\sqrt{\ln T}}{\sqrt{N}}\right) &\leq \mathbb{P}\left(\max_t \max_{s\in\{1,2,3,\ldots,C_k\}} |z'_s\mathcal{Q}_t z_s| \geq \frac{\sqrt{\ln T}}{\sqrt{T}}\right) \\ &\leq \sum_{s=1}^{C_K} \mathbb{P}\left(\max_t |z'_s\mathcal{Q}_t z_s| \geq \sqrt{\frac{\ln T}{N}}\right). \end{aligned}$$

Now, we are in the position to bound the probabilities $\mathbb{P}\left(\max_t |z'_s\mathcal{Q}_t z_s| \geq \sqrt{\frac{\ln T}{N}}\right)$. For any fixed $z_s \in S^{K-1}$,

$$z'_s\mathcal{Q}_t z_s = z'_s\left(\widehat{\Sigma}_{t,x} - \Sigma_x\right)z_s = \frac{1}{N}\sum_{i=1}^N \left[(z'_s x^*_{it})^2 - z'_s\Sigma_x z_s\right] = \frac{1}{N}\sum_{i=1}^N \left(W^2_{it} - EW^2_{it}\right),$$

where $W_{it} = z'_s x^*_{it}$. Thus, it is equivalent to bound $\mathbb{P}\left(\max_t \left|\frac{1}{N}\sum_{i=1}^N \left(W^2_{it} - EW^2_{it}\right)\right| \geq \sqrt{\frac{\ln T}{N}}\right)$.

Similar to arguments stated in Zhu (2017), because entries of $\{x_{it}\}$ have exponential-type tails as assumed in Assumption A1.8, it follows, by Lemma C.3(1) in Zhu (2017), that entries of $\{W_{it}\}$ also have exponential-type tails with parameters that only depend on the constants in Assumption A1.8, and further, by Lemma C.3(3)-(4) in Zhu (2017), $\{W_{it}^2 - EW_{it}^2\}$ also have exponential-type tails with parameter $\left(b^\star, \frac{\nu}{2}\right)$ with $b^\star \equiv 2^{\frac{1}{\nu}} (b+\mu)^2 (\ln(\|z\|_0 + 2))^{\frac{2}{\nu}} + m^*$ where $\mu \equiv Ex_{it}$, $\|z\|_0$ represents the total number of non-zero elements in z, $m^* \equiv E\left(z' x_{it}^*\right)^2$, and (b, ν) stated in Assumption A1.8. We now use Theorem 1 in Merlevède et al. (2011) under Assumption A1.8, it follows that, let a_T be the bound to be determined, and $\gamma \equiv \nu/2$, then, there exist positive constants $C1$, $C2$, $C3$, $C4$ and C_5 depending only on $b^\star$, $\nu \in (0,1)$,

$$
\begin{aligned}
& P\left(\max_t \left|\sum_{i=1}^N \left(W_{it}^2 - EW_{it}^2\right)\right| > a_T x\right) \leq \sum_{t=1}^T \mathbb{P}\left(\left|\sum_{i=1}^N \left(W_{it}^2 - EW_{it}^2\right)\right| > a_T x\right) \\
\leq \; & NT \exp\left(-C_1 a_T^\gamma x^\gamma\right) + T \exp\left(-\frac{C_2 a_T^2 x^2}{1 + C_3 N}\right) \\
& + T \exp\left[-\frac{C_4 a_T^2 x^2}{N} \exp\left(C_5 \left(a_T x\right)^{\gamma/(1-\gamma)} \left(\log a_T x\right)^{-\gamma}\right)\right].
\end{aligned}
$$

Thus, choose $a_T = d_* \sqrt{N \ln T}$, by elementary computations, we can choose large constants $a_*, b^*, d_* > 0$ such that $\forall x \geq a_*$

$$
\begin{aligned}
\mathbb{P}\left(\left(d_* \sqrt{N \ln T}\right)^{-1} \max_t \left|\sum_{i=1}^N \left(W_{it}^2 - EW_{it}^2\right)\right| > x\right) &= \mathbb{P}\left(\max_t \left|\sum_{i=1}^N \left(W_{it}^2 - EW_{it}^2\right)\right| > a_n x\right) \\
&\leq b^* \exp\left(-x^\gamma\right).
\end{aligned}
$$

It concludes that $\max_t \left\|\Sigma_x - \frac{1}{N}\sum_i x_{it}^* x_{it}^{*\prime}\right\|_2 = O_p\left(\sqrt{\frac{\ln T}{N}}\right)$. And thus, it follows that $\left\|\widehat{\Sigma}_{t,x} - \Sigma_x\right\|_{op} = O_p\left(\sqrt{\frac{\ln T}{N}}\right)$, which implies that $\max_t \left\|Q_{Nt}^{-1} - \Sigma_x^{-1}\right\|_{op} = O_p\left(\sqrt{\frac{\ln T}{N}}\right) = o_p(1)$.

(ii) For G_{1Nt}, we have $G_{1Nt} = \frac{1}{N}\sum_i \left(x_{it}^* x_{it}^{*\prime} - \Sigma_x\right)\lambda_i \equiv \frac{1}{N}\sum_i G_{1it}$, note that $\{x_{it}, \lambda_i\}$ have exponential-type tails as assumed in Assumption A1.8, then, by using Lemma C.3 in Zhu (2017) again, $\{G_{1it}\}_{i\in\{1,\ldots,N\}}$ also have have exponential-type tails by following those arguments of (i) above, and then, in the view of proofs for (i), we can conclude that

$$
\max_t \left\|\frac{1}{N}\sum_i \left(x_{it}^* x_{it}^{*\prime} - \Sigma_x\right)\lambda_i\right\|_2 = O_p\left(\sqrt{\frac{\ln T}{N}}\right).
$$

(iii) For G_{2Nt}, we have $G_{2Nt} = \frac{1}{N}\sum_i x_{it}^* u_{it} \equiv \frac{1}{N}\sum_i G_{2it}$, note that $\{x_{it}, u_{it}\}$ have exponential-type tails as assumed in Assumption A1.8, then, by using Lemma C.3 in Zhu (2017) again, $\{G_{2it}\}_{i\in\{1,\ldots,N\}}$ also have have exponential-type tails by following arguments for (i) and (ii) just above, and then, in the view of proofs for (i) and (ii), we can conclude that

$$\max_t \left\| \frac{1}{N}\sum_i x_{it}^* u_{it} \right\|_2 = O_p\left(\sqrt{\frac{\ln T}{N}}\right).$$

Then, Collecting the results shown in (i), (ii) and (iii) above, we then conclude that

$$\max_t \|\widehat{\gamma}_t - \gamma_t\|_\infty = O_p\left(\sqrt{\frac{\ln T}{N}}\right).$$

Uniform Consistency of $\widehat{\lambda}_i$

It can be seen that details of proofs for the uniform consistency of $\widehat{\lambda}_i$ are very similar to those shown in the last subsection. For the completeness, we give the main steps below and omit some repeated arguments for brevity.

Note that $\|\widehat{\lambda}-\lambda\|_\infty = \max_i \|\widehat{\lambda}_i - \lambda_i\|_\infty \le \max_i \|\widehat{\lambda}_i - \lambda_i\|_2$, then it suffices to bound $\max_i \|\widehat{\lambda}_i - \lambda_i\|_2$ instead. Neglecting those terms that are of smaller orders, we have

$$\widehat{\lambda}_i - \lambda_i \;\; = \;\; -\left(\frac{1}{NT}\sum_i\sum_s x_{is}^* x_{is}^{*\prime}\right)^{-1} \frac{\xi_{NT}^{\lambda}}{\sqrt{N}} + \left(\frac{1}{T}\sum_t x_{it}^* x_{it}^{*\prime}\right)^{-1}\left(\frac{\Psi_{iT}^{\gamma}}{\sqrt{T}} + \frac{\zeta_{iT}^{\gamma}}{\sqrt{T}}\right) + o_p\left(\delta_{NT}^{-1}\right),$$

where $\xi_{NT}^{\gamma} = \frac{1}{\sqrt{NT}}\sum_{j\neq i}\sum_t x_{jt}^* x_{jt}^{*\prime}\lambda_j$, $\zeta_{iT}^{\gamma} = \frac{1}{\sqrt{T}}\sum_t \left(x_{it}^* x_{it}^{*\prime} - \Sigma_x\right)\gamma_t$, and $\Psi_{iT}^{\gamma} = \frac{1}{\sqrt{T}}\sum_t x_{it}^* u_{it}$.

We then focus on the dominant terms below. Obviously, the first term at the right-hand side of the above equality does not involve taking the maximum over t, and based on Theorem 1.3.1, it is of order $O_p\left(N^{-1/2}\right)$, which will be the dominated term when N and T are comparable and diverge to infinity simultaneously. For second term at the right-hand side of the above equality, we have

$$\begin{aligned}\left(\frac{1}{T}\sum_t x_{it}^* x_{it}^{*\prime}\right)^{-1}\left(\frac{\Psi_{iT}^{\gamma}}{\sqrt{T}}+\frac{\zeta_{iT}^{\gamma}}{\sqrt{T}}\right) &= \left(\frac{1}{T}\sum_t x_{it}^* x_{it}^{*\prime}\right)^{-1}\left(\frac{1}{T}\sum_t \left(x_{it}^* x_{it}^{*\prime}-\Sigma_x\right)\gamma_t+\frac{1}{T}\sum_t x_{it}^* u_{it}\right)\\ &\equiv Q_{iT}^{-1}\left(G_{1iT}+G_{2iT}\right).\end{aligned}$$

Then, we have

$$\begin{aligned}\max_i\|\widehat{\lambda}_i-\lambda_i\|_2 &= \max_i\left\|Q_{iT}^{-1}\left(G_{1iT}+G_{2iT}\right)\right\|_2\\ &\le \left(\max_i\left\|Q_{iT}^{-1}\right\|_{op}\right)\left(\max_i\|G_{1iT}\|_2+\max_i\|G_{2iT}\|_2\right)\\ &\le \left(\left\|\Sigma_x^{-1}\right\|_{op}+\max_i\left\|Q_{iT}^{-1}-\Sigma_x^{-1}\right\|_{op}\right)\left(\max_i\|G_{1iT}\|_2+\max_i\|G_{2iT}\|_2\right).\end{aligned}$$

Our proofs below aim to show: (i) $\max_i\left\|Q_{iT}^{-1}-\Sigma_x^{-1}\right\|_{op}=O_p\left(\sqrt{\frac{\ln N}{T}}\right)$; (ii) $\max_i\|G_{1iT}\|_2=O_p\left(\sqrt{\frac{\ln N}{T}}\right)$; (iii) $\max_i\|G_{2iT}\|_2=O_p\left(\sqrt{\frac{\ln N}{T}}\right)$.

(i) For $Q_{iT}=\frac{1}{T}\sum_t x_{it}^* x_{it}^{*\prime}$, because

$$\begin{aligned}\left\|Q_{iT}^{-1}-\Sigma_x^{-1}\right\|_{op} &= \left\|Q_{iT}^{-1}(\Sigma_x-Q_{iT})\Sigma_x^{-1}\right\|_{op}\\ &\le \left\|Q_{iT}^{-1}\right\|_{op}\|\Sigma_x-Q_{iT}\|_{op}\left\|\Sigma_x^{-1}\right\|_{op}=\frac{\|\Sigma_x-Q_{iT}\|_{op}}{s_{min}\left(Q_{iT}\right)s_{min}\left(\Sigma_x\right)}.\end{aligned}$$

It follows that

$$\begin{aligned}\max_i\left\|Q_{iT}^{-1}-\Sigma_x^{-1}\right\|_{op} &\le \frac{\max_i\|\Sigma_x-Q_{iT}\|_{op}}{\min_i s_{min}\left(Q_{iT}\right)s_{min}\left(\Sigma_x\right)}\\ &\le \frac{\max_i\|\Sigma_x-Q_{iT}\|_{op}}{s_{min}\left(\Sigma_x\right)\left(s_{min}\left(\Sigma_x\right)-\max_i\|\Sigma_x-Q_{iT}\|_{op}\right)}=O_p(1)\max_i\|\Sigma_x-Q_{iT}\|_{op},\end{aligned}$$

where the last equality follows under Assumption A1.4(i) if $\max_i\|\Sigma_x-Q_{iT}\|_F=o_p(1)$ also holds, which will be shown as below. Given the fact that

$$\max_i \|\Sigma_x - Q_{iT}\|_{op} \;=\; \max_i \left\|\Sigma_x - \frac{1}{T}\sum_t x_{it}^* x_{it}^{*\prime}\right\|_{op} \leq \max_i \left\|\Sigma_x - \frac{1}{T}\sum_t x_{it}^* x_{it}^{*\prime}\right\|_{op}.$$

Use the equivalent definition of operator norm for a symmetric $K \times K$ matrix A again. Denote $\widehat{\Sigma}_X = \frac{1}{T}\sum_t x_{it}^* x_{it}^{*\prime}$, then, $\left\|\widehat{\Sigma}_X - \Sigma_x\right\|_{op} \equiv \|Q_i\|_{op} = \max_{v\in S^{k-1}, \|v\|=1} |v' Q_i v|$. By constructing a ϵ-net covering of S^{K-1}, we have $\left\|\widehat{\Sigma}_X - \Sigma_x\right\|_{op} \leq \frac{1}{1-2\epsilon}\max_{z\in N_\epsilon} |z' Q_i z|$, where N_ϵ denote the set of points within ϵ-net covering. Particularly, let $\epsilon = \frac{1}{4}$, it follows that $|N_\epsilon| \leq 9^K = C_K < \infty$ and,

$$\left\|\widehat{\Sigma}_X - \Sigma_x\right\|_{op} \leq 2\max_{z\in N_\epsilon} |z' Q_i z| = 2 \max_{s\in\{1,2,3,\ldots,C_K\}} |z_s' Q_i z_s|,$$

where $\{z_1, ..., z_{C_K}\}$ is a ϵ-net covering of S_{K-1}. Then, it suffices to bound the following quantity,

$$\begin{aligned}
\mathbb{P}\left(\max_i \left\|\widehat{\Sigma}_X - \Sigma_x\right\|_{op} \geq \frac{2\sqrt{\ln N}}{\sqrt{T}}\right) &\leq \mathbb{P}\left(\max_i \max_{s\in\{1,2,3,\ldots,C_k\}} |z_s' Q_i z_s| \geq \frac{2\sqrt{\ln N}}{\sqrt{T}}\right) \\
&\leq \sum_{s=1}^{C_k} \mathbb{P}\left(\max_i |z_s' Q_i z_s| \geq \frac{2\sqrt{\ln N}}{\sqrt{T}}\right).
\end{aligned}$$

Now, we are in the position to bound the probabilities $\mathbb{P}\left(\max_i |z_s' Q_i z_s| \geq \frac{\ln N}{\sqrt{T}}\right)$. For any fixed $z_s \in S^{K-1}$,

$$z_s' Q_i z_s \;= z_s'\left(\widehat{\Sigma}_X - \Sigma_x\right) z_s = \tfrac{1}{T}\textstyle\sum_{t=1}^T \left[(z_s' x_{it}^*)^2 - z_s' \Sigma_x z_s\right] = \tfrac{1}{T}\textstyle\sum_{t=1}^T \left(W_{it}^2 - EW_{it}^2\right),$$

where $W_{it} = z_s' x_{it}^*$. Thus, it is equivalent to bound $\mathbb{P}\left(\max_i \left|\frac{1}{T}\sum_{t=1}^T \left(W_{it}^2 - EW_{it}^2\right)\right| \geq \sqrt{\frac{\ln N}{T}}\right)$.

Following the arguments for (i) in the last subsection, we can readily show

$$\max_i \left\|\Sigma_x - \frac{1}{T}\sum_t x_{it}^* x_{it}^{*\prime}\right\|_2 = O_p\left(\sqrt{\frac{\ln N}{T}}\right).$$

Collecting all above immediate results, we conclude that $\max_i \left\|Q_{iT}^{-1} - \Sigma_x^{-1}\right\|_2 = O_p\left(\sqrt{\frac{\ln N}{T}}\right)$.

(ii) For G_{1iT}, we have

$$\begin{aligned}\max_i \|G_{1iT}\|_2 &\leq \max_i \left\| \frac{1}{T}\sum_t \left(x_{it}^* x_{it}^{*\prime} - \Sigma_x\right)\gamma_t \right\|_2 \\ &= O_p\left(\sqrt{\frac{\ln N}{T}}\right),\end{aligned}$$

where the second line follows by mimicking those arguments in (i) in the last section under Assumptions A1.1 and A1.8.

(iii) For $G_{2iT} = \frac{1}{T}\sum_t x_{it}^* u_{it}$, by mimicking those arguments in (i) in the last section under Assumptions A1.1, A1.2 and A1.8, it is straightforward to have $\max_i \|G_{2iT}\|_2 = O_p\left(\sqrt{\frac{\ln N}{T}}\right)$.

Collecting immediate results from (i) to (iii) above, it follows $\max_i \|\widehat{\lambda}_i - \lambda_i\|_2 = O_p\left(\sqrt{\frac{\ln N}{T}}\right)$ directly.

A.1.6 Proofs of Theorem 1.4.1

In a nutshell, our proof strategy follows arguments of proofs for Theorem 3 in Castagnetti et al. (2015). We will give main steps and arguments, and refer readers to Castagnetti et al. (2015) for omitted details.

***Part I*. The asymptotic null distribution of $\mathcal{T}_\lambda$**

Note that $\lambda_i = 0$ for all i under H_{0a}, we can write $\widehat{\lambda}_i = \widehat{\lambda}_i - \lambda_i$. Then, according to the decomposition in proofs for Theorem 1.3.3, we have

$$\widehat{\lambda}_i - \lambda_i = -\Sigma_x^{-1}(\xi_{NT}^\lambda/\sqrt{N}) + \Sigma_x^{-1}(\zeta_{iT}^\lambda/\sqrt{T}) + \Sigma_x^{-1}(\Psi_{iT}^\lambda/\sqrt{T}) + o_p\left(\delta_{NT}^{-1}\right).$$

And for $\widehat{\mathcal{V}}_{\lambda i}$, the consistent estimator of $\mathcal{V}_{\lambda i}$, we can write

$$\widehat{\mathcal{V}}_{\lambda i}^{-1} = \mathcal{V}_{\lambda i}^{-1} - \mathcal{V}_{\lambda i}^{-1}\left(\widehat{\mathcal{V}}_{\lambda i} - \mathcal{V}_{\lambda i}\right)\widehat{\mathcal{V}}_{\lambda i}^{-1} = \mathcal{V}_{\lambda i}^{-1} - \mathcal{V}_{\lambda i}^{-1}\left(\widehat{\mathcal{V}}_{\lambda i} - \mathcal{V}_{\lambda i}\right)\mathcal{V}_{\lambda i}^{-1} + o_p\left(\left\|\widehat{\mathcal{V}}_{\lambda i} - \mathcal{V}_{\lambda i}\right\|\right).$$

Neglecting the terms that is $o_p\left(\left\|\widehat{\mathcal{V}}_{\lambda i}-\mathcal{V}_{\lambda i}\right\|\right)$ further, under H_{0a}, we have

$$\begin{aligned}\left(\widehat{\lambda}_i-\lambda_i\right)'\widehat{\mathcal{V}}_{\lambda i}^{-1}\left(\widehat{\lambda}_i-\lambda_i\right) &= \left(\widehat{\lambda}_i-\lambda_i\right)'\mathcal{V}_{\lambda i}^{-1}\left(\widehat{\lambda}_i-\lambda_i\right)+\left(\widehat{\lambda}_i-\lambda_i\right)'\mathcal{V}_{\lambda i}^{-1}\left(\widehat{\mathcal{V}}_{\lambda i}-\mathcal{V}_{\lambda i}\right)\mathcal{V}_{\lambda i}^{-1}\left(\widehat{\lambda}_i-\lambda_i\right)\\ &\equiv \left(\widehat{\lambda}_i-\lambda_i\right)'\mathcal{V}_{\lambda i}^{-1}\left(\widehat{\lambda}_i-\lambda_i\right)+I_{\lambda i}.\end{aligned}$$

Now, we are in the position to show the desired results below by following four steps.

Step 1. It is straightforward to see the fact $\max_{1\le i\le N}\left(\widehat{\lambda}_i-\lambda_i\right)'\mathcal{V}_{\lambda i}^{-1}\left(\widehat{\lambda}_i-\lambda_i\right)=O_p(\ln N)$ by Theorem 1.3.3 and corresponding arguments of proofs for Theorem 3 in Castagnetti et al. (2015).

Step 2. we are going to show $\max_{1\le i\le N} I_{\lambda i}=o_p\left(\ln N\right)$ as below. Given that

$$\max_{1\le i\le N} I_{\lambda i}\le\max_{1\le i\le N}\left\|\left(\widehat{\lambda}_i-\lambda_i\right)'\mathcal{V}_{\lambda i}^{-1/2}\right\|_2^2\max_{1\le i\le N}\left\|\mathcal{V}_{\lambda i}^{-1/2}\left(\widehat{\mathcal{V}}_{\lambda i}-\mathcal{V}_{\lambda i}\right)\mathcal{V}_{\lambda i}^{-1/2}\right\|_F.$$

It is straightforward to see that the first term at the right-hand side of the above inequality is of order $O_p(\ln N)$ as shown in *Step 1* above. So it suffices to show $\max_{1\le i\le N}\left\|\mathcal{V}_{\lambda i}^{-1/2}\left(\widehat{\mathcal{V}}_{\lambda i}-\mathcal{V}_{\lambda i}\right)\mathcal{V}_{\lambda i}^{-1/2}\right\|_2=o_p\left(1\right)$.

Given that

$$\max_{1\le i\le N}\left\|\mathcal{V}_{\lambda i}^{-1/2}\left(\widehat{\mathcal{V}}_{\lambda i}-\mathcal{V}_{\lambda i}\right)\mathcal{V}_{\lambda i}^{-1/2}\right\|_F \le \max_{1\le i\le N}\left\|\mathcal{V}_{\lambda i}^{-1}\right\|_F\max_{1\le i\le N}\left\|\widehat{\mathcal{V}}_{\lambda i}-\mathcal{V}_{\lambda i}\right\|_F.$$

Obviously, for ant consistent estimator of $\mathcal{V}_{\lambda i}$, we have $\widehat{\mathcal{V}}_{\lambda i}=\mathcal{V}_{\lambda i}+o_p\left(1\right)$ for each i, and the fact that $\mathcal{V}_{\lambda i}$ exists and is non-singular according to proofs for Theorem 1.3.3 under Assumptions imposed for each i. By following arguments of proofs for Proposition 1.3.8, $\max_{1\le i\le N}\left\|\mathcal{V}_{\lambda i}^{-1}\right\|_F=O_p\left(1\right)$. holds uniformly over i immediately by simple calculations as long as $\max_{1\le i\le N}\left\|\widehat{\mathcal{V}}_{\lambda i}-\mathcal{V}_{\lambda i}\right\|_F=o_p\left(1\right)$ holds.

If we use the consistent estimate of $\mathcal{V}_{\lambda i}$ defined in Corollary 1.3.5, then $\widehat{\mathcal{V}}_{\lambda i}=\widehat{Q}_{\tilde{x}}^{-1}\left(\widehat{V}_{\lambda}+\widehat{\Omega}_{\lambda i}\right)\widehat{Q}_{\tilde{x}}^{-1}$, $\widehat{V}_{\lambda}$ is an consistent estimator of covariance matrix of V_{λ} defined in Theorem 1.3.2, $\widehat{Q}_{\tilde{x}}=\frac{1}{T}\sum_t\tilde{x}_{it}\tilde{x}_{it}'$, and

$$\widehat{\Omega}_{\lambda i}=\frac{1}{T}\widehat{\Sigma}_{i,x\gamma}+\frac{1}{T}\widehat{\widehat{\Sigma}}_{i,xu},$$

Since V_λ has nothing to do with taking maximum over i by construction in Corollary 1.3.2, to show $\max_{1\leq i\leq N}\left\|\widehat{\mathcal{V}}_{\lambda i}-\mathcal{V}_{\lambda i}\right\|_F=o_p(1)$ holds, it is sufficient to show that $\max_{1\leq i\leq N}\left\|\widehat{\Omega}_{\lambda i}-\Omega_{\lambda i}\right\|_F=o_p(1)$ holds. Then, by triangular inequality, it suffices to show $\max_{1\leq i\leq N}\left\|\widehat{\Sigma}_{i,x\gamma}-\Sigma_{i,x\gamma}\right\|_F=o_p(1)$ and $\max_{1\leq i\leq N}\left\|\widehat{\tilde{\Sigma}}_{i,xu}-\tilde{\Sigma}_{i,xu}\right\|=o_p(1)$ separately. By direct calculations, and similar arguments as those in proofs for Proposition 1.3.8, the desired result follows. We now give the sketch of proofs for $\max_{1\leq i\leq N}\left\|\widehat{\Sigma}_{i,x\gamma}-\Sigma_{i,x\gamma}\right\|_F=o_p(1)$ below as the illustrative purpose.

Because $\widehat{\Sigma}_{i,x\gamma}$ is the HAC estimator of $\Sigma_{i,x\gamma}$, we can show that $\max_{1\leq i\leq N}\left\|\widehat{\Sigma}_{i,x\gamma}-\widehat{\Sigma}^{\star}_{i,x\gamma}\right\|_F=o_p(1)$ with $\widehat{\Sigma}^{\star}_{i,x\gamma}\equiv\frac{1}{T}\sum_{t,s}\left(x^*_{jt}x^{*\prime}_{it}-\Sigma_x\right)\gamma_t\gamma_s'\left(x^*_{is}x^{*\prime}_{is}-\Sigma_x\right)$ under mild conditions. Further, by straightforward calculations, we have

$$
\begin{aligned}
\widehat{\Sigma}_{i,x\gamma}-\Sigma^{\star}_{i,x\gamma} &= \frac{1}{T}\sum_{t,s}\left(x^*_{jt}x^{*\prime}_{it}-\Sigma_x\right)\gamma_t\left(\widehat{\gamma}_s-\gamma_s\right)'\left(x^*_{is}x^{*\prime}_{is}-\Sigma_x\right)\\
&\quad+\frac{1}{T}\sum_{t,s}\left(x^*_{it}x^{*\prime}_{it}-\Sigma_x\right)\left(\widehat{\gamma}_t-\gamma_t\right)\gamma_s'\left(x^*_{is}x^{*\prime}_{is}-\Sigma_x\right)\\
&\quad+\frac{1}{T}\sum_{t,s}\left(x^*_{it}x^{*\prime}_{it}-\Sigma_x\right)\left(\widehat{\gamma}_t-\gamma_t\right)\left(\widehat{\gamma}_s-\gamma_s\right)'\left(x^*_{is}x^{*\prime}_{is}-\Sigma_x\right)+\text{dominated terms}\\
&\equiv \mathcal{R}_{1i}+\mathcal{R}_{2i}+\mathcal{R}_{3i}+\text{dominated terms}.
\end{aligned}
$$

It suffices to bound $\max_{1\leq i\leq N}\|\mathcal{R}_{1i}\|_F$, $\max_{1\leq i\leq N}\|\mathcal{R}_{2i}\|_F$, and $\max_{1\leq i\leq N}\|\mathcal{R}_{3i}\|_F$ in turns. For $\|\mathcal{R}_{1i}\|_F$, we have

$$
\|\mathcal{R}_{1i}\|_F \leq \left\|\frac{1}{T}\sum_{t,s}\left(x^*_{it}x^{*\prime}_{it}-\Sigma_x\right)\gamma_t\iota_K\left(x^*_{is}x^{*\prime}_{is}-\Sigma_x\right)\right\|_{op}\max_{1\leq t\leq T}\|\widehat{\gamma}_t-\gamma_t\|_2,
$$

where ι_K stands for K-dimensional vector with 1 as the entry for all elements. It follows that

$$
\begin{aligned}
\max_{1\leq i\leq N}\|\mathcal{R}_{1i}\|_2 &\leq \max_{1\leq i\leq N}\left\|\frac{1}{T}\sum_{t,s}\left(x^*_{it}x^{*\prime}_{it}-\Sigma_x\right)\gamma_t\iota_K\left(x^*_{is}x^{*\prime}_{is}-\Sigma_x\right)\right\|_{op}\max_{1\leq t\leq T}\|\widehat{\gamma}_t-\gamma_t\|_2\\
&\equiv \max_{1\leq i\leq N}\left\|\check{R}_{1i}\right\|_{op}\max_{1\leq t\leq T}\|\widehat{\gamma}_t-\gamma_t\|_2
\end{aligned}
$$

Then, under Assumption A1.1, by similar arguments for the proofs of Proposition 1.3.8, we can readily show that $\max_{1\leq i\leq N}\left\|\check{R}_{1i}\right\|_{op}=O_p\left(\sqrt{\ln N}\right)$, and by Proposition 1.3.8, we then have $\max_{1\leq t\leq T}\|\widehat{\gamma}_t-\gamma_t\|_2=$

$O_p\left(\sqrt{N^{-1}\ln T}\right)$, as (N,T) go to infinity jointly, we have $\max_{1\le i\le N}\|\mathcal{R}_{1i}\| = O_p\left(\sqrt{N^{-1}\ln T\ln N}\right) = o_p(1)$.

By similar arguments for $\max_{1\le i\le N}\|\mathcal{R}_{1i}\|$, we can also show that $\max_{1\le i\le N}\|\mathcal{R}_{2i}\|$ and $\max_{1\le i\le N}\|\mathcal{R}_{3i}\|$ are both of order $o_p(1)$. Collecting these immediate results together, $\max_{1\le i\le N}\left\|\widehat{\Sigma}_{i,x\gamma} - \Sigma_{i,x\gamma}\right\|_F = o_p(1)$ holds.

Obviously, $\max_{1\le i\le N}\left\|\widehat{\tilde{\Sigma}}_{i,xu} - \tilde{\Sigma}_{i,xu}\right\| = o_p(1)$ can be justified by mimicking above arguments with necessary modifications.

Step 3. Now, we consider the covariance between $\sqrt{T}\left(\widehat{\lambda}_i - \lambda_i\right)$ and $\sqrt{T}\left(\widehat{\lambda}_j - \lambda_j\right)$. Recall that

$$\widehat{\lambda}_i - \lambda_i = -\Sigma_x^{-1}(\xi_{NT}^{\lambda}/\sqrt{N}) + \Sigma_x^{-1}(\zeta_{iT}^{\lambda}/\sqrt{T}) + \Sigma_x^{-1}(\Psi_{iT}^{\lambda}/\sqrt{T}) + o_p\left(\delta_{NT}^{-1}\right)$$

where $\xi_{NT}^{\lambda} = \frac{1}{\sqrt{NT}}\sum_{j\neq i}\sum_t x_{jt}^* x_{jt}^{*\prime}\lambda_j$, $\zeta_{iT}^{\lambda} = \frac{1}{\sqrt{T}}\sum_t\left(x_{it}^* x_{it}^{*\prime} - \Sigma_x\right)\gamma_t$, and $\Psi_{iT}^{\lambda} = \frac{1}{\sqrt{T}}\sum_t x_{it}^* u_{it}$

Under H_{0a} and Assumption A1.9, the above decomposition simplifies to the below:

$$\sqrt{T}\left(\widehat{\lambda}_i - \lambda_i\right) = \Sigma_x^{-1}\frac{1}{\sqrt{T}}\sum_t x_{it}^* u_{it} + o_p(1) \equiv \zeta_{iT}^{\lambda} + o_p(1)$$

Neglecting those smaller-order terms, the covariance between ζ_{iT}^{λ} and ζ_{jT}^{λ} is then proportion to the quantity, $E\left(\frac{1}{T}\sum_{t,s} x_{it}^* u_{it} u_{js} x_{is}^{*\prime}\right)$, and

$$E\left(\frac{1}{T}\sum_{t,s} x_{it}^* u_{it} u_{js} x_{is}^{*\prime}\right)\ln N = \frac{1}{T}\sum_{t,s} E\left(x_{it}^* u_{it} u_{js} x_{js}^{*\prime}\right)\ln N \to 0,$$

as (N,T) go to infinity jointly under Assumption A1.8. Thus, $Cov\left(\zeta_{iT}^{\lambda}, \zeta_{jT}^{\lambda}\right) = 0$ as (N,T) goes to infinity for all $i \neq j$. It concludes that ζ_{iT}^{λ} and ζ_{jT}^{λ} are asymptotic independent for $i \neq j$ because of asymptotic normality of ζ_{iT}^{λ} and ζ_{jT}^{λ}, and so $\sqrt{T}\left(\widehat{\lambda}_i - \lambda_i\right)$ and $\sqrt{T}\left(\widehat{\lambda}_j - \lambda_j\right)$ are. Meanwhile, the above result also indicates that the Berman's condition holds for $\{\sqrt{T}\left(\widehat{\lambda}_i - \lambda_i\right)\}_{i=1}^N$.

Step 4. By following corresponding arguments of proofs for Theorem 3 in Castagnetti et al. (2015), the asymptotic distribution of $\mathcal{T}_{\lambda}$ is the Gumbel distribution after being rescaled appropriately as shown in Theorem 1.4.1.

***Part II*. The asymptotic local power properties of $\mathcal{T}_\lambda$**

Now, we turn to analyze the local power of $\mathcal{T}_\lambda$ under local alternatives $H_{1a,NT}$. Note that under $H_{1a,NT}$, we have $\widehat{\lambda}_i = \widehat{\lambda}_i - \lambda_i + \lambda_i$. Similar to *Part I* above, neglecting terms that is $o_p\left(\left\|\widehat{\mathcal{V}}_{\lambda i} - \mathcal{V}_{\lambda i}\right\|\right)$, under $H_{1a,NT}$, we then have

$$
\begin{aligned}
& \left(\widehat{\lambda}_i - \lambda_i + \lambda_i\right)' \widehat{\mathcal{V}}_{\lambda i}^{-1} \left(\widehat{\lambda}_i - \lambda_i + \lambda_i\right) \\
= & \left(\widehat{\lambda}_i - \lambda_i + \lambda_i\right)' \mathcal{V}_{\lambda i}^{-1} \left(\widehat{\lambda}_i - \lambda_i + \lambda_i\right) + \left(\widehat{\lambda}_i - \lambda_i + \lambda_i\right)' \mathcal{V}_{\lambda i}^{-1} \left(\widehat{\mathcal{V}}_{\lambda i} - \mathcal{V}_{\lambda i}\right) \mathcal{V}_{\lambda i}^{-1} \left(\widehat{\lambda}_i - \lambda_i + \lambda_i\right) \\
= & \left(\widehat{\lambda}_i - \lambda_i\right)' \mathcal{V}_{\lambda i}^{-1} \left(\widehat{\lambda}_i - \lambda_i\right) + \lambda_i' \mathcal{V}_{\lambda i}^{-1} \lambda_i + 2\left(\widehat{\lambda}_i - \lambda_i\right)' \mathcal{V}_{\lambda i}^{-1} \lambda_i \\
& + \left(\widehat{\lambda}_i - \lambda_i\right)' \mathcal{V}_{\lambda i}^{-1} \left(\widehat{\mathcal{V}}_{\lambda i} - \mathcal{V}_{\lambda i}\right) \mathcal{V}_{\lambda i}^{-1} \left(\widehat{\lambda}_i - \lambda_i\right)' \\
& + \lambda_i' \mathcal{V}_{\lambda i}^{-1} \left(\widehat{\mathcal{V}}_{\lambda i} - \mathcal{V}_{\lambda i}\right) \mathcal{V}_{\lambda i}^{-1} \lambda_i + 2\left(\widehat{\lambda}_i - \lambda_i\right)' \mathcal{V}_{\lambda i}^{-1} \left(\widehat{\mathcal{V}}_{\lambda i} - \mathcal{V}_{\lambda i}\right) \mathcal{V}_{\lambda i}^{-1} \lambda_i \\
\equiv & \left(\widehat{\lambda}_i - \lambda_i\right)' \mathcal{V}_{\lambda i}^{-1} \left(\widehat{\lambda}_i - \lambda_i\right) + I_1^\lambda + I_2^\lambda + 2I_3^\lambda + I_4^\lambda + 2I_5^\lambda,
\end{aligned}
$$

***Step 1*.** It is straightforward to see the fact $\max_{1\le i\le N} \left(\widehat{\lambda}_i - \lambda_i\right)' \mathcal{V}_{\lambda i}^{-1} \left(\widehat{\lambda}_i - \lambda_i\right) = O_p(\ln N)$ by Theorem 1.3.3 and corresponding arguments of proofs for Theorem 3 in Castagnetti et al. (2015).

***Step 2*.** In this step, we figure out the order of $\max_{1\le i\le N} \left\|I_1^\lambda\right\|_2$,..., $\max_{1\le i\le N} \left\|I_5^\lambda\right\|_2$ in turns below.

(1) $\max_{1\le i\le N} \left\|I_1^\lambda\right\|_2 \le \max_{1\le i\le N} \|\lambda_i\|_2^2 \left\|\mathcal{V}_{\lambda i}^{-1}\right\|_{op} \le \max_{1\le i\le N} T \|\lambda_i\|_2^2 = O\left(T\|\lambda\|_\infty^2\right)$, where $\|\lambda\|_\infty^2 \equiv \max_{1\le i\le N}\{\|\lambda_1\|_2^2, .., \|\lambda_N\|_2^2\}$, and $\left\|\left(T^{-1}\mathcal{V}_{\lambda i}\right)^{-1}\right\|_{op} = O(1)$ holds uniformly over i by construction of $\mathcal{V}_{\lambda i}$ stated in Theorem 1.3.3 and the corresponding proofs for Theorem 1.3.3 in the Appendix A.

(2) By the immediate results in *Step 1* and (1) just above,

$$
\begin{aligned}
\max_{1\le i\le N} \left\|I_2^\lambda\right\|_2 & \le \left(\max_{1\le i\le N} \left\|\mathcal{V}_{\lambda i}^{-1/2}\left(\widehat{\lambda}_i - \lambda_i\right)\right\|_2\right) \left(\max_{1\le i\le N} \|\lambda_i\|_2\right) \left(\max_{1\le i\le N} \left\|\mathcal{V}_{\lambda i}^{-1/2}\right\|_{op}\right) \\
& = O_p\left(\sqrt{(\ln N)T}\, \|\lambda\|_\infty\right),
\end{aligned}
$$

where $\|\lambda\|_\infty \equiv \max_{1\le i\le N}\{\|\lambda_1\|_2, .., \|\lambda_N\|_2\}$.

(3) $\max_{1\leq i\leq N}\left\|I_3^{\lambda}\right\|_2 \leq \max_{1\leq i\leq N}\left\|\left(\widehat{\lambda}_i-\lambda_i\right)'\mathcal{V}_{\lambda i}^{-1}\left(\widehat{\lambda}_i-\lambda_i\right)\right\|_2 \max_{1\leq i\leq N}\left\|\left(\widehat{\mathcal{V}}_{\lambda i}-\mathcal{V}_{\lambda i}\right)\mathcal{V}_{\lambda i}^{-1}\right\|_F = o_p\left(\ln N\right)$ holds by the immediate results of *Setp 2* in *Part I* in this section and (1) just above.

(4) $\max_{1\leq i\leq N}\left\|I_4^{\lambda}\right\|_2 \leq \max_{1\leq i\leq N}\left\|I_1^{\lambda}\right\|_2 \max_{1\leq i\leq N}\left\|\left(\widehat{\mathcal{V}}_{\lambda i}-\mathcal{V}_{\lambda i}\right)\mathcal{V}_{\lambda i}^{-1}\right\|_F = o_p\left(T\left\|\lambda\right\|_\infty^2\right)$ holds because such two terms are bounded by (1) and (3) just above.

(5) $\max_{1\leq i\leq N}\left\|I_5^{\lambda}\right\|_2 \leq \max_{1\leq i\leq N}\left\|I_2^{\lambda}\right\|_2 \max_{1\leq i\leq N}\left\|\left(\widehat{\mathcal{V}}_{\lambda i}-\mathcal{V}_{\lambda i}\right)\mathcal{V}_{\lambda i}^{-1}\right\|_F = o_p\left(\sqrt{(\ln N)T}\left\|\lambda\right\|_\infty\right)$ holds that such two terms are bounded by (2) and (3) just above.

Step 3. In Step 2 above, under $H_{1a,NT}$, it is obvious that last four terms are dominated by $\max_{1\leq i\leq N}\left\|I_1^{\lambda}\right\|_2$, while $\max_{1\leq i\leq N}\left\|I_1^{\lambda}\right\|_2$ also dominates $\max_{1\leq i\leq N}\left(\widehat{\lambda}_i-\lambda_i\right)'\mathcal{V}_{\lambda i}^{-1}\left(\widehat{\lambda}_i-\lambda_i\right)$ under $H_{1a,NT}$. Therefore, it follows,

$$\mathcal{T}_\lambda = \max_{1\leq i\leq N}\left(\widehat{\lambda}_i-\lambda_i\right)'\mathcal{V}_{\lambda i}^{-1}\left(\widehat{\lambda}_i-\lambda_i\right) + \max_{1\leq i\leq N} I_1^{\lambda} + o_p\left(T\left\|\lambda\right\|_\infty^2\right) \stackrel{d}{=} \mathcal{T}_\lambda^0 + \mathcal{T}_\lambda^{NC},$$

where $\mathcal{T}_\lambda^0$ denotes the null distribution of $\mathcal{T}_\lambda$ and $\mathcal{T}_\lambda^{NC}$ denotes the non-centrality parameter such that $\mathcal{T}_\lambda^{NC} = O_p\left(T\left\|\lambda\right\|_\infty^2\right)$ and diverges to infinity as long as $T\left\|\lambda\right\|_\infty^2/\ln N \to +\infty$.

Under $H_{1a,NT}$, it holds that, as long as (N,T) go to infinity jointly, $P\{\mathcal{T}_\lambda > c_{\alpha,N}\} = P\{\mathcal{T}_\lambda^0 > c_{\alpha,N} - \mathcal{T}_\lambda^{NC}\} \to 1$ holds because $c_{\alpha,N} - \mathcal{T}_\lambda^{NC} \to -\infty$ as $(N,T)\to(\infty,\infty)$ with $c_{\alpha,N} = O(\ln N)$.

A.1.7 Proofs of Theorem 1.4.2

The proofs of Theorem 1.4.2 are very similar to those arguments of proofs for Theorem 1.4.1 in the previous section. So we will give the main steps and key arguments below for brevity.

***Part I*. The asymptotic null distribution of $\mathcal{T}_\gamma$**

Under H_{0b}, $\widehat{\gamma}_t = \widehat{\gamma}_t - \gamma_t$, and according to the decomposition shown in proofs of Theorem 1.3.4,

$$\widehat{\gamma}_t - \gamma_t = -\Sigma_x^{-1}(\xi_{NT}^{\gamma}/\sqrt{T}) + \Sigma_x^{-1}(\zeta_{Nt}^{\gamma}/\sqrt{N}) + \Sigma_x^{-1}(\Psi_{Nt}^{\gamma}/\sqrt{N}) + o_p\left(\delta_{NT}^{-1}\right).$$

And for $\widehat{\mathcal{V}}_{\gamma t}$, the consistent estimator of $\mathcal{V}_{\gamma t}$, we can write

$$\widehat{\mathcal{V}}_{\gamma t}^{-1} = \mathcal{V}_{\gamma t}^{-1} - \mathcal{V}_{\gamma t}^{-1}\left(\widehat{\mathcal{V}}_{\gamma t} - \mathcal{V}_{\gamma t}\right)\widehat{\mathcal{V}}_{\gamma t}^{-1} = \mathcal{V}_{\gamma t}^{-1} - \mathcal{V}_{\gamma t}^{-1}\left(\widehat{\mathcal{V}}_{\gamma t} - \mathcal{V}_{\gamma t}\right)\mathcal{V}_{\gamma t}^{-1} + o_p\left(\left\|\widehat{\mathcal{V}}_{\gamma t} - \mathcal{V}_{\gamma t}\right\|\right).$$

Then, under H_{0b}, neglecting the terms that is $o_p\left(\left\|\widehat{\mathcal{V}}_{\gamma t} - \mathcal{V}_{\gamma t}\right\|\right)$,

$$\begin{aligned}(\widehat{\gamma}_t - \gamma_t)' \widehat{\mathcal{V}}_{\gamma t}^{-1}(\widehat{\gamma}_t - \gamma_t) &= (\widehat{\gamma}_t - \gamma_t)' \mathcal{V}_{\gamma t}^{-1}(\widehat{\gamma}_t - \gamma_t) + (\widehat{\gamma}_t - \gamma_t)' \mathcal{V}_{\gamma t}^{-1}\left(\widehat{\mathcal{V}}_{\gamma t} - \mathcal{V}_{\gamma t}\right)\mathcal{V}_{\gamma t}^{-1}(\widehat{\gamma}_t - \gamma_t) \\ &\equiv (\widehat{\gamma}_t - \gamma_t)' \mathcal{V}_{\gamma t}^{-1}(\widehat{\gamma}_t - \gamma_t) + I_{\gamma t}\end{aligned}$$

Now, we are in the position to show the desired results below by following four steps.

Step 1. It is straightforward to see the fact $\max_{1\le t\le T}(\widehat{\gamma}_t - \gamma_t)' \mathcal{V}_{\gamma t}^{-1}(\widehat{\gamma}_t - \gamma_t) = O_p(\ln T)$ by Theorem 1.3.4 and proofs of Theorem 4 in Castagnetti et al. (2015).

Step 2. we are going to show $\max_{1\le t\le T} I_{\gamma t} = o_p(\ln T)$ as below,

$$\max_{1\le t\le T} I_{\gamma t} \le \max_{1\le t\le T}\left\|(\widehat{\gamma}_t - \gamma_t)' \mathcal{V}_{\gamma t}^{-1/2}\right\|_2^2 \max_{1\le t\le T}\left\|\left(\widehat{\mathcal{V}}_{\gamma t} - \mathcal{V}_{\gamma t}\right)\mathcal{V}_{\gamma t}^{-1}\right\|_F$$

Thus, it suffices to show $\max_{1\le t\le T}\left\|\left(\widehat{\mathcal{V}}_{\gamma t} - \mathcal{V}_{\gamma t}\right)\mathcal{V}_{\gamma t}^{-1}\right\|_F = o_p(1)$, which can be shown by following arguments for $\max_{1\le i\le N}\left\|\left(\widehat{\mathcal{V}}_{\lambda i} - \mathcal{V}_{\lambda i}\right)\mathcal{V}_{\lambda i}^{-1}\right\| = o_p(1)$ in the previous section just above.

Step 3. We show $\sqrt{N}(\widehat{\gamma}_t - \gamma_t)$ and $\sqrt{N}(\widehat{\gamma}_s - \gamma_s)$ are asymptotically independent. Recall that

$$\widehat{\gamma}_t - \gamma_t = -\Sigma_x^{-1}(\xi_{NT}^{\gamma}/\sqrt{T}) + \Sigma_x^{-1}(\zeta_{Nt}^{\gamma}/\sqrt{N}) + \Sigma_x^{-1}(\Psi_{Nt}^{\gamma}/\sqrt{N}) + o_p\left(\delta_{NT}^{-1}\right),$$

where $\xi_{iNT}^{\gamma} = \frac{1}{N\sqrt{T}}\sum_i\sum_{s\neq t} x_{is}^{*}x_{is}^{*\prime}\gamma_s$, $\zeta_{iT}^{\gamma} = \frac{1}{\sqrt{N}}\sum_i\left(x_{it}^{*}x_{it}^{*\prime} - \Sigma_x\right)\lambda_i$, and $\Psi_{iT}^{\gamma} = \frac{1}{\sqrt{N}}\sum_i x_{it}^{*}u_{it}$

Under H_{0b} and Assumption A1.9, the expression of $\sqrt{N}(\widehat{\gamma}_t - \gamma_t)$ simplifies to the below:

$$\sqrt{N}(\widehat{\gamma}_t - \gamma_t) = \Sigma_x^{-1}\frac{1}{\sqrt{N}}\sum_i x_{it}^{*}u_{it} + o_p(1) = \zeta_{Nt}^{\gamma} + o_p(1)$$

Then, it is straightforward to see that the covariance between ζ_{Nt}^{λ} and ζ_{Ns}^{λ} is proportion to the quantity, $E\left(\frac{1}{N}\sum_{i,j}x_{it}^{*}x_{js}^{*\prime}u_{it}u_{js}\right)$ when the smaller-order terms are ignored.

Then, let $\tau = t-s$, under Assumption A1.1 and A1.2, $\lim_{\tau,N\to\infty}\frac{1}{N}\sum_{i,j}E\left(x_{it}^{*}u_{it}u_{jt-\tau}x_{jt-\tau}^{*\prime}\right)\ln T = 0$ as $(N,T)\to\infty$. It concludes ζ_{Nt}^{λ} and ζ_{Ns}^{λ} are asymptotically independent due to the asymptotic normality of ζ_{Nt}^{λ} and ζ_{Ns}^{λ}, and so $\sqrt{N}\left(\widehat{\gamma}_t-\gamma_t\right)$ and $\sqrt{N}\left(\widehat{\gamma}_s-\gamma_s\right)$ are.

Step 4. Following corresponding arguments of proofs for Theorem 4 in Castagnetti et al. (2015), the asymptotic distribution of $\mathcal{T}_\gamma$ under H_{0b} will follow the Gumbel distribution after being rescaled appropriately as shown in Theorem 1.4.2.

***Part II*. The asymptotic local power properties of $\mathcal{T}_\gamma$**

The analysis of the power of test under local alternatives $H_{1b,NT}$ is similar to those in the proof of Theorem 1.4.1 in the previous section just above. So the sketch of proofs are given here for the brevity. Under $H_{1b,NT}$, we can write $\widehat{\gamma}_t = \widehat{\gamma}_t - \gamma_t + \gamma_t$, then, neglecting terms that are of order $o_p\left(\left\|\widehat{\mathcal{V}}_{\gamma t}-\mathcal{V}_{\gamma t}\right\|\right)$, we then have following decomposition,

$$
\begin{aligned}
& \left(\widehat{\gamma}_t-\gamma_t+\gamma_t\right)^{\prime}\widehat{\mathcal{V}}_{\gamma t}^{-1}\left(\widehat{\gamma}_t-\gamma_t+\gamma_t\right) \\
= & \left(\widehat{\gamma}_t-\gamma_t+\gamma_t\right)^{\prime}\mathcal{V}_{\gamma t}^{-1}\left(\widehat{\gamma}_t-\gamma_t+\gamma_t\right)+\left(\widehat{\gamma}_t-\gamma_t+\gamma_t\right)^{\prime}\mathcal{V}_{\gamma t}^{-1}\left(\widehat{\mathcal{V}}_{\gamma t}-\mathcal{V}_{\gamma t}\right)\mathcal{V}_{\gamma t}^{-1}\left(\widehat{\gamma}_t-\gamma_t+\gamma_t\right) \\
= & \left(\widehat{\gamma}_t-\gamma_t\right)^{\prime}\mathcal{V}_{\gamma t}^{-1}\left(\widehat{\gamma}_t-\gamma_t\right)+\gamma_t^{\prime}\mathcal{V}_{\gamma t}^{-1}\gamma_t+2\left(\widehat{\gamma}_t-\gamma_t\right)^{\prime}\mathcal{V}_{\gamma t}^{-1}\gamma_t+\left(\widehat{\gamma}_t-\gamma_t\right)^{\prime}\mathcal{V}_{\gamma t}^{-1}\left(\widehat{\mathcal{V}}_{\gamma t}-\mathcal{V}_{\gamma t}\right)\mathcal{V}_{\gamma t}^{-1}\left(\widehat{\gamma}_t-\gamma_t\right) \\
& +\gamma_t^{\prime}\mathcal{V}_{\gamma t}^{-1}\left(\widehat{\mathcal{V}}_{\gamma t}-\mathcal{V}_{\gamma t}\right)\mathcal{V}_{\gamma t}^{-1}\gamma_t+2\left(\widehat{\gamma}_t-\gamma_t\right)^{\prime}\mathcal{V}_{\gamma t}^{-1}\left(\widehat{\mathcal{V}}_{\gamma t}-\mathcal{V}_{\gamma t}\right)\mathcal{V}_{\gamma t}^{-1}\gamma_t \\
= & \left(\widehat{\gamma}_t-\gamma_t\right)^{\prime}\mathcal{V}_{\gamma t}^{-1}\left(\widehat{\gamma}_t-\gamma_t\right)+I_1^{\gamma}+I_2^{\gamma}+I_3^{\gamma}+I_4^{\gamma}+I_5^{\gamma}.
\end{aligned}
$$

Step 1. It is straightforward to see the fact that $\max_{1\le t\le T}\left(\widehat{\gamma}_t-\gamma_t\right)^{\prime}\mathcal{V}_{\gamma t}^{-1}\left(\widehat{\gamma}_t-\gamma_t\right) = O_p(\ln T)$ by Theorem 1.3.4 and proofs of Theorem 4 in Castagnetti et al. (2015).

Step 2. Similar to corresponding arguments of the analyses of power properties in proofs for Theorem 1.4.1 in the previous section, we can readily obtain that

(1) $\max_{1\le t\le T} I_1^{\gamma} = O_p\left(N\left\|\gamma\right\|_{\infty}^2\right)$, where $\left\|\gamma\right\|_{\infty}^2 = \max_t\left\|\gamma_t\right\|_2^2$.

(2) $\max_{1\le i\le N} I_2^{\gamma} = O_p\left(\sqrt{(\ln T)N}\left\|\gamma\right\|_{\infty}\right)$, where $\left\|\gamma\right\|_{\infty} = \max_t\left\|\gamma_t\right\|_2$

(3) $\max\limits_{1\leq i\leq N} I_3^{\gamma} = o_p\left(\ln T\right)$,

(4) $\max\limits_{1\leq i\leq N} I_4^{\gamma} = o_p\left(N\left\|\gamma_t\right\|_{\infty}^2\right)$,

(5) $\max\limits_{1\leq i\leq N} I_5^{\gamma} = o_p\left(\sqrt{(\ln T)N}\left\|\gamma\right\|_{\infty}\right)$.

Step 3. According to the immediate results in *Step 2* just above, under $H_{1b,NT}$, last four terms are dominated by $\max\limits_{1\leq i\leq N} I_1^{\gamma}$ while $\max\limits_{1\leq i\leq N} I_1^{\gamma}$ also dominates $\max\limits_{1\leq i\leq N}\left(\widehat{\gamma}_t-\gamma_t\right)' \mathcal{V}_{\lambda i}^{-1}\left(\widehat{\gamma}_t-\gamma_t\right)$ provided that $N\left\|\gamma\right\|_{\infty}^2 / \ln T \rightarrow \infty$. Therefore, it follows:

$$\mathcal{T}_{\gamma} = \max_{1\leq t\leq T}\left(\widehat{\gamma}_t-\gamma_t\right)' \mathcal{V}_{\gamma t}^{-1}\left(\widehat{\gamma}_t-\gamma_t\right) + \max_{1\leq t\leq T} I_1^{\gamma} + o_p\left(N\left\|\gamma\right\|_{\infty}^2\right) \overset{d}{=} \mathcal{T}_{\gamma}^0 + \mathcal{T}_{\gamma}^{NC},$$

where $\mathcal{T}_{\gamma}^0$ denotes the null distribution of $\mathcal{T}_{\gamma}$ and $\mathcal{T}_{\gamma}^{NC}$ denotes the non-centrality parameter such that $\mathcal{T}_{\gamma}^{NC} = O_p\left(N\left\|\gamma\right\|_{\infty}^2\right)$ and tends to $+\infty$ if $N\left\|\gamma\right\|_{\infty}^2 / \ln T \rightarrow +\infty$

Under $H_{1b,NT}$, as $(N,T)\rightarrow(\infty,\infty)$, it holds that $P\{\mathcal{T}_{\gamma} > c_{\alpha,T}\} = P\{\mathcal{T}_{\gamma}^0 > c_{\alpha,N} - \mathcal{T}_{\gamma}^{NC}\} \rightarrow 1$ holds as $c_{\alpha,T} - \mathcal{T}_{\gamma}^{NC} \rightarrow -\infty$ with $c_{\alpha,T} = O(\ln T)$.

A.1.8 Proofs of Theorem 1.4.3

Denote the original sample $\mathcal{W}_{NT} \equiv \{\left(y_{it}, x_{it}\right), i = 1,\ldots,N, t = 1,\ldots,T\}$. Let $P^*(\cdot)$ denote the probability measure induced by the wild bootstrap conditional on $\mathcal{W}_{NT}$, and $E^*(\cdot)$ and $Var^*(\cdot)$ denote the expectation and variance with respect to $P^*(\cdot)$. Let $O_{P^*}(\cdot)$ and $o_{P^*}(\cdot)$ denote the probability order under $P^*(.)$.

(a). We now prove that the null distribution of $\mathcal{T}_{\lambda}$ by $\mathcal{T}_{\lambda}^*$ as below.

Part I: The Validity of Block Wild Bootstrap Scheme

Let $\widehat{\lambda}_i^*$, $\mathcal{V}_{\lambda i}^*$ and $\widehat{\mathcal{V}}_{\lambda i}^*$ denote the bootstrap analogue of $\widehat{\lambda}_i$, $\mathcal{V}_{\lambda i}$ and $\widehat{\mathcal{V}}_{\lambda i}$ respectively. It follows that

$$\mathcal{T}_{\lambda}^* \equiv \max_{1\leq i\leq N} \widehat{\lambda}_i^{*\prime}\left(\widehat{\mathcal{V}}_{\lambda i}^*\right)^{-1}\widehat{\lambda}_i^*.$$

For Theorem 1.3.3, we have proved that $\mathcal{T}_\lambda$ will convergence to the Gumbel distribution after being rescaled appropriately as $N, T \to \infty$ under the null hypothesis and Assumptions A1.1-A1.5, A1.8 and A1.9. By Polya-Cantelli lemma, it follows that

$$\sup_{x \in \mathbb{R}} \left| P\left(A_N \mathcal{T}_\lambda \leq x + B_N\right) - \Psi(x) \right| \xrightarrow{P} 0.$$

where $\Psi(x)$ denotes the Gumbel distribution that location parameter equals 0 and scale parameter equals 1. Then the desired result follows if the following statement holds

$$\sup_{x \in \mathbb{R}} \left| P^*\left(A_N \mathcal{T}_\lambda^* \leq x + B_N\right) - \Psi(x) \right| \xrightarrow{P} 0$$

To this end, we show the desired results above by imitating the arguments of proofs for Theorem 4.1. Specifically, our proofs below consist of two parts:

(1) Show the asymptotic normality of $\widehat{\lambda}_i^*$

(2) Show the asymptotic distribution of $\mathcal{T}_\lambda^* \equiv \max_{1 \leq i \leq N} \widehat{\lambda}_i^{*\prime} \left(\widehat{\mathcal{V}}_{\lambda i}^*\right)^{-1} \widehat{\lambda}_i^*$.

Part (1). Conditional on $\mathcal{W}_{NT}$, $\widehat{\ddot{u}}_{it}^*$ are independent across i, and are independent of x_{js} for all i, j, t, s because these objects are fixed in the fixed-design bootstrap world. Note that fore each i, $E^*\left(\widehat{\ddot{u}}_{it}^*\right) = \widehat{\ddot{u}}_{it} E^*\left(w_{it}\right) = 0$ and $E^*\left[\left(\widehat{\ddot{u}}_{it}^*\right)^2\right] = \widehat{\ddot{u}}_{it}^{*2} E\left(w_{it}^2\right) = \widehat{\ddot{u}}_{it}^2$, and these will simplify the proofs in the bootstrap world.

Observing that $\widehat{\ddot{y}}_{it}^* = \tilde{x}_{it}' \lambda_i^* + \widehat{\ddot{u}}_{it}^*$ given that $\lambda_i^* = 0$ for all i because the null hypothesis is maintained in the bootstrap world. Similar to the proofs of Theorem 1.3.3, in the bootstrap world, we also can readily show that, for each i,

$$\widehat{\lambda}_i^* - \lambda_i^* = \left(T^{-1} \sum_t \tilde{x}_{it} \tilde{x}_{it}'\right)^{-1} \left(T^{-1} \sum_t \tilde{x}_{it} \widehat{\ddot{u}}_{it}^*\right) \equiv \left(T^{-1} \sum_t \tilde{x}_{it} \tilde{x}_{it}'\right)^{-1} \left(T^{-1/2} \zeta_{iT}^{\lambda *}\right),$$

where $\widehat{\ddot{u}}_{it}^* = \widehat{\ddot{u}}_{it} w_{it}$ for each $i = 1, \ldots, N$ and $t = 1, \ldots, T$, and $\zeta_{iT}^{\lambda *} = T^{-1/2} \sum_t \tilde{x}_{it} \widehat{\ddot{u}}_{it}^*$.

It is straightforward to see $\frac{1}{T} \sum_t \tilde{x}_{it} \tilde{x}_{it}' = O_p(1)$ by the law of large numbers and is bounded from below under under Assumption A1.1. Then, by continuous mapping theorem, we only need to show the asymptotic normality of $\zeta_{iT}^{\lambda *}$.

We define that, for the j-th block of the cross-section unit i,

$$\begin{aligned} \mathcal{Z}_{ij}^{*} &\equiv \Omega^{*-1/2} b_T^{-1/2} \sum_{l=1}^{b_T} \tilde{x}_{i,(j-1)b_T+l} \widehat{\ddot{u}}_{i,(j-1)b_T+l}^{*} \\ &= \Omega^{*-1/2} b_T^{-1/2} \sum_{l=1}^{b_T} \tilde{x}_{i,(j-1)b_T+l} \widehat{\ddot{u}}_{i,(j-1)b_T+l} w_{i,(j-1)b_T+l} \\ &= \Omega^{*-1/2} b_T^{-1/2} \sum_{l=1}^{b_T} \tilde{x}_{i,(j-1)b_T+l} \widehat{\ddot{u}}_{i,(j-1)b_T+l} \xi_{ij}, \end{aligned}$$

where $i = 1, \ldots, N$ and $j = 1, \ldots, L_T$, and $w_{is} = \xi_{ij}$ if $(j-1)b_T < s \leq jb_T$ holds by construction as stated in Step 2 and 3 in Algorithm BWB, and

$$\Omega^{*} = T^{-1} Var^{*} \left(\sum_{j=1}^{L_T} \sum_{l=1}^{b_T} \tilde{x}_{i,(j-1)b_T+l} \widehat{\ddot{u}}_{i,(j-1)b_T+l} \xi_{ij} \right).$$

Then, we can write

$$\zeta_{iT}^{\lambda *} = T^{-1/2} \sum_{t} \tilde{x}_{it} \widehat{\ddot{u}}_{it}^{*} = L_T^{-1/2} \sum_{j=1}^{L_T} \mathcal{Z}_{ij}^{*},$$

where $\{\mathcal{Z}_{ij}\}$ are independent across j for each i conditional on $\mathcal{W}_{NT}$ for $j = 1, \ldots, L_T$ with $E^{*}\left(\mathcal{Z}_{ij}^{*}\right) = 0$ and $Var(L_T^{-1/2} \sum_{j=1}^{L_T} \mathcal{Z}_{ij}^{*}) = I_K$ by the definition of Ω^{*} displayed just above.

To establish the asymptotic normality of $\zeta_{iT}^{\lambda *} = L_T^{-1/2} \sum_{j=1}^{L_T} \mathcal{Z}_{ij}^{*}$, it suffices to show that the Lyapunov condition is satisfied, namely, $\sum_{j=1}^{L_T} E^{*} \left\| \mathcal{Z}_{ij}^{*} / \sqrt{L_T} \right\|_2^4 = o(1)$. By the definition of $\mathcal{Z}_{ij}^{*}$ above, we have

$$L_T^{-2} \sum_{j=1}^{L_T} E^{*} \left\| \mathcal{Z}_{ij} \right\|_2^4 \leq \left\| \Omega^{*-1/2} \right\|_F^4 T^{-2} \sum_{j=1}^{L_T} E^{*} \left\| \sum_{l=1}^{b_T} \tilde{x}_{i,(j-1)b_T+l} \widehat{\ddot{u}}_{i,(j-1)b_T+l}^{*} \right\|_2^4$$

Note that the requirements for b_T below can generalized indeed as stated in Theorem 1.4.3(a).

We claim that $\Omega^* \xrightarrow{P} \widehat{\widehat{\Sigma}}_{i,xu}$, where $\widehat{\widehat{\Sigma}}_{i,xu}$ is defined in Corollary 1.3.5, then $\left\|\Omega^{*-1/2}\right\|_F^4 = O_p(1)$ holds, and prove this claim later. Thus, we only need to show remaining terms above are $o(1)$ as whole. Noting that

$$\begin{aligned}
&\widehat{\ddot{u}}^*_{i,(j-1)b_T+l} \\
= &\left[u_{it} + \ddot{x}_{it}'(\beta - \widehat{\beta}) + \tilde{x}_{it}'\left(\lambda_i - \widehat{\lambda}_i\right) - \frac{1}{N}\sum_j \tilde{x}_{jt}'\left(\lambda_j - \widehat{\lambda}_j\right) + \breve{x}_{it}'\left(\gamma_t - \widehat{\gamma}_t\right) - \frac{1}{T}\sum_s \breve{x}_{is}'\left(\gamma_s - \widehat{\gamma}_s\right)\right]\xi_{ij}.
\end{aligned}$$

Then, by C_r-inequality, we have that

$$T^{-2}\sum_{j=1}^{L_T} E^*\left\|\sum_{l=1}^{b_T} \tilde{x}^*_{i,(j-1)b_T+l}\widehat{\ddot{u}}^*_{i,(j-1)b_T+l}\right\|_2^4 \le 4\left(M_{1T} + M_{2T} + M_{3T} + M_{4T} + M_{5T} + M_{6T}\right),$$

where,

$$\begin{aligned}
M_{1T} &= 6^3\left(\frac{1}{T}\right)^2 \sum_{j=1}^{L_T}\left\|\sum_{l=1}^{b_T}\tilde{x}_{i,(j-1)b_T+l}u_{i,(j-1)b_T+l}\right\|_2^4 E^*\left|\xi_{ij}\right|^4, \\
M_{2T} &= 6^3\left(\frac{1}{T}\right)^2 \sum_{j=1}^{L_T}\left\|\sum_{l=1}^{b_T}\tilde{x}_{i,(j-1)b_T+l}\ddot{x}'_{i,(j-1)b_T+l}(\beta - \widehat{\beta})\right\|_2^4 E^*\left|\xi_{ij}\right|^4, \\
M_{3T} &= 6^3\left(\frac{1}{T}\right)^2 \sum_{j=1}^{L_T}\left(\left\|\sum_{l=1}^{b_T}\tilde{x}_{i,(j-1)b_T+l}\tilde{x}'_{i,(j-1)b_T+l}\left(\lambda_i - \widehat{\lambda}_i\right)\right\|_2\right)^4 E^*\left|\xi_{ij}\right|^4, \\
M_{4T} &= 6^3\left(\frac{1}{T}\right)^2 \sum_{j=1}^{L_T}\left\|\frac{1}{N}\sum_h\sum_{l=1}^{b_T}\tilde{x}_{i,(j-1)b_T+l}\tilde{x}'_{h,(j-1)b_T+l}\left(\lambda_h - \widehat{\lambda}_h\right)\right\|_2^4 E^*\left|\xi_{ij}\right|^4, \\
M_{5T} &= 6^3\left(\frac{1}{T}\right)^2 \sum_{j=1}^{L_T}\left\|\sum_{l=1}^{b_T}\tilde{x}_{i,(j-1)b_T+l}\breve{x}'_{i,(j-1)b_T+l}\left(\gamma_{(j-1)b_T+l} - \widehat{\gamma}_{(j-1)b_T+l}\right)\right\|_2^4 E^*\left|\xi_{ij}\right|^4, \\
M_{6T} &= 6^3\left(\frac{1}{T}\right)^2 \sum_{j=1}^{L_T}\left(\left\|\frac{1}{T}\sum_s\sum_{l=1}^{b_T}\tilde{x}_{i,(j-1)b_T+l}\breve{x}'_{is}\left(\gamma_s - \widehat{\gamma}_s\right)\right\|_2\right)^4 E^*\left|\xi_{ij}\right|^4.
\end{aligned}$$

Below we show that M_{1T} to M_{6T} are all of order $o_{p^*}(1)$ terms below provided that $E^*\left|\xi_{ij}\right|^4 < \infty$ by the condition stated in Algorithm BWB and note that $T = b_T L_T$.

For M_{1T}, by direct moment calculations, $\left\|\sum_{l=1}^{b_T}\tilde{x}_{i,(j-1)b_T+l}u_{i,(j-1)b_T+l}\right\|_2^4 = O_p\left(b_T^4\right)$ holds under

Assumption A1.1, it follows $M_{1T} = O_p\left(b_T^4 L_T T^{-2}\right) = O_p\left(b_T^3 T^{-1}\right)$.

For M_{2T}, using the fact that $\widehat{\beta} - \beta = O_p\left(N^{-1/2} + T^{-1/2}\right)$, in bootstrap world, by direct calculation, use triangular and Cauchy-Schwartz inequality, we can readily show as below by neglecting those terms in the expansions of M_{2T} that are of smaller order,

$$\begin{aligned} M_{2T} &= 6^3 \left(\frac{1}{T}\right)^2 \sum_{j=1}^{L_T} \left\| \sum_{l=1}^{b_T} \tilde{x}_{i,(j-1)b_T+l} \tilde{x}'_{i,(j-1)b_T+l} (\beta - \widehat{\beta}) \right\|_2^4 E^* \left|\xi_{ij}\right|^4 \\ &\leq 6^3 \left\|\beta - \widehat{\beta}\right\|_2^4 \left(\frac{1}{T}\right)^2 \sum_{j=1}^{L_T} \left\| \sum_{l=1}^{b_T} \tilde{x}_{i,(j-1)b_T+l} \tilde{x}'_{i,(j-1)b_T+l} \right\|_F^4 E^* \left|\xi_{ij}\right|^4 \\ &= o_p(1) O_p\left(T^{-2} L_T b_T^4\right) O(1) = O_p\left(b_T^3 T^{-3/2} + N^{-1/2} b_T^3 T^{-1}\right), \end{aligned}$$

where $\left\| \sum_{l=1}^{b_T} \tilde{x}_{i,(j-1)b_T+l} \tilde{x}'_{i,(j-1)b_T+l} \right\|_F^4 = O_p\left(b_T^4\right)$ by direct calculations under Assumption A1.1.

For M_{3T} we can show $M_{3T} = O_p\left(b_T^3 T^{-3}\right)$ readily by following above arguments for M_{1T} and M_{2T}.

For M_{4T}, we have

$$\begin{aligned} M_{4T} &\leq 6^3 \left(\frac{1}{T}\right)^2 \sum_{j=1}^{L_T} \left\| \sum_{l=1}^{b_T} \tilde{x}_{i,(j-1)b_T+l} \tilde{x}'_{h,(j-1)b_T+l} \right\|_F^4 \left(\frac{1}{N} \sum_{h \neq i} \left\| \lambda_h - \widehat{\lambda}_h \right\|_2^2 \right)^2 E^* \left|\xi_{ij}\right|^4 \\ &= T^{-2} L_T O_p\left(b_T^4\right) O_p\left(T^{-2} (\ln N)^2\right) O(1) \\ &= O_p\left((\ln N)^2 T^{-4} L_T b_T^4\right) = O_p\left(\ln N)^2 T^{-3} b_T^3\right), \end{aligned}$$

where the second line holds because $\left\| \sum_{l=1}^{b_T} \tilde{x}_{i,(j-1)b_T+l} x^{*\prime}_{h,(j-1)b_T+l} \right\|_F^4 = O\left(b_T^4\right)$ by direct moments calculations under Assumption A1.1, and $\max_i \left\| \lambda_i - \widehat{\lambda}_i \right\| = O_p\left(\sqrt{\ln N / T}\right)$ by Proposition 1.3.8.

For M_{5T} and M_{6T}, similar to arguments for bounding M_{4T} above, we can readily show that

$$M_{5T} = O_p\left(N^{-2} (\ln T)^2 T^{-1} b_T^3\right), \text{ and } M_{6T} = O_p\left(N^{-2} (\ln T)^2 T^{-1} b_T^3\right),$$

holds by under Assumption A1.1 and Proposition 1.3.8.

Therefore, if b_T is set to be $o(T^{1/3})$, M_{1T} to M_{6T} are all of order $o_{p^*}(1)$. After verifying the Lyapunov condition, $\zeta_{iT}^{\lambda*} \xrightarrow{d} N(0, \Omega^*)$ follows directly by the CLT.

Now, we are in the position to show the consistency of Ω^* constructed above. By construction, we have

$$\begin{aligned}
\Omega^* &= T^{-1} Var^* \left(\sum_{j=1}^{L_T} \sum_{l=1}^{b_T} \tilde{x}_{i,(j-1)b_T+l} \widehat{\ddot{u}}_{i,(j-1)b_T+l} \xi_{ij} \right) \\
&= T^{-1} \sum_{j=1}^{L_T} \left[\sum_{l=1}^{b_T} \sum_{k=1}^{b_T} \left(\tilde{x}_{i,(j-1)b_T+l} \widehat{\ddot{u}}_{i,(j-1)b_T+l} \right) \left(\tilde{x}_{i,(j-1)b_T+l} \widehat{\ddot{u}}_{i,(j-1)b_T+k} \right)' \right] E^* \left(\xi_{ij}^2 \right) \\
&= L_T^{-1} \sum_{j=1}^{L_T} \left[b_T^{-1} \sum_{l=1}^{b_T} \sum_{k=1}^{b_T} \left(\tilde{x}_{i,(j-1)b_T+l} \widehat{\ddot{u}}_{i,(j-1)b_T+l} \right) \left(\tilde{x}_{i,(j-1)b_T+l} \widehat{\ddot{u}}_{i,(j-1)b_T+k} \right)' \right] \qquad \text{(A.1.1)}
\end{aligned}$$

By direct moments calculations, $\Omega^* \xrightarrow{P^*} \widehat{\tilde{\Sigma}}_{i,xu}$ holds given that $1/b_T + b_T/T^{1/3} \to 0$ as $T \to \infty$, where $\widehat{\tilde{\Sigma}}_{i,xu}$ is given in Corollary 1.3.5. We give the sketch of the main steps and omit the details because we can follow the arguments for verifying the Lyapunov condition just above to get the desired result. Recall that

$$\begin{aligned}
& \widehat{\ddot{u}}^*_{i,(j-1)b_T+l} \\
= & \left[u_{it} + \ddot{x}_{it}'(\beta - \widehat{\beta}) + \tilde{x}_{it}' \left(\lambda_i - \widehat{\lambda}_i \right) - \frac{1}{N} \sum_j \tilde{x}_{jt}' \left(\lambda_j - \widehat{\lambda}_j \right) + \breve{x}_{it}' \left(\gamma_t - \widehat{\gamma}_t \right) - \frac{1}{T} \sum_s \breve{x}_{is}' \left(\gamma_s - \widehat{\gamma}_s \right) \right] \xi_{ij}.
\end{aligned}$$

By simple calculations, it can be seen that

$$\begin{aligned}
\Omega^* &= L_T^{-1} \sum_{j=1}^{L_T} \left[b_T^{-1} \sum_{l=1}^{b_T} \sum_{k=1}^{b_T} \left(x^*_{i,(j-1)b_T+l} u_{i,(j-1)b_T+l} \right) \left(x_{i,(j-1)b_T+l} u_{i,(j-1)b_T+k} \right)' \right] + \text{dominated terms} \\
&\equiv L_T^{-1} \sum_{j=1}^{L_T} \Omega_j^* + \text{dominated terms}.
\end{aligned}$$

Intuitively, as $b_T \to \infty$, $\Omega_j^* \xrightarrow{p} \tilde{\Sigma}_{i,xu}$, and so does Ω^*. Meanwhile, $\widehat{\tilde{\Sigma}}_{i,xu}$ is constructed as the HAC estimator of $\tilde{\Sigma}_{i,xu}$. Therefore, it also can be shown that $\left\| \Omega^* - \widehat{\tilde{\Sigma}}_{i,xu} \right\|_F = o_p(1)$.

So the asymptotic normality of $\widehat{\lambda}_i^*$ holds in the view of those arguments of the proof for Theorem 3.2 in Lahiri (2013).

Part (2). Now, we derive the asymptotic distribution of $\mathcal{T}_\lambda^*$ in bootstrap world.

To this end, we denote $\widehat{\mathcal{V}}_{\lambda i} \equiv T^{-1}Q_{\tilde{x}}^{-1}\widehat{\widehat{\Sigma}}_{i,xu}Q_{\tilde{x}}^{-1}$, and $\widehat{\mathcal{V}}_{\lambda i}^* = T^{-1}Q_{\tilde{x}}^{-1}\widehat{\widehat{\Sigma}}_{i,xu}^*Q_{\tilde{x}}^{-1}$, where $Q_{\tilde{x}} = T^{-1}\sum_t \tilde{x}_{it}\tilde{x}_{it}'$,$\widehat{\widehat{\Sigma}}_{i,xu}^*$ is calculated by the same way as $\widehat{\widehat{\Sigma}}_{i,xu}$, then we can follow those arguments of proofs for Theorem 1.4.1 and just give the outline here for brevity.

$$\widehat{\mathcal{V}}_{\lambda i}^{*-1} = \widehat{\mathcal{V}}_{\lambda i}^{-1} - \widehat{\mathcal{V}}_{\lambda i}^{-1}\left(\widehat{\mathcal{V}}_{\lambda i}^* - \widehat{\mathcal{V}}_{\lambda i}\right)\widehat{\mathcal{V}}_{\lambda i}^{*-1} + o_p\left(\left\|\widehat{\mathcal{V}}_{\lambda i}^* - \widehat{\mathcal{V}}_{\lambda i}\right\|\right).$$

Similarly, neglecting those terms that are $o_p\left(\left\|\widehat{\mathcal{V}}_{\lambda i}^* - \widehat{\mathcal{V}}_{\lambda i}\right\|\right)$,

$$\begin{aligned}
&\left(\widehat{\lambda}_i^* - \lambda_i^*\right)'\widehat{\mathcal{V}}_{\lambda i}^{*-1}\left(\widehat{\lambda}_i^* - \lambda_i^*\right)\\
=\ &\left(\widehat{\lambda}_i^* - \lambda_i^*\right)'\widehat{\mathcal{V}}_{\lambda i}^{-1}\left(\widehat{\lambda}_i^* - \lambda_i\right) + \left(\widehat{\lambda}_i^* - \lambda_i^*\right)'\widehat{\mathcal{V}}_{\lambda i}^{-1}\left(\widehat{\mathcal{V}}_{\lambda i}^* - \widehat{\mathcal{V}}_{\lambda i}\right)\widehat{\mathcal{V}}_{\lambda i}^{-1}\left(\widehat{\lambda}_i^* - \lambda_i^*\right)'\\
\equiv\ &\left(\widehat{\lambda}_i^* - \lambda_i^*\right)'\mathcal{V}_{\lambda i}^{-1}\left(\widehat{\lambda}_i^* - \lambda_i^*\right) + I_{\lambda i}^*.
\end{aligned}$$

By repeating the exact same arguments as those in the proofs of Theorem 1.4.1, it is straightforward to see that $\max_{1\le i\le N}\left(\widehat{\lambda}_i^* - \lambda_i^*\right)'\mathcal{V}_{\lambda i}^{-1}\left(\widehat{\lambda}_i^* - \lambda_i^*\right) = O_{p^*}(\ln N)$ holds.

Note that $\max_{1\le i\le N} I_{\lambda i}^* = o_p(\ln N)$ holds if the fact

$$\begin{aligned}
\max_{1\le i\le N} I_{\lambda i}^* &\le \max_{1\le i\le N}\left\|\left(\widehat{\lambda}_i^* - \lambda_i^*\right)'\mathcal{V}_{\lambda i}^{-1}\left(\widehat{\lambda}_i^* - \lambda_i^*\right)\right\|_2 \max_{1\le i\le N}\left\|\left(\widehat{\mathcal{V}}_{\lambda i}^* - \widehat{\mathcal{V}}_{\lambda i}\right)\widehat{\mathcal{V}}_{\lambda i}^{-1}\right\|_2\\
&= O_{p^*}(\ln N)o_{p^*}(1) = o_{p^*}(\ln N).
\end{aligned}$$

holds. Again, it suffices to show $\max_{1\le i\le N}\left\|\left(\widehat{\mathcal{V}}_{\lambda i}^* - \mathcal{V}_{\lambda i}^*\right)\mathcal{V}_{\lambda i}^{*-1}\right\|_2 = o_p(1)$, such proofs follow the exact same argument for Theorem 1.4.1 in the bootstrap world, and those details are not repeated here again.

Part II: The Validity of Wild Bootstrap Scheme

Note that we construction $\mathcal{T}_\lambda$ by the observations $\{y_{it}, x_{it}\}_{t=1}^N$ for each i.

Because the wild bootstrap is the special case of the Block Wild Bootstrap when $b_T = 1$. We just introduce some notations here and the details can imitate the arguments in Part I just above.

Particularly, by similar notations used in Part I above, denote

$$\begin{aligned}\mathcal{Z}_{ij}^{*} \equiv \Omega^{*-1/2}\tilde{x}_{i,(j-1)+1}\widehat{\ddot{u}}_{i,(j-1)+1}^{*} &= \Omega^{*-1/2}\tilde{x}_{i,(j-1)+1}\widehat{\ddot{u}}_{i,(j-1)+1}w_{it} \\ &= \Omega^{*-1/2}\tilde{x}_{i,(j-1)+1}\widehat{\ddot{u}}_{i,(j-1)+1}\xi_{ij},\end{aligned}$$

for $i = 1, \ldots, N$ and $j = 1, \ldots, T$, where $\Omega^{*} = T^{-1}Var^{*}\left(\sum_{j=1}^{T}\tilde{x}_{i,(j-1)+l}\widehat{\ddot{u}}_{i,(j-1)+l}\xi_{ij}\right)$ and $w_{it} = \xi_{ij}$ if $t = j$ by constructions in Algorithm BWB.

Then, we can write $T^{-1/2}\sum_t \tilde{x}_{it}\widehat{\ddot{u}}_{it}^{*} = T^{-1/2}\sum_{j=1}^{T}\mathcal{Z}_{ij}^{*}$, where $\mathcal{Z}_{ij}$ are independent across j conditional on $\mathcal{W}_{NT}$ for $j = 1, \ldots, T$ with $E^{*}\left(\mathcal{Z}_{ij}^{*}\right) = 0$ and $Var(T^{-1/2}\sum_{j=1}^{T}\mathcal{Z}_{ij}^{*}) = I$ by the construction of Ω^{*}.

The asymptotic normality of $T^{-1/2}\sum_{j=1}^{T}\mathcal{Z}_{ij}^{*}$ requires to verify the Lyapunov condition, namely, $\sum_{j=1}^{T}E^{*}\left\|\mathcal{Z}_{ij}^{*}/\sqrt{T}\right\|_2^4 = o(1)$.

With these notations, the remaining steps are the same as those in Part I just above.

(b). The Validity of Wild Bootstrap Scheme

Note that we construction $\mathcal{T}_\gamma$ by the observations $\{y_{it}, x_{it}\}_{i=1}^{N}$ for each t.

Observing that $\widehat{\ddot{y}}_{it}^{*} = \breve{x}_{it}'\gamma_t^{*} + \widehat{\ddot{u}}_{it}^{*}$ given that $\gamma_t^{*} = 0$ for all t because the null hypothesis is maintained in the bootstrap world. Similar to the proofs of Theorem 1.3.4, in the bootstrap world, we also can readily show that, for each t,

$$\widehat{\gamma}_t^{*} - \gamma_t^{*} = (N^{-1}\textstyle\sum_i \breve{x}_{it}\breve{x}_{it}')^{-1}\left(N^{-1}\sum_i \breve{x}_{it}\widehat{\ddot{u}}_{it}^{*}\right) \equiv (N^{-1}\sum_i \breve{x}_{it}\breve{x}_{it}')^{-1}\left(T^{-1/2}\zeta_{Nt}^{\gamma *}\right),$$

where $\widehat{\ddot{u}}_{it}^{*} = \widehat{\ddot{u}}_{it}w_{it}$ for each $i = 1, \ldots, N$ and $t = 1, \ldots, T$, and $\zeta_{Nt}^{\gamma *} = N^{-1/2}\sum_i \breve{x}_{it}\widehat{\ddot{u}}_{it}^{*}$.

Similar to the arguments used for Part II of (a) just above, if we set block size $b_N = 1$, then, we can treat the wild bootstrap as the special case of block wild bootstrap. So proofs here are very similar to arguments for the proofs in (a) shown above.

Particularly, denote

$$\begin{aligned}\mathcal{Z}_{jt}^{*} \equiv \Omega^{*-1/2} \breve{x}_{(j-1)+1,t} \widehat{\ddot{u}}_{(j-1)+1,t}^{*} &= \Omega^{*-1/2} \breve{x}_{(j-1)+1,t} \widehat{\ddot{u}}_{(j-1)+1,t} w_{it} \\ &= \Omega^{*-1/2} \breve{x}_{(j-1)+1,t} \widehat{\ddot{u}}_{(j-1)+1,t} \xi_{jt},\end{aligned}$$

for $j = 1, \ldots, N$ and $t = 1, \ldots, T$, where $\Omega^{*} = N^{-1} Var^{*} \left(\sum_{i=1}^{N} \breve{x}_{i,(j-1)b_T+l} \widehat{\ddot{u}}_{i,(j-1)b_T+l} \xi_{jt} \right)$ converges to $\widehat{\breve{\Sigma}}_{Xu}$ that is defined in Corollary 1.3.6, and $w_{it} = \xi_{jt}$ if $i = j$ by constructions in Algorithm WB.

Then, we can write $N^{-1/2} \sum_{i=1}^{N} \breve{x}_{it} \widehat{\ddot{u}}_{it}^{*} = N^{-1/2} \sum_{j=1}^{N} \mathcal{Z}_{jt}^{*}$, where $\mathcal{Z}_{jt}$ are independent across j conditional on $\mathcal{W}_{NT}$ for $j = 1, \ldots, N$ with $E^{*}\left(\mathcal{Z}_{jt}^{*}\right) = 0$ and $Var(N^{-1/2} \sum_{j=1}^{N} \mathcal{Z}_{jt}^{*}) = I$ by the construction of Ω^{*} above.

The asymptotic normality of $N^{-1/2} \sum_{j=1}^{N} \mathcal{Z}_{jt}^{*}$ requires the verification of the Lyapunov condition, namely, $\sum_{j=1}^{N} E^{*} \left\| \mathcal{Z}_{jt}^{*} / \sqrt{N} \right\|_{2}^{2+d} = o(1)$ for some $d > 0$.

The remaining steps are very similar to proofs in (a) and thus omitted here.

A.2 Proofs of Propositions in A.1

This section is composed of three parts. Section S1 contains the proofs of theorems, corollaries and propositions in the main texts. Section S2 contains some technical lemmas that are used in the proofs in the main texts and Section S1. Section S3 provides some discussions on the theoretical results under the heterogeneity of x_{it}. We continues to use the notations defined at the end of Section 1 of the paper. Let $\sum_{j \neq i}$ and $\sum_{t \neq s}$ denote $\sum_{i=1}^{N} \sum_{j=1, j \neq i}^{N}$ and $\sum_{t=1}^{T} \sum_{s=1, t \neq s}^{T}$, respectively.

In this section, we prove Propositions A.1.1–A.1.3.

A.2.1 Proof of the Proposition A.1.1

To prove A.1.1, we need the following four lemmas that will be proved after we finish the proof of the proposition.

Lemma A.2.1. *Suppose Assumption A1.1 holds. Then*

(i) $\left\|\frac{1}{NT}\sum_i\sum_t x^*_{it}x^{*\prime}_{it}\right\|_F = O_p(1)$,

(ii) $\left\|\frac{1}{NT^2}\sum_i\sum_t\sum_s x^*_{it}x^{*\prime}_{is}\right\|_F = O_p(\frac{1}{T})$,

(iii) $\left\|\frac{1}{N^2T}\sum_i\sum_{j\neq i}\sum_t x^*_{it}x^{*\prime}_{jt}\right\|_F = O_p(\frac{1}{N})$,

(iv) $\left\|\frac{1}{N^2T^2}\sum_i\sum_{j\neq i}\sum_t\sum_s x^*_{it}x^{*\prime}_{js}\right\|_F = O_p(\frac{1}{NT})$.

Lemma A.2.2. *Suppose that Assumptions A1.1 and A1.2 hold. Then*

(i) $\left\|\frac{1}{NT}\sum_i\sum_t x^*_{it}u_{it}\right\| = O_p(\frac{1}{\sqrt{NT}})$,

(ii) $\left\|\frac{1}{NT^2}\sum_i\sum_t\sum_s x^*_{it}u_{is}\right\| = O_p(\frac{1}{\sqrt{NT^2}})$,

(iii) $\left\|\frac{1}{N^2T}\sum_i\sum_l\sum_t x^*_{lt}u_{it}\right\| = O_p(\frac{1}{\sqrt{N^2T}})$,

(iv) $\left\|\frac{1}{N^2T^2}\sum_i\sum_l\sum_t\sum_r x^*_{lt}u_{ir}\right\| = O_p(\frac{1}{NT})$.

Lemma A.2.3. *Under Assumption A1.1,*

(i) $\left\|\frac{1}{NT}\sum_i\sum_t x^*_{it}x^{*\prime}_{it}\lambda_i\right\| = O_p(\frac{1}{\sqrt{N}})$,

(ii) $\left\|\frac{1}{NT^2}\sum_i\sum_t\sum_s x^*_{it}x^{*\prime}_{is}\lambda_i\right\| = O_p(\frac{1}{T})$

(iii) $\left\|\frac{1}{N^2T}\sum_{l\neq i}\sum_i\sum_t x^*_{lt}x^{*\prime}_{it}\lambda_i\right\| = O_p(\frac{1}{\sqrt{N^2T}})$,

(iv) $\left\|\frac{1}{N^2T^2}\sum_{l\neq i}\sum_i\sum_t\sum_s x^*_{lt}x^{*\prime}_{is}\lambda_i\right\| = O_p(\frac{1}{NT})$.

Lemma A.2.4. *Under Assumptions A1.1,*

(i) $\left\|\frac{1}{NT}\sum_i\sum_t x^*_{it}x^{*\prime}_{it}\gamma_t\right\|_2 = O_p(\frac{1}{\sqrt{T}})$,

(ii) $\left\|\frac{1}{NT^2}\sum_i\sum_t\sum_r x^*_{ir}x^{*\prime}_{it}\gamma_t\right\| = O_p(\frac{1}{\sqrt{NT^2}})$,

(iii) $\left\|\frac{1}{N^2T}\sum_i\sum_{l\neq i}\sum_t x^*_{lt}x^{*\prime}_{it}\gamma_t\right\| = O_p(\frac{1}{\sqrt{N^2T}})$,

(iv) $\left\|\frac{1}{N^2T^2}\sum_i\sum_{l\neq i}\sum_t\sum_r x^*_{lr}x^{*\prime}_{it}\gamma_t\right\| = O_p(\frac{1}{NT})$.

Proof of Proposition A.1.1.

Recall that

$$\begin{aligned}\widehat{\beta}-\beta &= \left(\frac{1}{NT}\sum_i\sum_t \ddot{x}_{it}\ddot{x}'_{it}\right)^{-1}\left[\frac{1}{NT}\sum_i\sum_t \ddot{x}_{it}\left(\ddot{u}_{it}+\ddot{\theta}_{it}+\ddot{\varphi}_{it}\right)\right] \\ &\equiv A^{-1}_{1NT}\left(B_{1NT}+B_{2NT}+B_{3NT}\right),\end{aligned}$$

where A_{1NT}, B_{1NT}, B_{2NT}, and B_{3NT} are defined in the proof of Theorem 3.1. Note that

$$
\begin{aligned}
A_{1NT} &= \frac{1}{NT}\sum_i\sum_t\left(\tilde{x}_{it}-\frac{1}{N}\sum_j\tilde{x}_{jt}\right)\left(\tilde{x}_{it}-\frac{1}{N}\sum_j\tilde{x}_{jt}\right)' \\
&= \frac{1}{NT}\sum_i\sum_t\tilde{x}_{it}\tilde{x}_{it}'-\frac{1}{N^2T}\sum_{i,j}\sum_t\tilde{x}_{it}\tilde{x}_{jt}'\equiv A_{1NT1}-A_{1NT2}, \\
B_{1NT} &= \frac{1}{NT}\sum_i\sum_t\left(\tilde{x}_{it}-\frac{1}{N}\sum_j\tilde{x}_{jt}\right)\left(\tilde{u}_{it}-\frac{1}{N}\sum_j\tilde{u}_{jt}\right) \\
&= \frac{1}{NT}\sum_i\sum_t\tilde{x}_{it}\tilde{u}_{it}-\frac{1}{NT}\sum_i\sum_t\left(\frac{1}{N}\sum_j\tilde{x}_{jt}\right)\tilde{u}_{it}\equiv B_{1NT1}-B_{1NT2}, \\
B_{2NT} &= \frac{1}{NT}\sum_i\sum_t\left(\tilde{x}_{it}-\frac{1}{N}\sum_j\tilde{x}_{jt}\right)\left(\tilde{x}_{it}'\lambda_i-\frac{1}{N}\sum_j\tilde{x}_{jt}'\lambda_j\right) \\
&= \frac{1}{NT}\sum_i\sum_t\tilde{x}_{it}\tilde{x}_{it}'\lambda_i-\frac{1}{NT}\sum_i\sum_t(\frac{1}{N}\sum_j\tilde{x}_{jt})\tilde{x}_{it}'\lambda_i\equiv B_{2NT1}-B_{2NT2}, \\
B_{3NT} &= \frac{1}{NT}\sum_i\sum_t\left(\breve{x}_{it}-\frac{1}{T}\sum_s\breve{x}_{is}\right)\left(\breve{x}_{it}'\gamma_t-\frac{1}{T}\sum_r\breve{x}_{ir}'\gamma_r\right) \\
&= \frac{1}{NT}\sum_i\sum_t\breve{x}_{it}\breve{x}_{it}'\gamma_t-\frac{1}{NT^2}\sum_i\sum_{s,t}\breve{x}_{is}\breve{x}_{it}'\gamma_t\equiv B_{3NT1}-B_{3NT2}. \qquad \text{(A.2.1)}
\end{aligned}
$$

(i) Note that $\tilde{x}_{it}=\tilde{x}_{it}^*$ where $\tilde{x}_{it}^*=\tilde{x}_{it}-E\left(\tilde{x}_{it}\right).$ Then

$$
\begin{aligned}
A_{1NT} &= \frac{1}{NT}\sum_i\sum_t\tilde{x}_{it}^*\tilde{x}_{it}^{*\prime}-\frac{1}{N^2T}\sum_{i,j}\sum_t\tilde{x}_{it}^*\tilde{x}_{jt}^{*\prime} \\
&= \frac{1}{NT}\sum_i\sum_t x_{it}^*x_{it}^{*\prime}-\frac{1}{NT^2}\sum_i\sum_t\sum_s x_{it}^*x_{is}^{*\prime}-\frac{1}{N^2T}\sum_{i,j}\sum_t x_{it}^*x_{jt}^{*\prime}+\frac{1}{N^2T^2}\sum_{i,j}\sum_{t,s}x_{it}^*x_{js}^{*\prime} \\
&= \left(1-\frac{1}{N}\right)A_{1NT1}-\left(1-\frac{1}{N}\right)A_{1NT2}-A_{1NT3}+A_{1NT4},
\end{aligned}
$$

where the first equality holds by (A.2.1), $A_{1NT1}\equiv\frac{1}{NT}\sum_i\sum_t x_{it}^*x_{it}^{*\prime}$, $A_{1NT2}\equiv\frac{1}{NT^2}\sum_i\sum_{t,s}x_{it}^*x_{is}^{*\prime}$, $A_{1NT3}\equiv\frac{1}{N^2T}\sum_{j\neq i}\sum_t x_{it}^*x_{jt}^{*\prime}$, and $A_{1NT4}\equiv\frac{1}{N^2T^2}\sum_{j\neq i}\sum_{t,s}x_{it}^*x_{js}^{*\prime}$. By Lemma A.2.1, $A_{1NT1}=O_p\left(1\right)$, $A_{1NT2}=O_p(\frac{1}{T})$, $A_{1NT3}=O_p(\frac{1}{N})$, and $A_{1NT4}=O_p(\frac{1}{NT})$. In addition, it is standard to show that

$$
A_{1NT1}\xrightarrow{p}\lim_{(N,T)\to\infty}\frac{1}{NT}\sum_i\sum_t E\left(x_{it}^*x_{it}^{*\prime}\right)\equiv\Sigma_x.
$$

Then (i) follows by the Slutsky lemma.

(ii) Note that $B_{1NT} = B_{1NT1} - B_{1NT2}$ by (A.2.1). So it suffices to bound B_{1NT1} and B_{1NT2} in turn. Using $\tilde{x}_{it} = \tilde{x}^*_{it}$ again, we have

$$
\begin{aligned}
B_{1NT1} &= \frac{1}{NT}\sum_i \sum_t \tilde{x}^*_{it}\tilde{u}_{it} \\
&= \frac{1}{NT}\sum_i \sum_t \left(x^*_{it} - \frac{1}{T}\sum_r x^*_{ir}\right)(u_{it} - \frac{1}{T}\sum_q u_{iq}) \\
&= \frac{1}{NT}\sum_i \sum_t x^*_{it}u_{it} - \frac{1}{NT^2}\sum_i \sum_t \sum_r x^*_{it}u_{ir} \equiv B_{1NT1a} + B_{1NT1b}.
\end{aligned}
$$

Similarly,

$$
B_{1NT2} = \frac{1}{N^2T}\sum_{i,j}\sum_t x^*_{jt}u_{it} - \frac{1}{N^2T^2}\sum_{i,j}\sum_{t,s} x^*_{jt}u_{is} \equiv B_{1NT2a} + B_{1NT2b}.
$$

By Lemma A.2.2, $\|B_{1NTa}\|_2 = O_p(\frac{1}{\sqrt{NT}})$, $\|B_{1NTb}\|_2 = O_p(\frac{1}{T})$, $\|B_{2NTa}\|_2 = O_p(\frac{1}{\sqrt{N^2T}})$, and $\|B_{2NTb}\|_2 = O_p(\frac{1}{NT})$. Then (ii) follows.

(iii) Recall that $B_{2NT} = B_{2NT1} - B_{2NT2}$ by (A.2.1). Using $\tilde{x}_{it} = \tilde{x}^*_{it}$ again, we have following decompositions:

$$
\begin{aligned}
B_{2NT1} &= \frac{1}{NT}\sum_i \sum_t \tilde{x}^*_{it}\tilde{x}^{*\prime}_{it}\lambda_i \\
&= \frac{1}{NT}\sum_i \sum_t \left(x^*_{it} - \frac{1}{T}\sum_r x^*_{ir}\right)\left(x^*_{it} - \frac{1}{T}\sum_r x^*_{ir}\right)'\lambda_i \\
&= \frac{1}{NT}\sum_i \sum_t x^*_{it}x^{*\prime}_{it}\lambda_i - \frac{1}{NT^2}\sum_i \sum_{t,r} x^*_{it}x^{*\prime}_{ir}\lambda_i, \\
B_{2NT2} &= \frac{1}{NT}\sum_i \sum_t (\frac{1}{N}\sum_j \tilde{x}^*_{jt})\tilde{x}^{*\prime}_{it}\lambda_i \\
&= \frac{1}{N^2T}\sum_{i,j}\sum_t \left(x^*_{jt} - \frac{1}{T}\sum_r x^*_{jr}\right)\left(x^*_{it} - \frac{1}{T}\sum_r x^*_{ir}\right)'\lambda_i \\
&= \frac{1}{N^2T}\sum_{i,j}\sum_t x^*_{lt}x^{*\prime}_{it}\lambda_i - \frac{1}{N^2T^2}\sum_{i,j}\sum_{t,r} x^*_{jt}x^{*\prime}_{ir}\lambda_i.
\end{aligned}
$$

It follows that

$$\begin{aligned} B_{2NT} &= \frac{N-1}{N}\left(\frac{1}{NT}\sum_i\sum_t x_{it}^* x_{it}^{*\prime}\lambda_i - \frac{1}{NT^2}\sum_i\sum_{t,r} x_{it}^* x_{ir}^{*\prime}\lambda_i\right) \\ &\quad -\frac{1}{N^2T}\sum_{l\neq i}\sum_t x_{lt}^* x_{it}^{*\prime}\lambda_i + \frac{1}{N^2T^2}\sum_{l\neq i}\sum_{t,r} x_{lt}^* x_{ir}^{*\prime}\lambda_i \\ &\equiv \frac{N-1}{N}\left(B_{2NTa} - B_{2NTb}\right) - B_{2NTc} + B_{2NTd}. \end{aligned}$$

By Lemma A.2.3, $\|B_{2NTa}\| = O_p(\frac{1}{\sqrt{N}})$, $\|B_{2NTb}\| = O_p(\frac{1}{T})$, $\|B_{2NTc}\| = O_p(\frac{1}{N})$, $\|B_{2NTd}\| = O_p(\frac{1}{NT})$. Then $\|B_{2NT}\| = O_p(\frac{1}{\sqrt{N}})$.

(iv) Recall that $B_{3NT} = B_{3NT1} - B_{3NT2}$ by (A.2.1). Then, by direct calculations based on the fact $x_{it}^* = x_{it} - \mu$, and $\breve{x}_{it} = x_{it} - N^{-1}\sum_i x_{it}$, we have

$$\begin{aligned} B_{3NT} &= \frac{1}{NT}\sum_i\sum_t x_{it}^* x_{it}^*\gamma_t - \frac{1}{NT^2}\sum_i\sum_{t,r} x_{ir}^* x_{it}^*\gamma_t \\ &\quad -\frac{1}{N^2T}\sum_{i,l}\sum_t x_{lt}^* x_{it}^*\gamma_t + \frac{1}{N^2T^2}\sum_{i,l}\sum_{t,r} x_{lr}^* x_{it}^*\gamma_t \\ &= \frac{N-1}{N}\left(\frac{1}{NT}\sum_i\sum_t x_{it}^* x_{it}^*\gamma_t - \frac{1}{NT^2}\sum_i\sum_{t,r} x_{ir}^* x_{it}^*\gamma_t\right) \\ &\quad -\frac{1}{N^2T}\sum_{l\neq i}\sum_t x_{lt}^* x_{it}^*\gamma_t + \frac{1}{N^2T^2}\sum_{l\neq i}\sum_{t,r} x_{lr}^* x_{it}^*\gamma_t \\ &\equiv \frac{N-1}{N}\left(B_{3NTa} - B_{3NTb}\right) - B_{3NTc} + B_{3NTd}. \end{aligned}$$

By Lemma A.2.4, $\|B_{3NTa}\| = O_p(\frac{1}{\sqrt{T}})$, $\|B_{3NTb}\| = O_p(\frac{1}{T})$, $\|B_{3NTc}\| = O_p(\frac{1}{N})$, $\|B_{3NTd}\| = O_p(\frac{1}{NT})$. Then $\|B_{3NT}\| = O_p(\frac{1}{\sqrt{T}})$. ■

Proof of Lemma A.2.1.

(i) It is suffices to show $\left\|\frac{1}{NT}\sum_i\sum_t x_{it}^* x_{it}^{*\prime} - \Sigma_x\right\|_{\text{sp}} = o_p(1)$ with $\Sigma_x = E\left(x_{it}^* x_{it}^{*\prime}\right)$. Define

$w_{it} = z' x_{it}^*$, where $z \in \mathbb{R}^K$ such that $\|z\| = 1$, then

$$\begin{aligned} \left\| \frac{1}{NT} \sum_i \sum_t x_{it}^* x_{it}^{*\prime} - \Sigma_x \right\|_{\text{sp}} &= \left\| \frac{1}{NT} \sum_i \sum_t \left(w_{it}^2 - E w_{it}^2 \right) \right\| \\ &\equiv \left\| \frac{1}{NT} \sum_i \sum_t \psi_{it} \right\| . \end{aligned}$$

By construction, $\{\psi_{it}\}_{t=1}^T$ above are still mixing sequences with weak cross-sectional dependence under Assumption A1.1, we can readily obtain that $\frac{1}{NT} \sum_i \sum_t \psi_{it} = O_p \left(\frac{1}{\sqrt{NT}} \right)$. Therefore, the desired result follows provided that $m \leq \varphi_1 (\Sigma_x) \leq \varphi_K (\Sigma_x) \leq M$.

(ii) By direct calculations, we have

$$\begin{aligned} E \left\| \frac{1}{NT^2} \sum_i \sum_{t,s} x_{it}^* x_{is}^{*\prime} \right\|^2 &= \frac{1}{N^2 T^4} \sum_i \sum_{t,s,r,q} E \left(x_{iq}^{*\prime} x_{it}^* x_{is}^{*\prime} x_{ir}^* \right) + \frac{1}{N^2 T^4} \sum_{j \neq i} \sum_{t,s,r,q} E \left(x_{is}^{*\prime} x_{jr}^* x_{jq}^{*\prime} x_{it}^* \right) \\ &\equiv a_1 + a_2. \end{aligned}$$

For a_1, we have

$$a_1 = \frac{1}{N^2 T^4} \sum_i \sum_{t,s,r,q} E \left(x_{iq}^{*\prime} x_{it}^* x_{is}^{*\prime} x_{ir}^* \right) = \frac{1}{N^2 T^4} \sum_i E \left\| \sum_t x_{it}^* \right\|_2^4 = O \left(\frac{1}{NT^2} \right),$$

where the last equality holds because $E \left\| \sum_t x_{it}^* \right\|_2^4 = O(T^2)$ by Lemma A.3.1 under Assumption A1.1. Similarly,

$$\begin{aligned} a_2 &= \frac{1}{N^2 T^4} \sum_{j \neq i} \sum_{t,s,r,q} E \left(x_{is}^{*\prime} x_{jr}^* x_{jq}^{*\prime} x_{it}^* \right) = \frac{1}{N^2 T^4} \sum_{j \neq i} E \left[\left(\sum_s x_{is}^* \right)' \left(\sum_r x_{jr}^* \right) \right]^2 \\ &\leq \frac{1}{T^4} \left(\frac{1}{N} \sum_i \left[E \left\| \sum_s x_{is}^* \right\|^4 \right]^{1/2} \right) \left(\frac{1}{N} \sum_j \left[E \left\| \sum_r x_{jr}^* \right\|^4 \right]^{1/2} \right) = O \left(\frac{1}{T^2} \right), \end{aligned}$$

where inequality holds by Cauchy-Schwartz inequality, the final equality holds because $E \left\| \sum_t x_{it}^* \right\|_2^4 = O(T^2)$ for each i by Lemma A.3.1 under Assumption A1.1.

(iii) By direct calculations, we have

$$
\begin{aligned}
E\left\|\frac{1}{N^2 T}\sum_{j\neq i}\sum_{t} x_{it}^{*} x_{jt}^{*\prime}\right\|^2 &= \frac{1}{N^4 T^2}\sum_{j\neq i}\sum_{k\neq l}\sum_{t,s} E\left(x_{jt}^{*\prime} x_{ks}^{*} x_{ls}^{*\prime} x_{it}^{*}\right) \\
&\leq \frac{1}{N^2 T^2}\sum_{t,s}\left[E\left\|\frac{1}{\sqrt{N}}\sum_{i} x_{it}^{*}\right\|^4\right]^{1/2}\left[E\left\|\frac{1}{\sqrt{N}}\sum_{j} x_{js}^{*}\right\|^4\right]^{1/2} \\
&= \frac{1}{N^2 T^2} O\left(T^2\right) = O\left(\frac{1}{N^2}\right),
\end{aligned}
$$

where the inequality holds by Cauchy-Schwarz inequality, and the final bound holds under Assumption A1.1.

(iv) By direct calculations, we have

$$
\begin{aligned}
E\left\|\frac{1}{N^2 T^2}\sum_{j\neq i}\sum_{t,s} x_{it}^{*} x_{js}^{*\prime}\right\|^2 &= \frac{1}{N^4 T^4}\sum_{j_1\neq i_1}\sum_{j_2\neq i_2}\sum_{t,s,r,q} E\left(x_{j_1 s}^{*\prime} x_{i_2 r}^{*} x_{j_2 q}^{*\prime} x_{i_1 t}^{*}\right) \\
&\leq \frac{1}{N^2 T^4} E\left\|\sum_{t}\left(\frac{1}{\sqrt{N}}\sum_{i} x_{it}^{*}\right)\right\|_2^4 \\
&= \frac{1}{N^2 T^4} O\left(T^2\right) = O\left(\frac{1}{N^2 T^2}\right),
\end{aligned}
$$

where the last line follows because $E\left\|\sum_t\left(\frac{1}{\sqrt{N}}\sum_i x_{it}^{*}\right)\right\|_2^4 = O\left(T^2\right)$ by Lemma A.3.1 under Assumption A1.1. ■

Proof of Lemma A.2.2.

Note that $E(x_{it}^{*} u_{it}) = E(x_{it}^{*} u_{is}) = E(x_{lt}^{*} u_{it}) = E(x_{lt}^{*} u_{ir}) = 0$ under Assumption A1.2(iii).

(i) Let $\delta_* = \delta_1 \vee \delta_2$. Then, we have

$$
\begin{aligned}
& E\left\|\frac{1}{NT}\sum_i\sum_t x_{it}^* u_{it}\right\|^2 = \frac{1}{N^2T^2}\sum_{i,j}\sum_{t,s} E\left(u_{it}x_{it}^{*\prime}x_{js}^* u_{js}\right) \\
\leq & \frac{1}{N^2T^2}\left(\sum_{i,j}\sum_{t,s}\left[\alpha_{ij}(|t-s|)\right]^{\delta_*/(4+\delta_*)}\right)\left[E\left\|x_{it}^*\right\|^{(4+\delta_*)} E\left\|x_{js}^*\right\|^{(4+\delta_*)}\right]^{2/(4+\delta_*)} \\
& \left[E\left\|u_{it}^*\right\|^{(4+\delta_*)} E\left\|u_{js}^*\right\|^{(4+\delta_*)}\right]^{2/(4+\delta_*)} \\
= & O(1)\frac{1}{N^2T}\sum_{\tau=1}^{T}\sum_{i,j}\left[\alpha_{ij}(|\tau|)\right]^{\delta_*/(4+\delta_*)} = O\left(\frac{1}{NT}\right),
\end{aligned}
$$

where the second line holds by Davydov's inequality, the last lines above holds under Assumptions A1.1(i) and A1.2(i).

(ii) By moments calculations, we have,

$$
\begin{aligned}
& E\left\|\frac{1}{NT^2}\sum_i\sum_{t,s} x_{it}^* u_{is}\right\|^2 \\
\leq & \frac{1}{N^2T^2}\sum_{i,j} E\left[\left(\frac{1}{\sqrt{T}}\sum_s u_{is}\right)\left(\frac{1}{\sqrt{T}}\sum_t x_{it}^*\right)'\right]\left[\left(\frac{1}{\sqrt{T}}\sum_t x_{jr}^*\right)\left(\frac{1}{\sqrt{T}}\sum_q u_{jq}\right)\right] \\
\lesssim & \frac{1}{N^2T^2}\sum_{i,j}\alpha(0)^{\delta_*/(4+\delta_*)} = O\left(\frac{1}{NT^2}\right),
\end{aligned}
$$

where the second line holds because $\{T^{-1}\sum_{t,s} u_{is}\}$ are still α-mixing sequences with zero mean under Assumptions A1.1 and A1.2, then use Davydov's inequality and Lemma A.3.1 under Assumption A1.1(i) and A1.2(i) sequentially; and the final bound holds under Assumption A1.1(i) and A1.2(i).

(iii)–(iv) The proof is analogous to that of (ii) and thus omitted. ■

Proof of Lemma A.2.3.

(i) Note that $E\left(\frac{1}{NT}\sum_i\sum_t x_{it}^* x_{it}^{*\prime}\lambda_i\right) = 0$ under Assumption A1.1, then,

$$
\begin{aligned}
E\left\|\frac{1}{NT}\sum_i\sum_t x_{it}^* x_{it}^{*\prime}\lambda_i\right\|^2 &= E\left\|\frac{1}{NT}\sum_i\sum_t \left(x_{it}^* x_{it}^{*\prime} - \Sigma_x\right)\lambda_i + \Sigma_x\lambda_i\right\|^2 \\
&\leq \frac{1}{N^2T^2}\sum_{i,j} E\left(\lambda_i'\lambda_j\right)\left\|\sum_t \Sigma_x\right\|^2 \\
&\quad + \frac{1}{N^2T^2}\sum_{i,j}\sum_{t,s} E\lambda_i'\left(x_{it}^* x_{it}^{*\prime} - \Sigma_x\right)\left(x_{js}^* x_{js}^{*\prime} - \Sigma_x\right)\lambda_j \\
&\equiv a_3 + a_4 = \frac{1}{N^2T^2}O\left(NT^2\right) + \frac{1}{N^2T^2}O\left(NT\right) = O\left(\frac{1}{N}\right) + O\left(\frac{1}{NT}\right),
\end{aligned}
$$

where the second line follows by triangular inequality. Furthermore, $a_3 = O\left(1/N\right)$ because we have $\text{Var}\left(N^{-1/2}\sum_i \lambda_i\right) = O\left(1\right)$ under Assumption A1.1(iii), $\left\|\Sigma_x\right\|^2 < \infty$ under Assumption A1.1. For a_4 above, let $w_{it} = \lambda_i'\left(x_{it}^* x_{it}^{*\prime} - \Sigma_x\right)$, then, $a_4 = N^{-2}T^{-2}\sum_{i,j}\sum_{t,s} w_{it}w_{js}$ and $\{w_{it}\}$ are still mixing sequences that satisfy the Assumption A1.1(i), it follows that $a_4 = O(1/NT)$.

(ii) Note that $E\left(\frac{1}{NT^2}\sum_i\sum_t\sum_s x_{it}^* x_{is}^{*\prime}\lambda_i\right) = 0$ and $E\left(x_{it}^* x_{is}^{*\prime}\lambda_i\right) = 0$ under Assumption A1.1, then,

$$
\begin{aligned}
E\left\|\frac{1}{NT^2}\sum_i\sum_{t,s} x_{it}^* x_{is}^{*\prime}\lambda_i\right\|^2 &= \frac{1}{N^2T^2}\sum_{i,j} E\left(\frac{1}{T}\sum_{t,s}\lambda_i' x_{is}^* x_{it}^*\right)\left(\frac{1}{T}\sum_{r,q} x_{jr}^{*\prime} x_{jq}^*\lambda_j\right) \\
&\leq O\left(1\right)\frac{1}{N^2T^2}\sum_{i,j}\alpha_{ij}(0)^{\delta_1/(4+\delta_1)} \\
&= \frac{1}{N^2T^2}O\left(N\right) = O\left(\frac{1}{NT^2}\right),
\end{aligned}
$$

where the second line holds because $\{T^{-1}\sum_{t,s}\lambda_i' x_{is}^* x_{it}^*\}$ are still α-mixing sequences with zero mean under Assumption A1.1, then use Davydov's inequality and Lemma A.3.1 under Assumption A1.1 sequentially; and the final bound holds under Assumption A1.1(i).

(iii) Note that $E\left(\frac{1}{N^2T}\sum_{l\neq i}\sum_t x_{lt}^* x_{it}^{*\prime}\lambda_i\right) = 0$ and $E\left(x_{lt}^* x_{it}^{*\prime}\lambda_i\right) = 0$ under Assumption A1.1, then,

$$
\begin{aligned}
E\left\|\frac{1}{N^2T}\sum_{l\neq i}\sum_{t}x_{lt}^{*}x_{it}^{*\prime}\lambda_i\right\|^2 &= \frac{1}{N^2T^2}\sum_{t,s}E\left(\frac{1}{N}\sum_{l\neq i}\lambda_i^{\prime}x_{it}^{*}x_{lt}^{*}\right)\left(\frac{1}{N}\sum_{j\neq k}x_{js}^{*\prime}x_{ks}^{*}\lambda_k\right) \\
&\leq O(1)\frac{1}{N^2T^2}\sum_{t,s}\alpha(|t-s|))^{\delta_1/(4+\delta_1)} \\
&= \frac{1}{N^2T^2}O(T)=O\left(\frac{1}{N^2T}\right),
\end{aligned}
$$

where the second line holds because $\{N^{-1}\sum_i\sum_{l\neq i}\lambda_i^{\prime}x_{it}^{*}x_{lt}^{*}\}$ are still α-mixing sequences with zero mean under Assumption A1.1, then use Davydov's inequality and Lemma A.3.1 under Assumption A1.1 sequentially; and the final bound holds under Assumption A1.1(i).

(iv) Note that $E\left(\frac{1}{N^2T^2}\sum_{l\neq i}\sum_{t,s}x_{lt}^{*}x_{is}^{*\prime}\lambda_i\right)=0$ under Assumption A1.1, then,

$$
\begin{aligned}
& E\left\|\frac{1}{N^2T^2}\sum_{l\neq i}\sum_{t,s}x_{lt}^{*}x_{is}^{*\prime}\lambda_i\right\|^2 \\
=\ & \frac{1}{N^4T^2}\sum_{i,j}\sum_{t,s,r,q}E\left[\lambda_i^{\prime}x_{is}^{*}\left(\sum_{l\neq i}x_{lt}^{*\prime}\right)\left(\sum_{k\neq j}x_{kr}^{*}\right)x_{jq}^{*\prime}\lambda_j\right] \\
\leq\ & \frac{1}{N^4T^4}E\left(\left\|\sum_{l\neq i}\sum_{t}x_{lt}^{*}\right\|\left\|\sum_{k\neq j}\sum_{r}x_{kr}^{*}\right\|\left\|\sum_{i}\sum_{s}x_{is}^{*\prime}\lambda_i\right\|\left\|\sum_{j}\sum_{q}x_{jq}^{*\prime}\lambda_j\right\|\right) \\
\leq\ & \frac{1}{N^2T^4}\left[E\left\|\sum_{t}\left(\frac{1}{\sqrt{N}}\sum_{j\neq i}x_{jt}^{*}\right)\right\|^4\right]^{1/2}\left[E\left\|\frac{1}{\sqrt{N}}\sum_{i}\sum_{s}x_{is}^{*\prime}\lambda_i\right\|^4\right]^{1/2} \\
=\ & \frac{1}{N^2T^4}O(T)\,O(T)=O\left(\frac{1}{N^2T^2}\right),
\end{aligned}
$$

where the fourth line follows by repeating uses of Cauchy-Schwarz inequality, and the final bound holds by Lemma A.3.1 under Assumption A1.1(i). ■

Proof of Lemma A.2.4

(i) Note that $E\left(\frac{1}{NT}\sum_i\sum_t x_{it}^{*}x_{it}^{*\prime}\gamma_t\right)=0$ under Assumption A1.1. By direct calculations, we have

$$
\begin{aligned}
& E\left\|\frac{1}{NT}\sum_i\sum_t x_{it}^* x_{it}^{*\prime}\gamma_t\right\|^2 \\
= & \frac{1}{N^2T^2}\sum_{i,j}\sum_{t,s} E\left(\gamma_t' x_{jt}^* x_{jt}^{*\prime} x_{is}^* x_{is}^{*\prime}\gamma_s\right) \\
= & \frac{1}{N^2T^2}\sum_{i,j}\sum_{t,s} E\left[\gamma_t'\left(x_{jt}^* x_{jt}^{*\prime}-\Sigma_x\right)\left(x_{is}^* x_{is}^{*\prime}-\Sigma_x\right)\gamma_s+\gamma_t'\Sigma_x\Sigma_x\gamma_s\right] \\
\leq & \frac{1}{N^2T^2}\sum_{i,j}\sum_{t,s}\left\|E\left[\gamma_t'\left(x_{jt}^* x_{jt}^{*\prime}-\Sigma_x\right)\left(x_{is}^* x_{is}^{*\prime}-\Sigma_x\right)\gamma_s\right]\right\|+\frac{1}{N^2T^2}\sum_{i,j}\sum_{t,s} E\left(\gamma_t'\Sigma_x\Sigma_x\gamma_s\right) \\
\leq & \frac{1}{N^2T^2}\sum_{i,j}\sum_{t,s}\left\|E\gamma_t'\left(x_{jt}^* x_{jt}^{*\prime}-\Sigma_x\right)\left(x_{is}^* x_{is}^{*\prime}-\Sigma_x\right)\gamma_s\right\|+\frac{1}{N^2T^2}\sum_{i,j}\sum_{t,s} E\gamma_t'\gamma_s\left\|\Sigma_x\right\|^2 \\
= & \frac{1}{N^2T^2}O\left(NT+N^2T\right)=O\left(\frac{1}{NT}\right)+O\left(\frac{1}{T}\right),
\end{aligned}
$$

where the fourth line above holds because $\|\Sigma_x\|_{op}<\infty$ under Assumption A1.1(i) and $\operatorname{Var}\left(T^{-1/2}\sum_t\gamma_t\right)=O(1)$ under Assumption A1.1(iv), besides, by construction, $\left\{\gamma_t'\left(x_{jt}^* x_{jt}^{*\prime}-\Sigma_x\right)\right\}$ are still mixing sequences under Assumption A1.1(i), and thus,

$$
\sum_{t,s}\sum_i\sum_j\left\|E\gamma_t'\left(x_{jt}^* x_{jt}^{*\prime}-\Sigma_x\right)\left(x_{it}^* x_{it}^{*\prime}-\Sigma_x\right)\gamma_s\right\|_2\leq O(1)\sum_t\sum_i\sum_j(\alpha_{ij}(|t-s|))^{\delta_1/(4+\delta_1)}=O(NT).
$$

holds directly under Assumption A1.1(i).

(ii) Note that $E\left(\frac{1}{NT^2}\sum_i\sum_{t,r}x_{ir}^* x_{it}^{*\prime}\gamma_t\right)=0$ and $E\left(x_{ir}^* x_{it}^{*\prime}\gamma_t\right)=0$ under Assumption A1.1, then,

$$
\begin{aligned}
E\left\|\frac{1}{NT^2}\sum_i\sum_{t,r}x_{ir}^* x_{it}^{*\prime}\gamma_t\right\|_2^2 &= \frac{1}{N^2T^2}\sum_{i,j}E\left(\frac{1}{T}\sum_{t,r}\gamma_t' x_{it}^* x_{ir}^*\right)\left(\frac{1}{T}\sum_{s,q}x_{jq}^{*\prime}x_{js}^*\gamma_s\right) \\
&\leq O(1)\frac{1}{N^2T^2}\sum_{i,j}\alpha_{ij}(0)^{\delta_1/(4+\delta_1)} \\
&= \frac{1}{N^2T^2}O(N)=O\left(\frac{1}{NT^2}\right),
\end{aligned}
$$

where the second line holds because $\{T^{-1}\sum_{t,r}\gamma_t' x_{it}^* x_{ir}^*\}$ are still α-mixing sequences with zero mean under Assumption A1.1, then use Davydov's inequality and Lemma A.3.1 under Assumption A1.1

sequentially; and the final bound holds under Assumption A1.1(i).

(iii) Note that $E\left(\frac{1}{N^2T}\sum_{l\neq i}\sum_t x_{lt}^* x_{it}^{*\prime}\gamma_t\right) = 0$ and $E\left(x_{lt}^* x_{it}^{*\prime}\gamma_t\right) = 0$ under Assumption A1.1, then,

$$
\begin{aligned}
E\left\|\frac{1}{N^2T}\sum_{l\neq i}\sum_t x_{lt}^* x_{it}^{*\prime}\gamma_t\right\|_2^2 &= \frac{1}{N^2T^2}\sum_{t,s}E\left(\frac{1}{N}\sum_{l\neq i}\gamma_t' x_{it}^* x_{lt}^*\right)\left(\frac{1}{N}\sum_{j\neq k}x_{js}^{*\prime}x_{ks}^*\gamma_s\right) \\
&\leq \frac{1}{N^2T^2}\sum_{t,s}\alpha(|t-s|)^{\delta_1/(4+\delta_1)} \\
&= \frac{1}{N^2T^2}O\left(T\right) = O\left(\frac{1}{N^2T}\right),
\end{aligned}
$$

where the second line holds because $\{N^{-1}\sum_i\sum_{l\neq i}\gamma_t' x_{it}^* x_{lt}^*\}$ are still α-mixing sequences with zero mean under Assumption A1.1, then use Davydov's inequality and Lemma A.3.1 under Assumption A1.1 sequentially; and the final bound holds under Assumption A1.1(i).

(iv) The desired bounds can be obtained by following similar arguments for the proofs of (i)–(iii) above.

A.2.2 Proof of Proposition A.1.2: Stochastic Bound of $\widehat{\lambda}_i - \lambda_i$

Recall that

$$
\begin{aligned}
\widehat{\lambda}_i - \lambda_i &= \frac{N}{N-1}A_{2iT}^{-1}\left(-C_{1iT} + C_{2iT} + C_{3iT} + C_{4iT}\right) \\
&= O_p\left(1\right)\left[O_p\left(\left\|\widehat{\beta}-\beta\right\|_2\right) + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right)\right],
\end{aligned}
$$

where A_{2iT}, C_{1iT}, C_{2iT}, C_{3iT} and C_{4iT} are defined in Section A.3.1:

$$
\begin{aligned}
A_{2iT} &\equiv \frac{1}{T}\sum_t \tilde{x}_{it}\tilde{x}_{it}', \\
C_{1iT} &\equiv \frac{1}{T}\sum_t \tilde{x}_{it}\left(\tilde{x}_{it}-\frac{1}{N}\tilde{x}_{lt}\right)'\left(\widehat{\beta}-\beta\right) = \frac{1}{T}\sum_t \tilde{x}_{it}\tilde{x}_{it}'\left(\widehat{\beta}-\beta\right) - \frac{1}{NT}\sum_l\sum_t \tilde{x}_{it}\tilde{x}_{lt}'\left(\widehat{\beta}-\beta\right) \\
&\equiv C_{1iT1} - C_{1iT2}, \\
C_{2iT} &\equiv \frac{1}{T}\sum_t \tilde{x}_{it}\left(\breve{x}_{it}'\gamma_t - \frac{1}{T}\sum_s \breve{x}_{is}'\gamma_s\right) = \frac{1}{T}\sum_t \tilde{x}_{it}\breve{x}_{it}'\gamma_t - \frac{1}{T^2}\sum_{t,s}\tilde{x}_{it}\breve{x}_{is}'\gamma_s \equiv C_{2iT1} - C_{2iT2}, \\
C_{3iT} &\equiv \frac{1}{T}\sum_t \tilde{x}_{it}\left(\tilde{u}_{it} - \frac{1}{N}\sum_l \tilde{u}_{lt}\right) = \frac{1}{T}\sum_t \tilde{x}_{it}\tilde{u}_{it} - \frac{1}{NT}\sum_l\sum_t \tilde{x}_{it}\tilde{u}_{lt} \equiv C_{3iT1} - C_{3iT2}, \\
C_{4iT} &\equiv \frac{1}{NT}\sum_{l=1,l\neq i}^{N}\sum_t \tilde{x}_{it}\tilde{x}_{lt}'\lambda_l.
\end{aligned}
$$

To prove Proposition A.1.2, we need following four lemmas.

Lemma A.2.5. *Suppose Assumption A1.1 and A1.4 hold. Then for each i, we have*

(i) $\left\|\frac{1}{T}\sum_t x_{it}^* x_{it}^{*\prime}\right\|_F = O_p(1)$,

(ii) $\left\|\frac{1}{T^2}\sum_t\sum_s x_{it}^* x_{is}^{*\prime}\right\| = O_p(\frac{1}{T})$,

(iii) $\left\|\frac{1}{NT}\sum_{l\neq i}\sum_t x_{it}^* x_{lt}^{*\prime}\right\| = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$,

(iv) $\left\|\frac{1}{NT^2}\sum_{l\neq i}\sum_t\sum_s x_{it}^* x_{ls}^{*\prime}\right\| = O_p(\frac{1}{\sqrt{NT^2}})$.

Lemma A.2.6. *Under Assumptions A1.1, the following hold:*

(i) $\left\|\frac{1}{T}\sum_t x_{it}^* x_{it}^{*\prime}\gamma_t\right\| = O_p(\frac{1}{\sqrt{T}})$,

(ii) $\left\|\frac{1}{T^2}\sum_t\sum_r x_{ir}^* x_{it}^{*\prime}\gamma_t\right\| = O_p(\frac{1}{T})$,

(iii) $\left\|\frac{1}{NT}\sum_{l\neq i}\sum_t x_{it}^* x_{lt}^{*\prime}\gamma_t\right\| = O_p(\frac{1}{\sqrt{N^2T}})$,

(iv) $\left\|\frac{1}{NT^2}\sum_{l\neq i}\sum_t\sum_r x_{ir}^* x_{lt}^{*\prime}\gamma_t\right\| = O_p\left(\frac{1}{\sqrt{NT^2}}\right)$.

Lemma A.2.7. *Under Assumption A1.1 and A1.2,*

(i) $\left\|\frac{1}{T}\sum_t x_{it}^* u_{it}\right\| = O_p(\frac{1}{\sqrt{T}})$,

(ii) $\left\|\frac{1}{T^2}\sum_t\sum_r x_{it}^* u_{ir}\right\| = O_p(\frac{1}{T})$,

(iii) $\left\|\frac{1}{NT}\sum_l\sum_t x_{it}^* u_{lt}\right\| = O_p(\frac{1}{\sqrt{NT}})$,

(iv) $\left\| \frac{1}{NT^2} \sum_l \sum_t \sum_r x^*_{it} u_{lr} \right\| = O_p(\frac{1}{\sqrt{NT^2}})$.

Lemma A.2.8. *Under Assumption A1.1, the following hold:*

(i) $\left\| \frac{1}{NT} \sum_{l \neq i} \sum_t x^*_{it} x^{*\prime}_{lt} \lambda_l \right\| = O_p(\frac{1}{\sqrt{NT}})$,

(ii) $\left\| \frac{1}{NT^2} \sum_{l \neq i} \sum_t \sum_r x^*_{it} x^{*\prime}_{lr} \lambda_l \right\| = O_p(\frac{1}{\sqrt{NT^2}})$.

Proof of Proposition A.1.2.

(i) Note that $\tilde{x}_{it} = x_{it} - \frac{1}{T}\sum_r x_{ir}$, $\tilde{x}^*_{it} = x^*_{it} - \frac{1}{T}\sum_r x^*_{ir}$ with $x^*_{it} = x_{it} - \mu$, then, $\tilde{x}_{it} = \tilde{x}^*_{it}$, it follows that:

$$A_{2iT} = \frac{1}{T}\sum_t \tilde{x}_{it}\tilde{x}'_{it} = \frac{1}{T}\sum_t \tilde{x}^*_{it}\tilde{x}^{*\prime}_{it} = \frac{1}{T}\sum_t x^*_{it}x^{*\prime}_{it} - \frac{1}{T^2}\sum_{r_1}\sum_{r_2} x^*_{ir_1}x^{*\prime}_{ir_2} \equiv A_{2iT1} - A_{2iT2}.$$

Then, by law of large number, as $T \to \infty$, we have $A_{2T} \xrightarrow{p} \lim_{T\to\infty} \frac{1}{T}\sum_t E\left(x^*_{it}x^{*\prime}_{it}\right) = \Sigma_x$ by By Lemma A.2.5.

(ii) Note that $\tilde{x} = x_{it} - \frac{1}{T}\sum_r x_{ir}$, $\tilde{x}^* = x^*_{it} - \frac{1}{T}\sum_r x^*_{ir}$ with $x^*_{it} = x_{it} - \mu$, then, $\tilde{x}_{it} = \tilde{x}^*_{it}$, it follows that:

$$C_{1iT} = \left[\frac{N-1}{N}\left(\underbrace{\frac{1}{T}\sum_t x^*_{it}x^{*\prime}_{it}}_{C_{1iT1}} - \underbrace{\frac{1}{T^2}\sum_t\sum_r x^*_{it}x^{*\prime}_{ir}}_{C_{1iT2}} \right) - \underbrace{\frac{1}{NT}\sum_{l\neq i}\sum_t x^*_{it}x^{*\prime}_{lt}}_{C1iT3} + \underbrace{\frac{1}{NT^2}\sum_{l\neq i}\sum_t\sum_s x^*_{it}x^{*\prime}_{ls}}_{C1iT4} \right] \times \left(\widehat{\beta} - \beta\right),$$

It is clear that the order of C_{1iT} is determined by both $\{C_{1iTk}\}_{k=1}^4$ and $\left(\widehat{\beta} - \beta\right)$.

By Lemma A.2.5, $\|C_{1iT1}\|_2 = O_p(1)$, $\|C_{1iT2}\|_2 = O_p\left(\frac{1}{T}\right)$, $\|C_{1iT3}\|_2 = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{N}\right)$, $\|C_{1iT4}\|_2 = O_p\left(\frac{1}{\sqrt{NT^2}}\right)$. Thus, the order of $\left(\widehat{\beta} - \beta\right)$ will determine the order of C_{1T}.

For the sake of this fact, we denote the bound as: $\|C_{1iT}\|_2 = O_p\left(\left\|\widehat{\beta} - \beta\right\|_2\right)$, which means the order of C_{1iT} is totally determined by the order of $\left\|\widehat{\beta} - \beta\right\|_2$.

(iii) Recall that $C_{2iT} = C_{2iT1} - C_{2iT2} = \frac{1}{T}\sum_t \tilde{x}_{it}\breve{x}'_{it}\gamma_t - \frac{1}{T}\sum_s \left(\frac{1}{T}\sum_t \tilde{x}_{it}\right)\breve{x}'_{is}\gamma_s$. Note that

$\frac{1}{T}\sum_t \tilde{x}_{it} = 0$, $\frac{1}{T}\sum_t \tilde{x}^*_{it} = 0$. Then $C_{2iT2} = 0$. Thus, $C_{2T} = C_{2iT1}$, and we bound C_{2iT1} below,

$$\begin{aligned} C_{2iT1} &= \frac{1}{T}\sum_t \left(x^*_{it} - \frac{1}{T}\sum_r x^*_{ir} \right)\left(x^*_{it} - \frac{1}{N}\sum_l x^*_{lt} \right)\gamma_t \\ &= \frac{N-1}{N}\left[\frac{1}{T}\sum_t x^*_{it}x^{*\prime}_{it}\gamma_t - \frac{1}{T^2}\sum_{t,r} x^*_{ir}x^{*\prime}_{it}\gamma_t \right] - \frac{1}{NT}\sum_{l=1,l\neq i}^{N}\sum_t x^*_{it}x^{*\prime}_{lt}\gamma_t + \frac{1}{NT^2}\sum_{l=1,l\neq i}^{N}\sum_{t,r} x^*_{ir}x^{*\prime}_{lt}\gamma_t \\ &\equiv \frac{N-1}{N}\left(C_{2iTa} - C_{2iTb}\right) - C_{2iTc} - C_{2iTd}. \end{aligned}$$

By Lemma A.2.6, $\|C_{2Ta}\|_2 = O_p\left(\frac{1}{\sqrt{T}}\right)$, $\|C_{2iTb}\|_2 = O_p\left(\frac{1}{T}\right)$, $\|C_{2iTc}\|_2 = O_p\left(\frac{1}{\sqrt{N^2T}}\right)$, $\|C_{2iTd}\|_2 = O_p\left(\frac{1}{\sqrt{NT^2}}\right)$. Hence, it follows that $\|C_{2iT}\|_2 = O_p\left(\frac{1}{\sqrt{T}}\right)$.

(iv) Recall that $\tilde{x}_{it} = \tilde{x}^*_{it}$, $\tilde{x} = x_{it} - \frac{1}{T}\sum_r x_{ir}$, $\tilde{x}^* = x^*_{it} - \frac{1}{T}\sum_r x^*_{ir}$, and $x^*_{it} = x_{it} - \mu$. Note that,

$$\begin{aligned} C_{3iT} &= \frac{1}{T}\sum_t \tilde{x}_{it}\tilde{u}_{it} - \frac{1}{NT}\sum_l\sum_t \tilde{x}_{it}\tilde{u}_{lt} \\ &= \frac{1}{T}\sum_t x^*_{it}u_{it} - \frac{1}{T^2}\sum_t\sum_r x^*_{it}u_{ir} - \frac{1}{NT}\sum_l\sum_t x^*_{it}u_{lt} + \frac{1}{NT^2}\sum_l\sum_{t,r} x^*_{it}u_{lr} \\ &= C_{3iTa} - C_{3iTb} - C_{3iTc} + C_{3iTd}. \end{aligned}$$

By Lemma A.2.7, $\|C_{3iTa}\|_2 = O_p\left(\frac{1}{\sqrt{T}}\right)$, $\|C_{3iTb}\|_2 = O_p\left(\frac{1}{T}\right)$, $\|C_{3iTc}\|_2 = O_p\left(\frac{1}{\sqrt{NT}}\right)$, $\|C_{3iTd}\|_2 = O_p\left(\frac{1}{\sqrt{NT^2}}\right)$. Then, the desired bound holds.

(v) Recall that $\tilde{x}_{it} = \tilde{x}^*_{it}$, $\tilde{x} = x_{it} - \frac{1}{T}\sum_r x_{ir}$, $\tilde{x}^* = x^*_{it} - \frac{1}{T}\sum_r x^*_{ir}$, $x^*_{it} = x_{it} - \mu$. By direct calculations, we have

$$\begin{aligned} C_{4iT} &= \frac{1}{NT}\sum_{l=1,l\neq i}^{N}\sum_t \tilde{x}_{it}\tilde{x}^{\prime}_{lt}\lambda_l = \frac{1}{NT}\sum_{l=1,l\neq i}^{N}\sum_t \tilde{x}^*_{it}\tilde{x}^{*\prime}_{lt}\lambda_l \\ &= \frac{1}{NT}\sum_{l=1,l\neq i}^{N}\sum_t x^*_{it}x^{*\prime}_{lt}\lambda_l - \frac{1}{NT^2}\sum_{l=1,l\neq i}^{N}\sum_{t,r} x^*_{it}x^{*\prime}_{lr}\lambda_l \equiv C_{4iT1} - C_{4iT2}. \end{aligned}$$

By Lemma A.2.8, $\|C_{4iT1}\|_2 = O_p\left(\frac{1}{\sqrt{NT}}\right)$, $\|C_{4iTb}\|_2 = O_p\left(\frac{1}{\sqrt{NT^2}}\right)$. It follows that $\|C_{4iT2}\|_2 = O_p\left(\frac{1}{\sqrt{NT}}\right)$ ■

Proof of Lemma A.2.5.

(i) Following arguments of proofs for Lemma A.2.1(1), $\frac{1}{T}\sum_t x_{it}^* x_{it}^{*\prime} \xrightarrow{P} \lim_{T\to\infty} \frac{1}{T}\sum_t E\left(x_{it}^* x_{it}^{*\prime}\right) = \Sigma_x$ holds under Assumption A1.1, $\frac{1}{T}\sum_t x_{it}^* x_{it}^{*\prime} = O_p(1)$ follows directly Assumption A1.4.

(ii) By Cauchy-Schwarz inequality

$$E\left\|\frac{1}{T^2}\sum_{t,s} x_{it}^* x_{is}^{*\prime}\right\|^2 \leq \frac{1}{T^4} E\left\|\sum_t x_{it}^*\right\|^4 = \frac{1}{T^4} O\left(T^2\right) = O\left(\frac{1}{T^2}\right),$$

where the last line holds because $E\left\|\sum_t x_{it}^*\right\|^4 = O\left(T^2\right)$ by Lemma A.3.1 under Assumption A1.1.

(iii) Denote $\tau_{il} = E\left(x_{it}^* x_{lt}^{*\prime}\right)$ temporarily. Then

$$\begin{aligned}
& E\left\|\frac{1}{NT}\sum_{l=1,l\neq i}^{N}\sum_t x_{it}^* x_{lt}^{*\prime}\right\|^2 \\
\leq\ & 2E\left\|\frac{1}{NT}\sum_{l=1,l\neq i}^{N}\sum_t \left(x_{it}^* x_{lt}^{*\prime} - \tau_{il}\right)\right\|^2 + 2\left\|\frac{1}{NT}\sum_{l=1,l\neq i}^{N}\sum_t \tau_{il}\right\|^2 \\
=\ & \frac{2}{N^2T^2}\sum_{l=1,l\neq i}^{N}\sum_{k=1,k\neq i}^{N}\sum_{t,s} \mathrm{tr}\left(E\left[\left(x_{it}^* x_{lt}^{*\prime} - \tau_{il}\right)\left(x_{is}^* x_{ks}^{*\prime} - \tau_{ik}\right)\right]\right) + \frac{2}{N^2T^2}\left\|\sum_{l=1,l\neq i}^{N}\sum_t \tau_{il}\right\|^2 \\
\lesssim\ & \frac{1}{N^2T^2}\sum_{l=1,l\neq i}^{N}\sum_{k=1,k\neq i}^{N}\sum_{t,s}\left[\alpha_{kl}\left(|t-s|\right)\right]^{\delta_1/(4+\delta_1)} + \frac{1}{N^2T^2}\left[\sum_t \left[\alpha_{il}\left(0\right)\right]^{\delta_1/(4+\delta_1)}\right]^2 \\
=\ & \frac{1}{N^2T^2} O\left(NT + T^2\right) = O\left(\frac{1}{NT} + \frac{1}{N^2}\right),
\end{aligned}$$

where we use the fact that $\left\|\sum_t \sum_{k=1,k\neq i}^{N} \tau_{ik}\right\| \leq \sum_t \sum_{k=1,k\neq i}^{N} O(1)\alpha_{ik}\left(0\right)^{\delta_1/(4+\delta_1)} = O\left(T\right)$ under Assumption A1.1, meanwhile, by similar arguments for (28) in Gao and Hong (2008) and the covariance inequality for the mixing sequences, we can show that

$$\begin{aligned}
\sum_{l=1,l\neq i}^{N}\sum_{k=1,k\neq i}^{N}\sum_t\sum_s \left\|E\left(x_{it}^* x_{lt}^{*\prime} - \tau_{il}\right)\left(x_{is}^* x_{ks}^{*\prime} - \tau_{ik}\right)\right\| &\lesssim \sum_{l=1,l\neq i}^{N}\sum_{k=1,k\neq i}^{N}\sum_t\sum_s \alpha_{kl}\left(|t-s|\right)^{\delta_1/(4+\delta_1)} \\
&= O\left(NT\right)
\end{aligned}$$

under Assumption A1.1. Note that the bound obtained here under Assumption A1.1 is similar to Assumption A1(iv) in Li et al. (2016), which demonstrate the weak form of cross-sectional dependence and serial dependence among the sequences.

(iv) By Cauchy-Schwarz inequality,

$$
\begin{aligned}
E\left\|\frac{1}{NT^2}\sum_{l\neq i}\sum_{t,s}x_{it}^{*}x_{ls}^{*\prime}\right\| &\leq \frac{1}{N^{1/2}T^2}\left[E\left\|\sum_{s}\left(\frac{1}{\sqrt{N}}\sum_{l}x_{ls}^{*}\right)\right\|^2\right]^{1/2}\left[E\left\|\sum_{t}x_{it}^{*}\right\|^2\right]^{1/2} \\
&= \frac{1}{N^{1/2}T^2}O\left(T^{1/2}\right)O\left(T^{1/2}\right)=O\left(\frac{1}{N^{1/2}T}\right),
\end{aligned}
$$

where the last line follows by Lemma A.3.1 under Assumption A1.1. ■

Proof of Lemma A.2.6.

(i) Note that $E\left(\frac{1}{T}\sum_t x_{it}^{*}x_{it}^{*\prime}\gamma_t\right)=0$ under Assumption A1.1, then, by direct moments calculations, we have

$$
\begin{aligned}
E\left\|\frac{1}{T}\sum_{t}x_{it}^{*}x_{it}^{*\prime}\gamma_t\right\|^2 &= \frac{1}{T^2}\sum_{t,s}E\left(\gamma_t^{\prime}x_{it}^{*}x_{it}^{*\prime}x_{is}^{*}x_{is}^{*\prime}\gamma_s\right) \\
&= \frac{1}{T^2}\sum_{t,s}E\left[\gamma_t^{\prime}\left(x_{it}^{*}x_{it}^{*\prime}-\Sigma_x\right)\left(x_{is}^{*}x_{is}^{*\prime}-\Sigma_x\right)\gamma_s+\gamma_t^{\prime}\Sigma_x\Sigma_x\gamma_s\right] \\
&\leq \frac{1}{T^2}\sum_{t,s}\left\|E\left[\gamma_t^{\prime}\left(x_{it}^{*}x_{it}^{*\prime}-\Sigma_x\right)\left(x_{is}^{*}x_{is}^{*\prime}-\Sigma_x\right)\gamma_s\right]\right\|_2+\frac{1}{T^2}\sum_{t,s}E\left(\gamma_t^{\prime}\Sigma_x\Sigma_x\gamma_s\right) \\
&\leq \frac{1}{T^2}\sum_{t,s}\left\|E\gamma_t^{\prime}\left(x_{it}^{*}x_{it}^{*\prime}-\Sigma_x\right)\left(x_{is}^{*}x_{is}^{*\prime}-\Sigma_x\right)\gamma_s\right\|_2+\frac{1}{T^2}\sum_{t,s}E\left(\gamma_t^{\prime}\gamma_s\right)\left\|\Sigma_x\right\|_{\text{sp}}^2 \\
&= \frac{1}{T^2}\left[O\left(T\right)+O\left(T\right)\right]=O\left(\frac{1}{T}\right),
\end{aligned}
$$

where the fourth line above holds by the fact that $\text{Var}\left(T^{-1/2}\sum_t\gamma_t\right)=O\left(1\right)$ under Assumption A1.1(iv) and that $\{\gamma_t^{\prime}\left(x_{it}^{*}x_{it}^{*\prime}-\Sigma_x\right)\}$ are still mixing sequences under Assumption A1.1(i), and thus,

$$
\sum_{t,s}\left\|E\gamma_t^{\prime}\left(x_{jt}^{*}x_{jt}^{*\prime}-\Sigma_x\right)\left(x_{it}^{*}x_{it}^{*\prime}-\Sigma_x\right)\gamma_s\right\|\lesssim T\sum_{\tau=1}^{\infty}\alpha(\tau)^{\delta_1/(4+\delta_1)}=O(T).
$$

holds directly under Assumption A1.1(i).

(ii) Note that $E\left(\frac{1}{T^2}\sum_t\sum_r x_{ir}^* x_{it}^{*\prime}\gamma_t\right) = 0$ under Assumption A1.1, then, by direct moments calculations, we have

$$
\begin{aligned}
E\left\|\frac{1}{T^2}\sum_t\sum_r x_{ir}^* x_{it}^{*\prime}\gamma_t\right\|_2^2 &\leq \frac{1}{T^4}\left[E\left\|\sum_r x_{ir}^*\right\|_2^4\right]^{1/2}\left[E\left\|\sum_t \gamma_t' x_{it}^*\right\|_2^4\right]^{1/2} \\
&= O\left(\frac{1}{T^2}\right),
\end{aligned}
$$

where the second line follows by the use of Cauchy-Schwarz inequality, and the final bound holds by Lemma A.3.1 under Assumption A1.1(i).

(iii) Note that $E\left(\frac{1}{NT}\sum_{l\neq i}\sum_t x_{it}^* x_{lt}^{*\prime}\gamma_t\right) = 0$ and $E\left(x_{it}^* x_{lt}^{*\prime}\gamma_t\right) = 0$ under Assumption A1.1, then,

$$
\begin{aligned}
E\left\|\frac{1}{NT}\sum_{l=1,l\neq i}^{N}\sum_t x_{lt}^* x_{it}^{*\prime}\gamma_t\right\|^2 &= \frac{1}{N^2T^2}\sum_{t,s}\sum_{l=1,l\neq i}^{N}\sum_{j=1,j\neq i}^{N} E\left[\left(\gamma_t' x_{it}^* x_{lt}^*\right)\left(x_{js}^{*\prime} x_{is}^* \gamma_s\right)\right] \\
&\text{.... center} \\
&\lesssim \frac{1}{N^2T^2}\sum_{t,s}\alpha(|t-s|)^{\delta_1/(4+\delta_1)} \text{ NO} \\
&= \frac{1}{N^2T^2}O(T) = O\left(\frac{1}{N^2T}\right),
\end{aligned}
$$

where the second line holds because $\{\sum_{l=1,l\neq i}^{N}\gamma_t' x_{it}^* x_{lt}^*\}$ are still α-mixing sequences with zero mean under Assumption A1.1, then use Davydov's inequality and Lemma A.3.1 under Assumption A1.1 sequentially; and the final bound holds under Assumption A1.1(i).

(iv) By straightforward calculations, we have

$$
\begin{aligned}
E\left\|\frac{1}{NT^2}\sum_{l=1,l\neq i}^{N}\sum_{t,r}x_{ir}^{*}x_{lt}^{*\prime}\gamma_t\right\|^2 &= \frac{1}{N^2T^4}\sum_{l=1,l\neq i}^{N}\sum_{k=1,k\neq i}^{N}\sum_{t,r,s,q}E\left(\gamma_t^{\prime}x_{lt}^{*}x_{ir}^{*\prime}x_{is}^{*}x_{kq}^{*\prime}\gamma_q\right)\\
&\leq \frac{1}{NT^4}E\left[\left\|\sum_t\left(\frac{1}{\sqrt{N}}\sum_l\gamma_t^{\prime}x_{lt}^{*}\right)\right\|^2\left\|\sum_r x_{ir}^{*}\right\|^2\right]\\
&\leq \frac{1}{NT^4}\left[E\left\|\sum_r x_{ir}^{*}\right\|_2^4\right]^{1/2}\left[E\left\|\sum_t\left(\frac{1}{\sqrt{N}}\sum_l\gamma_t^{\prime}x_{lt}^{*}\right)\right\|_2^4\right]^{1/2}\\
&= \frac{1}{NT^4}O\left(T\right)O\left(T\right)=O\left(\frac{1}{NT^2}\right),
\end{aligned}
$$

where the third line holds by Cauchy-Schwarz inequality; and the last line follows by Lemma A.3.1 under Assumption A1.1. ■

Proof of Lemma A.2.7. Note that $E(x_{it}^{*}u_{it}) = E(x_{it}^{*}u_{ir}) = E(x_{it}^{*}u_{lt}) = E(x_{it}^{*}u_{lr}) = 0$ under Assumption A1.2(iii), then, the terms above are also mean zero. Thus, by Chebyshev's inequality, we just need to show following results:

(i) Let $\delta_* = \max\{\delta_1, \delta_2\}$, which indicates the larger value between δ_1 and δ_2. Then, we have

$$
\begin{aligned}
E\left\|\frac{1}{T}\sum_t x_{it}^{*}u_{it}\right\|_2^2 &= \frac{1}{T^2}\sum_t\sum_s E\left(u_{it}u_{is}x_{it}^{*\prime}x_{is}^{*}\right)\\
&\leq \frac{1}{T^2}\left(\sum_t\sum_s(\alpha(|t-s|))^{\delta_*/(4+\delta_*)}\right)\left[E\left\|x_{it}^{*}\right\|^{(4+\delta_*)}E\left\|x_{is}^{*}\right\|^{(4+\delta_*)}\right]^{2/(4+\delta_*)}\\
&\quad \left[E\left\|u_{it}^{*}\right\|^{(4+\delta_*)}E\left\|u_{is}^{*}\right\|^{(4+\delta_*)}\right]^{2/(4+\delta_*)}\\
&\leq O(1)\frac{1}{T^2}\sum_t\sum_s\left(\alpha(|t-s|)\right)^{\delta_*/(4+\delta_*)}\\
&= O\left(\frac{1}{T}\right),
\end{aligned}
$$

where the second line follows by Davydov's inequality for the mixing sequence and the final line holds under Assumption A1.1 and A1.2.

(ii) By Cauchy-Schwarz inequality

$$E\left\|\frac{1}{T^2}\sum_{t,s}x_{it}^*u_{is}\right\|^2 \leq \frac{1}{T^4}\left[E\left\|\sum_s u_{is}\right\|^4\right]^{1/2}\left[E\left\|\sum_t x_{it}^*\right\|^4\right]^{1/2} = O\left(\frac{1}{T^2}\right),$$

where we use the fact that $E\left|\sum_s u_{is}\right|^4 = O(T^2)$ and $E\left\|\sum_t x_{it}^*\right\|^4 = O(T^2)$ by Lemma A.3.1 under Assumption A1.1 and A1.2.

(iii)–(iv) The proofs are similar to those of (i)–(ii) and thus omitted. ■

Proof of Lemma A.2.8.

(i) The proof is analogous to that of Lemma A.2.3(iii) and thus omitted.

(ii) By direct calculations, we have

$$\begin{aligned}
E\left\|\frac{1}{NT^2}\sum_{l=1,l\neq i}^{N}\sum_{t,r}x_{it}^*x_{lr}^{*\prime}\lambda_l\right\|^2 &= \frac{1}{N^2T^4}E\left[\sum_{t,s}\left(x_{is}^{*\prime}x_{it}^*\right)\right]\left[\sum_{l=1,l\neq i}^{N}\sum_{k=1,k\neq i}^{N}\sum_{r,q}\left(\lambda_l'x_{lq}^*x_{kr}^{*\prime}\lambda_k\right)\right] \\
&\leq \frac{1}{N^2T^4}\left(E\left\|\sum_{l=1,l\neq i}^{N}\sum_r \lambda_l'x_{lr}^*\right\|^2\left\|\sum_t x_{it}^*\right\|^2\right) \\
&\leq \frac{1}{N^2T^4}\left[E\left\|\sum_{l=1,l\neq i}^{N}\sum_r \lambda_l'x_{lr}^*\right\|^4\right]^{1/2}\left[E\left\|\sum_t x_{it}^*\right\|^4\right]^{1/2} \\
&= \frac{1}{N^2T^4}O\left(NT\right)O\left(T\right) = O\left(\frac{1}{NT^2}\right),
\end{aligned}$$

where the final line holds by Lemma A.3.1 under Assumption A1.1. ■

A.2.3 Proof of Proposition A.1.3: Stochastic Bound of $\widehat{\gamma}_t - \gamma_t$

Recall that

$$\widehat{\gamma}_t - \gamma_t = \frac{T}{T-1}A_{3Nt}^{-1}\left(-D_{1Nt} + D_{2Nt} + D_{3Nt} + D_{4Nt}\right),$$

where A_{3Nt}, D_{1Nt}, D_{2Nt}, D_{3Nt} and D_{4Nt} are as defined in Appendix A:

$$
\begin{aligned}
A_{3Nt} &\equiv \frac{1}{N}\sum_i \breve{x}_{it}\breve{x}_{it}', \\
D_{1Nt} &\equiv \frac{1}{N}\sum_i \breve{x}_{it}\left(\tilde{x}_{it} - \frac{1}{N}\tilde{x}_{lt}\right)'\left(\widehat{\beta}-\beta\right) = \frac{1}{N}\sum_i \breve{x}_{it}\tilde{x}_{it}'\left(\widehat{\beta}-\beta\right) - \frac{1}{N^2}\sum_{i,l}\breve{x}_{it}\tilde{x}_{lt}'\left(\widehat{\beta}-\beta\right) \\
&\equiv D_{1Nt1} - D_{1Nt2}, \\
D_{2Nt} &\equiv \frac{1}{N}\sum_i \breve{x}_{it}\left(\tilde{x}_{it}'\lambda_i - \frac{1}{N}\sum_l \tilde{x}_{lt}'\lambda_l\right) = \frac{1}{N}\sum_i \breve{x}_{it}\tilde{x}_{it}'\lambda_i - \frac{1}{N^2}\sum_{i,l}\breve{x}_{it}\tilde{x}_{ls}'\lambda_l \equiv D_{2Nt1} - D_{2Nt2}, \\
D_{3Nt} &\equiv \frac{1}{N}\sum_i \breve{x}_{it}\left(\tilde{u}_{it} - \frac{1}{N}\sum_l \tilde{u}_{lt}\right) = \frac{1}{N}\sum_i \breve{x}_{it}\tilde{u}_{it} - \frac{1}{N^2}\sum_l\sum_i \breve{x}_{it}\tilde{u}_{lt} \equiv D_{3Nt1} - D_{3Nt2}, \\
D_{4Nt} &\equiv \frac{1}{NT}\sum_i\sum_{t\neq s}\breve{x}_{it}\breve{x}_{is}'\gamma_s.
\end{aligned}
$$

To prove Propositon A.1.3, we need the following four lemmas.

Lemma A.2.9. *Under Assumption A1.1 and A1.4,*

(i) $\left\|\frac{1}{N}\sum_i x_{it}^* x_{it}^{*\prime}\right\| = O_p(1)$,

(ii) $\left\|\frac{1}{N^2}\sum_i\sum_{l\neq i} x_{it}^* x_{lt}^{*\prime}\right\| = O_p(\frac{1}{N})$,

(iii) $\left\|\frac{1}{NT}\sum_i\sum_r x_{it}^* x_{ir}^{*\prime}\right\| = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T}\right)$,

(iv) $\left\|\frac{1}{N^2T}\sum_i\sum_{l\neq i}\sum_r x_{lt}^* x_{ir}^{*\prime}\right\| = O_p(\frac{1}{\sqrt{N^2T}})$.

Lemma A.2.10. *Under Assumption A1.1,*

(i) $\left\|\frac{1}{N}\sum_i x_{it}^* x_{it}^{*\prime}\lambda_i\right\| = O_p(\frac{1}{\sqrt{N}})$,

(ii) $\left\|\frac{1}{N^2}\sum_i\sum_{l\neq i} x_{lt}^* x_{it}^{*\prime}\lambda_i\right\| = O_p(\frac{1}{N})$,

(iii) $\left\|\frac{1}{NT}\sum_i\sum_r x_{it}^* x_{ir}^{*\prime}\lambda_i\right\| = O_p(\frac{1}{\sqrt{NT}})$,

(iv) $\left\|\frac{1}{N^2T}\sum_i\sum_{i\neq l}\sum_r x_{lt}^* x_{ir}^{*\prime}\lambda_i\right\| = O_p(\frac{1}{\sqrt{N^2T}})$.

Lemma A.2.11. *Under Assumptions A1.1 and A1.2,*

(i) $\left\|\frac{1}{N}\sum_i x_{it}^* u_{it}\right\| = O_p(\frac{1}{\sqrt{N}})$,

(ii) $\left\|\frac{1}{N^2}\sum_i\sum_l x_{lt}^* u_{it}\right\| = O_p(\frac{1}{N})$,

(iii) $\left\|\frac{1}{NT}\sum_i\sum_r x_{it}^* u_{ir}\right\| = O_p(\frac{1}{\sqrt{NT}})$,

(iv) $\left\|\frac{1}{N^2T}\sum_i\sum_l\sum_r x_{lt}^* u_{ir}\right\| = O_p(\frac{1}{\sqrt{N^2T}})$.

Lemma A.2.12. *Under Assumption A1.1,*

(i) $\left\| \frac{1}{NT} \sum_i \sum_{s \neq t} x_{it}^* x_{is}^{*\prime} \gamma_s \right\| = O_p(\frac{1}{\sqrt{NT^2}})$,

(ii) $\left\| \frac{1}{N^2T} \sum_i \sum_{l \neq i} \sum_{s \neq t} x_{it}^* x_{ls}^{*\prime} \gamma_s \right\| = O_p(\frac{1}{\sqrt{N^2T}})$.

Proof of Proposition A.1.3.

(i) Recall that $x_{it}^* = x_{it} - \mu$. By Lemma A.2.9,

$$
\begin{aligned}
A_{3Nt} &= \frac{1}{N} \sum_i \breve{x}_{it} \breve{x}_{it}^{\prime} \\
&= \frac{1}{N} \sum_i \left(x_{it}^* - \frac{1}{N} \sum_{l_1} x_{l_1 t}^* \right) \left(x_{it}^* - \frac{1}{N} \sum_{l_2} x_{l_2 t}^* \right)^{\prime} \\
&= \left(1 - \frac{1}{N} \right) \frac{1}{N} \sum_i x_{it}^* x_{it}^{*\prime} - \frac{1}{N^2} \sum_i \sum_{l \neq i} x_{it}^* x_{lt}^{*\prime} \\
&\equiv \frac{N-1}{N} A_{3N1} - A_{3N2} \xrightarrow{p} \lim_{N \to \infty} \frac{1}{N} \sum_i E\left(x_{it}^* x_{it}^{*\prime} \right) = \Sigma_x,
\end{aligned}
$$

Obviously, $A_{3N1} \xrightarrow{p} \lim\limits_{N \to \infty} \frac{1}{N} \sum_i E\left(x_{it}^* x_{it}^{*\prime} \right) = \Sigma_x$ and $A_{3N2} =$

(ii) Note that $D_{1NT2} = 0$ by the fact $\frac{1}{N} \sum_i \breve{x}_{it} = 0$, then $D_{1Nt} = D_{1Nt1}$. We can decompose D_{1NT1} as below,

$$
\begin{aligned}
D_{1Nt1} &= \frac{1}{N} \sum_i \breve{x}_{it} \tilde{x}_{it}^{\prime} \left(\widehat{\beta} - \beta \right) \\
&= \left[\frac{1}{N} \sum_i \left(x_{it}^* - \frac{1}{N} \sum_l x_{lt}^* \right) x_{it}^{*\prime} - \frac{1}{NT} \sum_i \sum_r \left(x_{it}^* - \frac{1}{N} \sum_l x_{lt}^* \right) x_{ir}^{*\prime} \right] \left(\widehat{\beta} - \beta \right) \\
&= \left[\frac{N-1}{N} \left(\frac{1}{N} \sum_i x_{it}^* x_{it}^{*\prime} - \frac{1}{NT} \sum_i \sum_r x_{it}^* x_{ir}^{*\prime} \right) - \frac{1}{N^2} \sum_i \sum_{l \neq i} x_{lt}^* x_{it}^{*\prime} + \frac{1}{N^2T} \sum_i \sum_{l \neq i} \sum_r x_{lt}^* x_{ir}^{*\prime} \right] \\
&\quad \times \left(\widehat{\beta} - \beta \right) \\
&\equiv \left[\frac{N-1}{N} \left(D_{1Nt11} - D_{1Nt12} \right) - D_{1Nt13} - D_{1Nt14} \right] \times \left(\widehat{\beta} - \beta \right),
\end{aligned}
$$

It is clear that the order of D_{1T} is determined by both $\{D_{1Nt1k}\}_{k=1}^4$ and $\left(\widehat{\beta} - \beta \right)$. By Lemma A.2.9, $\|D_{1Nt11}\|_2 = O_p(1)$, D_{1Nt12} to D_{1Nt14} are of order $o_p(1)$ as $(N, T) \to \infty$. So, we denote the order of D_{1N} as $D_{1N} = O_p\left(\left\|\widehat{\beta} - \beta\right\|\right)$, which means the order of D_{1N} is totally determined by the order of $\left\|\widehat{\beta} - \beta\right\|$.

(iii) Recall that $D_{2Nt1} = D_{2Nt1} - D_{2Nt2} = \frac{1}{N}\sum_i \breve{x}_{it}\tilde{x}_{it}'\lambda_i - \frac{1}{N}\sum_l \left(\frac{1}{N}\sum_i \breve{x}_{it}\right)\tilde{x}_{ls}'\lambda_l$ Note that $\frac{1}{N}\sum_i \breve{x}_{jt} = 0$ because $\breve{x}_{jt} = \frac{1}{N}\sum_j x_{jt}$, and therefore $D_{2Nt2} = 0$. Then, we can further decompose $D_{2Nt} = D_{2Nt1}$ as below,

$$
\begin{aligned}
D_{2Nt} &= \frac{1}{N}\sum_i \breve{x}_{it}\tilde{x}_{it}'\lambda_i = \frac{1}{N}\sum_i \left(x_{it}^* - \frac{1}{N}\sum_l x_{lt}^*\right)\left(x_{it}^* - \frac{1}{T}\sum_r x_{ir}^*\right)'\lambda_i \\
&= \left(1 - \frac{1}{N}\right)\left[\frac{1}{N}\sum_i x_{it}^* x_{it}^{*\prime}\lambda_i - \frac{1}{NT}\sum_i\sum_r x_{it}^* x_{ir}^{*\prime}\lambda_i\right] - \frac{1}{N^2}\sum_i\sum_{l\neq i} x_{lt}^* x_{it}^{*\prime}\lambda_i \\
&\quad + \frac{1}{N^2 T}\sum_i\sum_{i\neq l}\sum_r x_{lt}^* x_{ir}^{*\prime}\lambda_i \\
&\equiv \frac{N-1}{N}\left(D_{2Nt11} - D_{2Nt12}\right) - D_{2Nt13} - D_{2Nt14},
\end{aligned}
$$

According to Lemma A.2.10, $\|D_{2Nt11}\|_2 = O_p(\frac{1}{\sqrt{N}})$, $\|D_{2Nt12}\|_2 = O_p(\frac{1}{\sqrt{NT}})$,$\|D_{2Nt13}\|_2 = O_p(\frac{1}{N})$, $\|D_{2Nt14}\|_2 = O_p(\frac{1}{\sqrt{N^2 T}})$. It follows that $\|D_{2Nt}\|_2 = O_p(\frac{1}{\sqrt{N}})$.

(iv) Recall that $D_{3Nt} = D_{3Nt1} - D_{3Nt2} = \frac{1}{N}\sum_i \breve{x}_{it}\tilde{u}_{it} - \frac{1}{N^2}\sum_l\sum_i \breve{x}_{it}\tilde{u}_{lt}$. Use the fact that $\frac{1}{N}\breve{x}_{it} = 0$ again, and thus $D_{3Nt2} = 0$, then $D_{3Nt} = D_{3Nt1}$. We can further decompose D_{3Nt1} as below

$$
\begin{aligned}
D_{3Nt1} &= \frac{1}{N}\sum_i \left(x_{it}^* - \frac{1}{N}\sum_l x_{lt}^*\right)\left(u_{it} - \frac{1}{T}\sum_r u_{ir}\right) \\
&= \frac{1}{N}\sum_i x_{it}^* u_{it} - \frac{1}{N^2}\sum_i\sum_l x_{lt}^* u_{it} - \frac{1}{NT}\sum_i\sum_r x_{it}^* u_{ir} + \frac{1}{N^2 T}\sum_i\sum_l\sum_r x_{lt}^* u_{ir} \\
&\equiv D_{3Nt11} - D_{3Nt12} - D_{3Nt13} + D_{3Nt14},
\end{aligned}
$$

According to Lemma A.2.11, $\|D_{3Nt11}\|_2 = O_p(\frac{1}{\sqrt{N}})$, $\|D_{3Nt12}\|_2 = O_p(\frac{1}{N})$, $\|D_{3Nt13}\|_2 = O_p(\frac{1}{\sqrt{NT}})$, $\|D_{3Nt14}\|_2 = O_p(\frac{1}{\sqrt{N^2 T}})$. Thus, it follows that $\|D_{3Nt}\|_2 = O_p(\frac{1}{\sqrt{N}})$.

(v) Note that $D_{4N} \equiv \frac{1}{NT}\sum_i \sum_{s\neq t} \breve{x}_{it}\breve{x}_{is}'\gamma_s$, then we can decompose D_{4N} as follows:

$$\begin{aligned} D_{4N} &= \frac{1}{NT}\sum_i \sum_{s\neq t}\left(x_{it}^* - \frac{1}{N}\sum_{l_1} x_{l_1 t}^*\right)\left(x_{is}^* - \frac{1}{N}\sum_{l_2} x_{l_2 s}^*\right)'\gamma_s \\ &= \left(1-\frac{1}{N}\right)\frac{1}{NT}\sum_i\sum_{s\neq t} x_{it}^* x_{is}^{*\prime}\gamma_s - \frac{1}{N^2T}\sum_i\sum_{l\neq i}\sum_{s\neq t} x_{it}^* x_{ls}^{*\prime}\gamma_s \\ &\equiv \frac{N-1}{N}D_{4N1} - D_{4N2}, \end{aligned}$$

According to Lemma A.2.12, $\|D_{4Nt1}\|_2 = O(\frac{1}{\sqrt{NT^2}})$ and $\|D_{4N2}\|_2 = O(\frac{1}{\sqrt{N^2T}})$. Then, $\|D_{4N}\|_2 = O(\frac{1}{\sqrt{N^2T}}) + O(\frac{1}{\sqrt{NT^2}})$. ■

Proof of Lemma A.2.9

(i) As in Lemma A.2.1(i) and A.2.5(i), we have $\frac{1}{N}\sum_i x_{it}^* x_{it}^{*\prime} \xrightarrow{p} \lim_{N\to\infty}\frac{1}{N}\sum_i E\left(x_{it}^* x_{it}^{*\prime}\right) = \Sigma_x$ under Assumption A1.1. Then the result follows.

(ii) By direct calculations, we have

$$\begin{aligned} E\left\|\frac{1}{N^2}\sum_i\sum_{l\neq i} x_{it}^* x_{lt}^{*\prime}\right\|^2 &= E\left\|\frac{1}{N^2}\sum_{i,l} x_{it}^* x_{lt}^{*\prime} - \frac{1}{N^2}\sum_i x_{it}^* x_{li}^{*\prime}\right\|^2 \\ &\lesssim \frac{1}{N^2}E\left\|\frac{1}{\sqrt{N}}\sum_i x_{it}^*\right\|^4 + O\left(\frac{1}{N^2}\right) = O\left(\frac{1}{N^2}\right). \end{aligned}$$

(iii) The proof is analogous to that of Lemma A.2.5(iii).

(iv) By direct calculations, we have

$$\begin{aligned} E\left\|\frac{1}{N^2T}\sum_{k\neq i}\sum_r x_{kt}^* x_{ir}^{*\prime}\right\|^2 &= E\left\|\frac{1}{N^2T}\sum_{k,i}\sum_r x_{kt}^* x_{ir}^{*\prime} - \frac{1}{N^2T}\sum_i\sum_r x_{it}^* x_{ir}^{*\prime}\right\|^2 \\ &\lesssim \frac{1}{N^2T^2}E\left\|\sum_s\left(\frac{1}{\sqrt{N}}\sum_i x_{is}^*\right)\right\|^2 E\left\|\frac{1}{\sqrt{N}}\sum_i x_{it}^*\right\|^2 \\ &\quad + E\left\|\frac{1}{N^2T}\sum_i\sum_r x_{it}^* x_{ir}^{*\prime}\right\|^2 \\ &= \frac{1}{N^2T^2}O(T)\,O(1) + O\left(\frac{1}{N^2T^2}\right) = O\left(\frac{1}{N^2T}\right), \end{aligned}$$

where the last line holds by Lemma A.3.1 and Assumption A1.1. ■

Proof of Lemma A.2.10

(i) Note that $E\left(\frac{1}{N}\sum_i x_{it}^* x_{it}^{*\prime}\lambda_i\right) = 0$ under Assumption A1.1, then,

$$\begin{aligned} E\left\|\frac{1}{N}\sum_i x_{it}^* x_{it}^{*\prime}\lambda_i\right\|^2 &= E\left\|\frac{1}{N}\sum_i \left(x_{it}^* x_{it}^{*\prime} - \Sigma_x\right)\lambda_i + \Sigma_x\lambda_i\right\|^2 \\ &\leq \frac{2}{N^2}\sum_{i,j} E\left(\lambda_i'\lambda_j\right)\|\Sigma_x\|_{\text{sp}}^2 + \frac{2}{N^2}\sum_{i,j} E\left[\lambda_i'\left(x_{it}^* x_{it}^{*\prime} - \Sigma_x\right)\left(x_{jt}^* x_{jt}^{*\prime} - \Sigma_x\right)\lambda_j\right] \\ &= \frac{1}{N^2}O(N) + \frac{1}{N^2}O(N) = O\left(\frac{1}{N}\right), \end{aligned}$$

where we use the fact that $\text{Var}\left(N^{-1/2}\sum_i \lambda_i\right) = O(1)$ under Assumption A1.1(iii), and $\{\lambda_i'\left(x_{it}^* x_{it}^{*\prime} - \Sigma_x\right)\}$ is a mixing sequence satisfying Assumption A1.1(i).

(ii) Note that $E\left(\frac{1}{N^2}\sum_{l\neq i} x_{lt}^* x_{it}^{*\prime}\lambda_i\right) = 0$ and $E\left(x_{lt}^* x_{it}^{*\prime}\lambda_i\right) = 0$ under Assumption A1.1, then,

$$\begin{aligned} E\left\|\frac{1}{N^2}\sum_{l\neq i} x_{lt}^* x_{it}^{*\prime}\lambda_i\right\|^2 &= \frac{1}{N^2}E\left(\frac{1}{N}\sum_{l\neq i}\lambda_i' x_{it}^* x_{lt}^*\right)\left(\frac{1}{N}\sum_{j\neq k} x_{jt}^{*\prime} x_{kt}^*\lambda_k\right) \\ &\lesssim \frac{1}{N^2}\left[\alpha(0)^{\delta_1/(4+\delta_1)}\right] = O\left(\frac{1}{N^2}\right), \end{aligned}$$

where the second line holds because $\{N^{-1}\sum_{l\neq i}\lambda_i' x_{it}^* x_{lt}^*\}$ is an α-mixing sequence with zero mean under Assumption A1.1, then use Davydov's inequality and Lemma A.3.1 under Assumption A1.1 sequentially; and the final bound holds under Assumption A1.1(i).

(iii) Note that $E\left(\frac{1}{NT}\sum_i\sum_r x_{it}^* x_{ir}^{*\prime}\lambda_i\right) = 0$ and $E\left(x_{it}^* x_{ir}^{*\prime}\lambda_i\right) = 0$ under Assumption A1.1. Then

$$\begin{aligned} E\left\|\frac{1}{NT}\sum_i\sum_r x_{it}^* x_{ir}^{*\prime}\lambda_i\right\|^2 &= \frac{1}{N^2T^2}\sum_{i,j} E\left(\frac{1}{T}\sum_r \lambda_i' x_{is}^* x_{it}^*\right)\left(\frac{1}{T}\sum_q x_{jt}^{*\prime} x_{jq}^*\lambda_j\right) \\ &\lesssim \frac{1}{N^2T}\sum_{i,j}\alpha_{ij}(0)^{\delta_1/(4+\delta_1)} = O\left(\frac{1}{NT}\right), \end{aligned}$$

where the second line holds because $\{T^{-1}\sum_r \lambda_i' x_{ir}^* x_{it}^*\}$ are still α-mixing sequences with zero mean under Assumption A1.1, then use Davydov's inequality and Lemma A.3.1 under Assumption A1.1

sequentially; and the final bound holds under Assumption A1.1(i).

(iv) Note that $E\left(\frac{1}{N^2T}\sum_{i\neq l}\sum_r x_{lt}^{*}x_{ir}^{*\prime}\lambda_i\right)=0$ under Assumption A1.1. Then

$$
\begin{aligned}
& E\left\|\frac{1}{N^2T}\sum_{l\neq i}\sum_{r} x_{lt}^{*}x_{ir}^{*\prime}\lambda_i\right\|^2 \\
= & \frac{1}{N^4T^2}\sum_{i,j}\sum_{r,q}E\left[\lambda_i^{\prime}x_{ir}^{*}\left(\sum_{l\neq i}x_{lt}^{*\prime}\right)\left(\sum_{k\neq j}x_{kt}^{*}\right)x_{jq}^{*\prime}\lambda_j\right] \text{ to be changed} \\
\leq & \frac{1}{N^4T^2}E\left(\left\|\sum_{l}x_{lt}^{*}\right\|\left\|\sum_{k\neq j}x_{kt}^{*}\right\|\left\|\sum_{i}\sum_{r}x_{ir}^{*\prime}\lambda_i\right\|_2\left\|\sum_{j}\sum_{q}x_{jq}^{*\prime}\lambda_j\right\|_2\right) \quad \text{NO} \\
\leq & \frac{1}{N^2T^2}\left[E\left\|\frac{1}{\sqrt{N}}\sum_{j\neq i}x_{jt}^{*}\right\|_2^4\right]^{1/2}\left[E\left\|\frac{1}{\sqrt{N}}\sum_{i}\sum_{r}x_{ir}^{*\prime}\lambda_i\right\|_2^4\right]^{1/2} \\
= & \frac{1}{N^2T^2}O\left(1\right)O\left(T\right)=O\left(\frac{1}{N^2T}\right),
\end{aligned}
$$

where the fourth line follows by repeating uses of Cauchy-Schwarz inequality, and the final bound holds by Lemma A.3.1 under Assumption A1.1(i). ■

Proof of Lemma A.2.11. By Chebyshev' inequality, we just need to show following results

(i) Note that $E\left(\frac{1}{N}\sum_i x_{it}^{*}u_{it}\right)=0$ and $E\left(x_{it}^{*}u_{it}\right)=0$ under Assumption A1.2(ii). Then

$$
\begin{aligned}
E\left\|\frac{1}{N}\sum_{i}x_{it}^{*}u_{it}\right\|^2 &= \frac{1}{N^2}\sum_{i,j}E\left(x_{it}^{*}u_{it}x_{jt}^{*}u_{jt}\right) \\
&\lesssim \frac{1}{N^2}\sum_{i,j}\alpha_{ij}(0)^{\delta_*/(4+\delta_*)}=O\left(\frac{1}{N}\right),
\end{aligned}
$$

where the final bound holds because $\{x_{it}^{*}u_{it}\}$ are still mixing sequences under Assumptions A1.1 and A1.2.

(ii) Note that $E\left(\frac{1}{N^2}\sum_i\sum_l x_{lt}^{*}u_{it}\right)=0$ and $E\left(x_{lt}^{*}u_{it}\right)=0$ under Assumption A1.2(ii). Then

$$
E\left\|\frac{1}{N^2}\sum_{i}\sum_{l}x_{lt}^{*}u_{it}\right\|^2=\frac{1}{N^2}\left[E\left\|\frac{1}{\sqrt{N}}\sum_{i}x_{it}^{*}\right\|^4\right]^{1/2}\left[E\left\|\frac{1}{\sqrt{N}}\sum_{l}u_{lt}\right\|^4\right]^{1/2}=O\left(\frac{1}{N^2}\right),
$$

where the final bound holds under Assumptions A1.1 and A1.2.

(iii)–(iv) The proofs are similar to those of (i)–(ii) and thus omitted. ■

Proof of Lemma A.2.12. By Chebyshev inequality, it suffices to show as follows:

(i) Note that $E\left(\frac{1}{NT}\sum_i \sum_{s=1,s\neq t}^{T} x_{it}^* x_{is}^{*\prime}\gamma_s\right) = 0$ and $E\left(x_{it}^* x_{is}^{*\prime}\gamma_s\right) = 0$ under Assumption A1.1, then,

$$\begin{aligned}
E\left\|\frac{1}{NT}\sum_i \sum_{s=1,s\neq t}^{T} x_{it}^* x_{is}^{*\prime}\gamma_s\right\|^2 &= \frac{1}{N^2T^2}\sum_{i,j} E\left(\sum_{r=1,r\neq t}^{T}\gamma_r' x_{ir}^* x_{it}^*\right)\left(\sum_{s=1,s\neq t}^{T} x_{jt}^{*\prime} x_{js}^*\gamma_s\right) \\
&\leq O(1)\frac{1}{N^2T^2}\sum_{i,j}\alpha_{ij}(0)^{\delta_1/(4+\delta_1)} \\
&= \frac{1}{N^2T^2}O(N) = O\left(\frac{1}{NT^2}\right),
\end{aligned}$$

where the second line holds because $\{\sum_{s=1,s\neq t}^{T}\gamma_s' x_{is}^* x_{it}^*\}$ are still α-mixing sequences with zero mean under Assumption A1.1, and the final bound holds under Assumption A1.1(i).

(ii) The proof is similar to that of (i) and thus omitted. ■

A.3 Some Technical Lemmas

In this section, we present some lemmas that are used in the proof of the technical lemmas in the last section.

Lemma A.3.1. *(Theorem 4.1, Shao et al. (1996)) Let $2 < p < r \leq \infty$, $2 < v \leq r$ and $\{\xi_t,\ t \geq 1\}$ be an $\alpha-$mixing sequence of random variables with $E(\xi_t) = 0$ and $\|\xi_t\|_r \equiv (E|\xi_t|^r)^{1/r} < \infty$. Assume that $\alpha(s) \leq Cs^{-\theta}$ for some $C > 0$ and $\theta > 0$. Let $S_T = \sum_{t=1}^{T}\xi_t$. Then, for any $\varepsilon > 0$, there exists $K = K(\varepsilon, r, p, v, \theta, C) < \infty$ such that*

$$E|S_T|^p \leq K\left((TC_T)^{p/2}\max_{1\leq t\leq T}\|\xi_t\|_v^p + T^{(p-(r-p)\theta/r)\vee(1+\varepsilon)}\max_{1\leq t\leq T}\|\xi_t\|_r^p\right)$$

where $C_T = \left(\sum_{t=0}^{T}(t+1)^{2/(v-2)}\alpha(t)\right)^{(v-2)/v}$. *In particular,*

$$E\left|S_T\right|^p \leq K\left(T^{p/2}\max_{1\leq t\leq T}\|\xi_t\|_v^p + T^{1+\varepsilon}\max_{1\leq t\leq T}\|\xi_t\|_r^p\right)$$

if $\theta > v/(v-2)$ *and* $\theta \geq (p-1)r/(r-p)$, *and*

$$E\left|S_T\right|^p \leq KT^{p/2}\max_{1\leq t\leq T}\|\xi_t\|_r^p$$

if $\theta \geq pr/(2(r-p))$.

The lemma is adapted from Lemma S1.2(2) in Zhu (2017), and our proof strategy is the same as those in Zhu (2017).

Lemma A.3.2. *Under Assumption A1,* $\max_{i,t,k_1,k_2}\sum_{s=1}^{T}\|E(x_{it,k_1}x_{js,k_2})\| = O(1)$

Proof. By Corollary 16.2.4 of Athreya and Lahiri (2006), $|\mathbb{E}\left(x_{it,k_1}x_{is,k_2}\right)| \leq 4\left[2\alpha(|t-s|)\right]^{1/2}C^2$ $\forall i,t,s,k_1,k_2$. It follows that for some $\kappa > 2$,

$$\begin{aligned}\max_{i,t,k_1,k_2}\sum_{s=1}^{T}\left|\mathbb{E}\left(x_{it,k_1}x_{js,k_2}\right)\right| &\lesssim \max_t\sum_{s=1}^{T}\left[\alpha(|t-s|)\right]^{1/2} \lesssim \max_t\sum_{s=1}^{T}|t-s|^{-\kappa/2} \\ &\lesssim \sum_{\tau=1}^{\infty}\tau^{-\kappa/2} \leq \infty. \blacksquare\end{aligned}$$

Lemma A.3.3. *(Lemma C.3 in Zhu (2017)) The following hold.*

(1) Let $Z \in \mathbb{R}^{m_Z}$ *be a random vector whose jth entry is denoted by* Z_j. *Suppose that there exist constants* $b, \gamma > 0$ *such that* $\forall j \in [m_Z]$, Z_j *has an exponential-type tail with parameter* (b,γ) *Then for any nonrandom vector* $a \in \mathbb{R}^m$, $a'Z$ *has an exponential-type tail with parameter* $\left(b\|a\|_1\log^{1/\gamma}\left(\|a\|_0+2\right),\gamma\right)$

(2) Let $\{Z_j\}_{j=1}^{m_Z}$ *be a sequence of random variables. Suppose that constants* $b,\gamma > 0$ *satisfy that* $\forall j \in [m_Z]$, Z_j *has an exponential-type tail with parameter* (b,γ). *Let* $q > 0$ *be any nonran- dom number. Then there exists a constant* $C_{\gamma,q} > 0$ *depending only on* γ *and* q *such that* $\mathbb{E}\max_{1\leq j\leq m_Z}|Z_j|^q \leq C_{\gamma,q}m_Zb^q$ *and* $\mathbb{E}\left|Z_j\right|^q \leq C_{\gamma,q}b^q\ \forall j \in [m_Z]$

(3) Let Z_1 and Z_2 be two random variables having exponential-type tails with parameters (b_1, γ_1) and (b_2, γ_2), respectively. Then $\forall \gamma \in \left(0, \gamma_1\gamma_2(\gamma_1+\gamma_2)^{-1}\right)$, $Z_1 Z_2$ has an exponential-type tail with parameter $\left(2^{1/\gamma_0} b_1 b_2, \gamma_0\right)$, where $\gamma_0 = \gamma_1\gamma_2(\gamma_1+\gamma_2)^{-1}$.

(4) Let X have an exponential-type tail with parameter (b_X, γ_X). Then $\forall a \in \mathbb{R}, X - a$ has an exponential-type tail with parameter $(b_X + |a|, \gamma_X)$.

A.4 Additional Discussions on the Effects of Heterogeneity in $\{x_{it}\}$

In this section, we will show that the key results we established under the homogeneity of x_{it} are still valid with slight modifications.

We relax the homogeneity assumption on x_{it}. Namely, we drop Assumption A1.1(ii), and instead, assume

Assumption S1. *$E(x_{it}) = \mu_i$, $Var(x_{it}) = \Sigma_{i,x}$ and $Cov(x_{it}, x_{it-k}) = \Gamma_{i,k}$, which vary across i. Besides, we denote $d_i = \mu_i - \bar{\mu}$ such that $d_i = O(1)$ in general, where $\bar{\mu} = N^{-1}\sum_i \mu_i$.*

Recall that $\ddot{x}_{it} = x_{it} - \bar{x}_{i\cdot} - \bar{x}_{\cdot t} + \bar{x}_{\cdot\cdot}$, by simple and direct calculations, it follows $\ddot{x}_{it} = x^*_{it} - \bar{x}^*_{i\cdot} - \bar{x}^*_{\cdot t} + \bar{x}^*_{\cdot\cdot}$ under the heterogeneity assumption imposed above, where $x^*_{it} = x_{it} - \mu_i$, and $\bar{x}^*_{i\cdot}$, $\bar{x}^*_{\cdot t}$, and $\bar{x}^*_{\cdot\cdot}$ are defined as before based on x^*_{it}. Similarly, $\tilde{x}_{it} = x^*_{it} - \bar{x}^*_{i\cdot}$ also holds by direct calculations. Differently, we have $\breve{x}_{it} = x^*_{it} - \bar{x}^*_{\cdot t} + d_i$ now.

By careful calculations as those for homogeneous settings of x_{it}, we can readily show that most of results in the Appendix A still hold, the heterogeneity on $\{x_{it}\}$ only has impacts on the asymptotic behavior of $\widehat{\gamma}_t$ for $t = 1, \ldots, T$.

In particular, the heterogeneity of x_{it} we consider above have no impacts on the asymptotic distribution of $\widehat{\beta}$ due to the symmetric structure of (I.2 .2). To see this point, according to results in Section S1 and above direct calculations about $\ddot{x}_{it}$ and $\tilde{x}_{it}$, immediate results for the A_{1NT}, B_{1NT} and B_{2NT} are the exact same as those in homogeneity case even under Assumption S1. By repeated and tedious derivations, we can readily show the bounds for A_{1NT}, B_{1NT} and B_{2NT} in Section S1 still pertain if we impose some necessary but mild conditions as those Assumption A1.1.

We now turn to B_{3NT} below, recall that

$$B_{3NT} = B_{3NT1} - B_{3NT2} = \frac{1}{NT}\sum_i \sum_t \breve{x}_{it}\breve{x}_{it}'\gamma_t - \frac{1}{NT^2}\sum_i \sum_s \sum_t \breve{x}_{is}\breve{x}_{it}'\gamma_t.$$

In particular,

$$\begin{aligned} B_{3NT1} &= \frac{1}{NT}\sum_i \sum_t \breve{x}_{it}\breve{x}_{it}'\gamma_t \\ &= \frac{1}{NT}\sum_i \sum_t \left[x_{it}^* - \frac{1}{N}\sum_j x_{jt}^* + d_i\right]\left[x_{it}^* - \frac{1}{N}\sum_j x_{jt}^* + d_i\right]'\gamma_t \\ &= \frac{1}{NT}\sum_i \sum_t \left[x_{it}^* - \frac{1}{N}\sum_j x_{jt}^*\right]\left[x_{it}^* - \frac{1}{N}\sum_j x_{jt}^*\right]'\gamma_t + \frac{1}{NT}\sum_i \sum_t d_i d_i'\gamma_t \\ &\quad + \frac{2}{NT}\sum_i \sum_t d_i\left[x_{it}^* - \frac{1}{N}\sum_j x_{jt}^*\right]'\gamma_t \equiv B_{3NT1a} + B_{3NT1b} + B_{3NT1c}, \end{aligned}$$

Following the same arguments for homogeneity case, $B_{3NT1a} = O_p\left(T^{-1/2}\right)$, and by straightforward calculations, it is easy to obtain that $B_{3NT1b} = O_p\left(T^{-1/2}\right)$, and $B_{3NT1c} = O_p\left(N^{-1/2}T^{-1/2}\right)$ under some necessary but mild conditions as those Assumption A1.1. Therefore, $B_{3NT1} = O_p\left(T^{-1}\right)$ still holds as we show in the homogeneity case. And similarly, we can show $B_{3NT2} = o_p\left(T^{-1/2}\right)$, which is the same as the homogeneity case too. Therefore, under Assumption S1, the order of B_{3NT} does not change. Then we can rewrite B_{3NT} as follows,

$$B_{3NT} = \frac{1}{NT}\sum_i \sum_t \left(x_{it}^* x_{it}^{*\prime} + d_i d_i'\right)\gamma_t + o_p\left(\frac{1}{\sqrt{T}}\right).$$

Following arguments of proofs for Theorem 1.3.1 in the Appendix A, we can show that the asymptotic distribution of B_{3NT} is the normal distribution and asymptotically independent of the asymptotic distribution of B_{2NT}.

In summary, under heterogeneity of x_{it}, those bounds for the related terms still can be unchanged as those in homogeneity case under mild conditions, and we can show the asymptotic normality of $\widehat{\beta}-\beta$ by employing suitable central limit theorems and almost sure representation theorem.

Under heterogeneity case, the expansions of $\widehat{\lambda}_i-\lambda_i$ indeed does not change by direct calculations and noting the fact that $\tilde{x}_{it}=x_{it}^*-\bar{x}_{i\cdot}^*$ still holds under Assumption S1, then, we have

$$\begin{aligned}
&\widehat{\lambda}_i-\lambda_i\\
= & -\left[\left(\frac{1}{NT}\sum_j\sum_t x_{jt}^*x_{jt}^{*\prime}\right)^{-1}\frac{1}{NT}\sum_{j\neq i}\sum_t x_{jt}^*x_{jt}^{*\prime}\lambda_j+\left(\frac{1}{NT}\sum_j\sum_t x_{jt}^*x_{jt}^{*\prime}\right)^{-1}\frac{1}{NT}\sum_j\sum_t x_{jt}^*x_{jt}^{*\prime}\gamma_t\right]\\
&+\left(\frac{1}{T}\sum_t x_{it}^*x_{it}^{*\prime}\right)^{-1}\frac{1}{T}\sum_t x_{it}^*x_{it}^{*\prime}\gamma_t+\left(\frac{1}{T}\sum_t x_{it}^*x_{it}^{*\prime}\right)^{-1}\frac{1}{T}\sum_t x_{it}^*u_{it}+o_p\left(\delta_{NT}^{-1}\right).
\end{aligned}$$

Under Assumption S1, we can further rewrite $\widehat{\lambda}_i-\lambda_i$ as follows, namely,

$$\begin{aligned}
&\widehat{\lambda}_i-\lambda_i\\
= & -\left(\frac{1}{N}\sum_j\Sigma_{j,x}\right)^{-1}\frac{1}{NT}\sum_{j\neq i}\sum_t x_{jt}^*x_{jt}^{*\prime}\lambda_j-\left(\frac{1}{N}\sum_j\Sigma_{j,x}\right)^{-1}\frac{1}{NT}\sum_j\sum_t\left(x_{jt}^*x_{jt}^{*\prime}-\Sigma_{j,x}\right)\gamma_t\\
&+\Sigma_{i,x}^{-1}\frac{1}{T}\sum_t\left(x_{it}^*x_{it}^{*\prime}-\Sigma_{i,x}\right)\gamma_t+\left(\frac{1}{T}\sum_t x_{it}^*x_{it}^{*\prime}\right)^{-1}\frac{1}{T}\sum_t x_{it}^*u_{it}+o_p\left(\delta_{NT}^{-1}\right)\\
= & -\mathcal{R}_{1i}^{\lambda}-\mathcal{R}_{2i}^{\lambda}+\mathcal{R}_{3i}^{\lambda}+\mathcal{R}_{4i}^{\lambda}.
\end{aligned}$$

If we treat $\left(-\mathcal{R}_{2i}^{\lambda}+\mathcal{R}_{3i}^{\lambda}\right)$ as the whole, we can have similar results as shown in Theorem 1.3.3 by using suitable notations and appropriate central limit theorems under mild conditions.

Similarly, under Assumption S1, the asymptotic normality of $\widehat{\gamma}_t-\gamma_t$ still holds by using suitable notations and appropriate central limit theorems. To see this point, note that $\breve{x}_{it}=x_{it}^*-\bar{x}_{\cdot t}^*+d_i$ under Assumption S1 now, we can show $\widehat{\gamma}_t-\gamma_t\asymp A_{3Nt}^{-1}\left(\mathcal{R}_{1Nt}^{\gamma}+\mathcal{R}_{2Nt}^{\gamma}+\mathcal{R}_{3Nt}^{\gamma}\right)$ by direct calculations,

where

$$
\begin{aligned}
A_{3Nt} &= \frac{1}{N}\sum_i \breve{x}_{it}\breve{x}_{it}' = \frac{1}{N}\sum_i x_{it}^* x_{it}^{*\prime} + \frac{1}{N}\sum_i d_i d_i' + O_p\left(\frac{1}{\sqrt{N}}\right) \\
\mathcal{R}_{1Nt}^{\gamma} &= -\left(\frac{1}{N}\sum_i x_{it}^* x_{it}^{*\prime} + \frac{1}{N}\sum_i d_i d_i'\right)\left(\frac{1}{NT}\sum_i\sum_t x_{it}^* x_{it}^{*\prime}\right)^{-1}\left[\frac{1}{NT}\sum_i\sum_t x_{it}^* x_{it}^{*\prime}\lambda_i\right] \\
&\quad + \frac{1}{N}\sum_i x_{it}^* x_{it}^{*\prime}\lambda_i + \frac{1}{N}\sum_i d_i x_{it}^{*\prime}\lambda_i + o_p\left(\frac{1}{\sqrt{N}}\right) \\
\mathcal{R}_{2Nt}^{\gamma} &= -\left(\frac{1}{N}\sum_i x_{it}^* x_{it}^{*\prime} + \frac{1}{N}\sum_i d_i d_i'\right)\left(\frac{1}{NT}\sum_i\sum_t x_{it}^* x_{it}^{*\prime}\right)^{-1}\frac{1}{NT}\sum_i\sum_t x_{it}^* x_{it}^{*\prime}\gamma_t \\
&\quad + \frac{1}{NT}\sum_i\sum_{s\neq t} d_i d_i'\gamma_s + o_p\left(\frac{1}{\sqrt{N}}\right) \\
\mathcal{R}_{3Nt}^{\gamma} &= \frac{1}{N}\sum_i x_{it}^* u_{it} + \frac{1}{N}\sum_i d_i u_{it} + o_p\left(\frac{1}{\sqrt{N}}\right).
\end{aligned}
$$

Based on the above decomposition, we can still show that $\mathcal{R}_{1Nt}^{\gamma}$, $\mathcal{R}_{2Nt}^{\gamma}$, and $\mathcal{R}_{3Nt}^{\gamma}$ are asymptotically pair-wise uncorrelated, and A_{3Nt}^{-1} is nonsingular and well-defined under conditions. Following similar proofs as those for Theorem 1.3.4 in the Appendix A, it is straightforward to show $\widehat{\gamma}_t - \gamma_t$ will follow standard normal distribution by employing suitable central limit theorems and almost sure representation theorem under mild conditions.

Thus, the asymptotic normality of $\widehat{\beta}$, $\widehat{\lambda}_i$ and $\widehat{\gamma}_t$ will still hold under mild conditions as Theorems 1.3.1, 1.3.3 and 1.3.4 under the heterogeneity $\{x_{it}\}$. It is a standard exercise to follow proofs in the Appendix A to show the uniform consistency of $\widehat{\lambda}_i$ and $\widehat{\gamma}_t$, and asymptotic null distribution and local power properties for the max-type specification tests statistics after being rescaled appropriately under necessary but mild conditions.

Appendix B

Appendix to Chapter 2

B.1 Proofs of Main Results

To prove the main results in the paper, especially Theorem 2.3.1, Corollary 2.3.2, Theorem 2.3.3 and Theorem 2.3.4, we need some technical lemmas. Below we first state the technical lemmas whose proofs can be found on the online supplement, and then prove the main results in the paper.

B.1.1 Technical Lemmas

We first state some technical lemmas related to the consistency of the estimated mildly explosive factors under the null. In the case where the factors exhibit a unit root process, the results are relatively simple and similar to those in Bai and Ng (2004).

Consistency of the estimated mildly explosive factors under the null

Let $\mathcal{J}_T = (\boldsymbol{\rho}_0^0)^T \mathcal{J}_B$, where recall that $(\boldsymbol{\rho}_0^0)^T \equiv diag((\rho_{01}^0)^T, \ldots, \left(\rho_{0R_0}^0\right)^T)$ and $\mathcal{J}_B$ is a normalization constant used in the PCA estimation.

The following three lemmas hold under some mild assumptions when $\rho_{0,r} = 1 + \frac{c_r}{\kappa_T}$ with $c_r > 0$ being finite for $\forall r = 1, \ldots, R_0$, and $\rho_{0i} = 1$ for i.

Lemma B.1.1. *Suppose that Assumptions A2.1–A2.6 hold. Then $\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}$ is asymptotically invertible.*

Lemma B.1.2. *Suppose that Assumptions A2.1–A2.6 hold and the null hypothesis in (3.2 .5) holds. Then there exists $\boldsymbol{H}$ with asymptotic rank R_0 such that as $(N,T)\to\infty$,*

(a) $\frac{1}{T}\sum_{t=2}^{T}\left\|\boldsymbol{H}^{-1}\widehat{\boldsymbol{B}}_t-\boldsymbol{f}_t^0\right\|^2=O_p(N^{-p})+O_p\left(\left[\left(\rho_{01}^0\right)^{-2T}+\ldots+\left(\rho_{0R_0}^0\right)^{-2T}\right]T\right)$;

(b) $\left(\boldsymbol{H}^{-1}\widehat{\boldsymbol{B}}_t-\boldsymbol{f}_t^0\right)=O_p(N^{-p/2})$ *for each given* t;

(c) $\left(\widehat{\boldsymbol{\lambda}}_i-\boldsymbol{H}^{\prime -1}\boldsymbol{\lambda}_i^0\right)=O_p\left(\mathcal{J}_B^{-1}\left(1+\mathcal{J}_B^{-1}N^{-p/2}T^{1/2}\right)\right)$ *for each given* i.

Lemma B.1.3. *Under the assumptions of Lemma B.1.2,*

$$\max_{1\leq t\leq T}\frac{1}{\sqrt{T}}\left\|\sum_{s=2}^{t}\boldsymbol{H}^{-1}\widehat{\boldsymbol{B}}_s-\boldsymbol{f}_s^0\right\| \;=\; O_p\left(N^{-p/2}\right).$$

Remark B.1. Lemma B.1.2(a) reveals that under Assumptions A2.1-A2.6, $\hat{\boldsymbol{B}}_t$ can only estimate $\boldsymbol{f}_t^0=\boldsymbol{B}_t^0+\boldsymbol{u}_t$ instead of $\boldsymbol{B}_t$ consistently up to a rotation matrix $\boldsymbol{H}$. This is because the loading matrix for both $\boldsymbol{B}_t^0$ and $\boldsymbol{u}_t$ are the same as shown in (2.2 .10). However, for (2.2 .10), if we allow that $Z_{it}=\boldsymbol{\lambda}_{B,i}^{\prime}\boldsymbol{B}_t^0+\boldsymbol{\lambda}_{u,i}^{\prime}\boldsymbol{u}_t+\epsilon_{it}$ and further assume that $\|\boldsymbol{\Lambda}_B^{\prime}\boldsymbol{\Lambda}_u\|=O_p\left(N^{\upsilon}\right)$ and $\upsilon<p$, $\boldsymbol{B}_t^0$ and $\boldsymbol{u}_t$ can be identified separately by following the procedure that has been recently proposed by Peng et al. (2020).

Lemma B.1.4. *Under the assumptions of Lemma B.1.2,* $\left\|\boldsymbol{\Lambda}^0\boldsymbol{f}^{0\prime}-\widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{B}}^{\prime}\right\|^2=O_p\left(N+T\right)$.

Lemma B.1.4 is interesting in the sense that, under the null, if factors follow mildly explosive processes, common parts in (2.2 .10) can be estimated consistently for any $p>0$ because the rate does not involve p explicitly, which represents the intensity of factors in the current paper.

Consistency of estimated nonstationary factors under the null

In this subsection, we focus on the case with $p=1$, $\rho_{0,r}=1$ for $\forall r$, and $\rho_{0i}=1$ for all i. In this case, it is the similar case developed in Bai and Ng (2004), and Lemma B.1.5 and Lemma B.1.6

are presented for completeness and can readily obtained by following proofs in Bai and Ng (2004) directly under assumptions in the current paper.

Lemma B.1.5. *Let $\boldsymbol{f}_t$ be defined by (2.2 .9). Consider estimation of (2.2 .10) by the method of principal components and suppose that Assumptions A2.1-A2.6 hold. Then there exists an $\widetilde{\boldsymbol{H}}$ with rank R such that as $N, T \to \infty$, under the null hypothesis (2.2 .5), and $p = 1$*

(a) $\min\{N, T\} T^{-1} \sum_{t=1}^{T} \left\| \widetilde{\boldsymbol{H}}^{-1} \hat{\boldsymbol{B}}_t - \boldsymbol{f}_t^0 \right\|^2 = O_p(1)$

(b) $\min\left(\sqrt{N}, T\right) \left(\widetilde{\boldsymbol{H}}^{-1} \hat{\boldsymbol{B}}_t - \boldsymbol{f}_t^0\right) = O_p(1),$ *for each given t.*

(c) $\min\{\sqrt{T}, N\} \left(\hat{\boldsymbol{\lambda}}_i - \widetilde{\boldsymbol{H}}'^{-1} \boldsymbol{\lambda}_i^0\right) = O_p(1),$ *for each given i*

Lemma B.1.6. *Under the assumptions of Lemma B.1.5, and let $p = 1$, with the same $\widetilde{\boldsymbol{H}}$ in Lemma B.1.5,*

$$\max_{1 \leq t \leq T} \frac{1}{\sqrt{T}} \left\| \sum_{s=2}^{t} \widetilde{\boldsymbol{H}}^{-1} \hat{\boldsymbol{B}}_s - \boldsymbol{f}_s^0 \right\| = O_p\left(N^{-\frac{1}{2}}\right) + O_p\left(N^{-\frac{3}{4}}\right)$$

B.1.2 Proofs of Theorem 3.1

Mildly Explosive Factors Case

The structure of the proof follows the proofs of Lemmas 1, 2, 3 and 6 in Bai and Ng (2010) with a substatial amount of differences. to be emphasized. For clarity, we focus on the case where $\rho_{0,r} > 1$ for $r = 1, ..., R_0$ under the null. When $\rho_{0,r} = 1$ for each r, the proof is basically done in Bai and Ng (2010). The mixture case follows from the combination of arguments in these two separate cases.

Note that

$$\sqrt{N}T\left(\hat{\rho} - 1\right) = \sqrt{N}T \frac{tr\left(\hat{e}'_{-1} \Delta \hat{e}\right)}{tr\left(\hat{e}'_{-1} \hat{e}_{-1}\right)} = \left[\frac{1}{NT^2 \kappa_T} \sum_{i=1}^{N} \sum_{t=2}^{T} \hat{e}_{it}^2\right]^{-1} \frac{1}{\sqrt{N}T\kappa_T} \sum_{i=1}^{N} \sum_{t=2}^{T} \widehat{e}_{it-1} \Delta \widehat{e}_{it} \tag{B.1.1}$$

It suffices to prove the theorem by establishing the following results:

(i) $\frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=2}^{T} \hat{e}_{it}^2 = \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=2}^{T} e_{it}^2 + O_p(N^{1/2-p}) + O_p(\kappa_T N^{1/2} T^{-1}) = \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=2}^{T} e_{it}^2 + O_p(N^{1/2-p}) + o_p\left(1\right);$

(ii) $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}\widehat{e}_{it-1}\Delta\widehat{e}_{it}=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}e_{it-1}\Delta e_{it}+O_p(N^{-1/2}+T^{-1/2})+O_p(\kappa_T N^{1/2}T^{-1})=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}e_{it-1}\Delta e_{it}+o_p\left(1\right);$

(iii) $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}e_{it-1}\Delta e_{it}\xrightarrow{d}\mathcal{N}(0,\frac{1}{2}\overline{\sigma_\epsilon^4})\quad\text{as}\quad(N,T)\to\infty;$

(iv) $\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}e_{it}^2=\frac{1}{2}\overline{\sigma_\epsilon^2}+o_p(1).$

Given (i)–(iv), we have

$$\begin{aligned}\sqrt{N}T\left(\hat{\rho}-1\right) &= \left[\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}e_{it}^2+o_p\left(1\right)\right]^{-1}\left[\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}e_{it-1}\Delta e_{it}+o_p\left(1\right)\right]\\ &\xrightarrow{d}\mathcal{N}\left(0,2\left(\overline{\sigma_\epsilon^2}\right)^{-2}\overline{\sigma_\epsilon^4}\right).\end{aligned}$$

Below, we establish (i)–(iv) in order.

Step 1: We show (i).

Note that $\widehat{e}_{it}=e_{it}-e_{i1}+\boldsymbol{\lambda}_i^{0\prime}\boldsymbol{V}_t-\boldsymbol{d}_i'\widehat{\boldsymbol{F}}_t\equiv e_{it}+A_{it}$ where $\boldsymbol{V}_t=\sum_{s=2}^{t}\boldsymbol{v}_s$, $\boldsymbol{v}_t=\boldsymbol{H}^{-1}\widehat{\boldsymbol{B}}_t-\boldsymbol{f}_t^0$, $\boldsymbol{d}_i=\widehat{\boldsymbol{\lambda}}_i-\boldsymbol{H}'^{-1}\boldsymbol{\lambda}_i^0$, and $A_{it}=-e_{i1}+\boldsymbol{\lambda}_i^{0\prime}\boldsymbol{V}_t-\boldsymbol{d}_i'\widehat{\boldsymbol{F}}_t$. Then

$$\begin{aligned}\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\hat{e}_{it}^2 &= \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}e_{it}^2+\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}A_{it}^2+\frac{2}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}e_{it}A_{it}\\ &\equiv \mathbb{Q}_1+\mathbb{Q}_2+2\mathbb{Q}_3.\end{aligned}$$

We show in (iv) below that $\mathbb{Q}_1=\frac{1}{2}\overline{\sigma_\epsilon^2}+o_p(1)$. We will show below that $\mathbb{Q}_2=o_p\left(1\right)$. Then $\mathbb{Q}_3\leq\left(\mathbb{Q}_1\mathbb{Q}_2\right)^{1/2}=o_p\left(1\right)$ and the result in (i) follows. For $\mathbb{Q}_2$, we have

$$\begin{aligned}\mathbb{Q}_2 &= \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(e_{i1}-\boldsymbol{\lambda}_i^{0\prime}\boldsymbol{V}_t+\boldsymbol{d}_i'\widehat{\boldsymbol{F}}_t\right)^2\\ &\leq 3\left(\frac{1}{NT}\sum_{i=1}^{N}e_{i1}^2+\frac{1}{N}\sum_{i=1}^{N}\left\|\boldsymbol{\lambda}_i^0\right\|^2\frac{1}{T^2}\sum_{t=2}^{T}\left\|\boldsymbol{V}_t\right\|^2+\frac{1}{N}\sum_{i=1}^{N}\left\|\boldsymbol{d}_i\right\|^2\frac{1}{T^2}\sum_{t=2}^{T}\left\|\hat{\boldsymbol{F}}_t\right\|^2\right)\\ &\equiv 3\left(\mathbb{Q}_{2,1}+\mathbb{Q}_{2,2}+\mathbb{Q}_{2,3}\right).\end{aligned}$$

Under the null and Assumption A2.1(a), $\mathbb{Q}_{2,1}=\frac{1}{NT}\sum_{i=1}^{N}e_{i1}^2=O_p\left(T^{-1}\right)$ by Assumption A2.3(c) and Markov inequality. By Lemma B.1.3 and Assumption A2.2,

$$\mathbb{Q}_{2,2}=\frac{1}{N}\sum_{i=1}^{N}\left\|\boldsymbol{\lambda}_i^0\right\|^2\frac{1}{T^2}\sum_{t=2}^{T}\left\|\boldsymbol{V}_t\right\|^2=O_p\left(N^{p-1}\right)O_p\left(N^{-p}\right)=O_p\left(N^{-1}\right).$$

Following Lemma B.1.2(c), we can readily show that

$$\frac{1}{N}\sum_{i=1}^{N}\left\|\boldsymbol{d}_i\right\|^2=O_p\left(\mathcal{J}_B^{-2}\right),$$

where we use the fact that $\mathcal{J}_B^{-1}N^{-p/2}T^{1/2}=o(1)$ for a general choice of $\mathcal{J}_B$ such as $T^{1/2}$ or T . Using $\widehat{\boldsymbol{F}}_t=\sum_{s=1}^{t}\widehat{\boldsymbol{B}}_s=\sum_{s=1}^{t}(\widehat{\boldsymbol{B}}_s-\boldsymbol{H}\boldsymbol{f}_s^0)+\boldsymbol{H}\sum_{s=1}^{t}\boldsymbol{f}_s^0$, we have

$$\begin{aligned}
&\frac{1}{T^2}\sum_{t=1}^{T}\left\|\hat{\boldsymbol{F}}_t\right\|^2\\
\leq\ &\frac{2}{T^2}\sum_{t=1}^{T}\left\|\sum_{s=1}^{t}\left(\widehat{\boldsymbol{B}}_s-\boldsymbol{H}\boldsymbol{f}_s^0\right)\right\|^2+\frac{2}{T^2}\sum_{t=1}^{T}\left\|\sum_{s=1}^{t}\boldsymbol{H}\boldsymbol{f}_s^0\right\|^2\\
\leq\ &\left\|\boldsymbol{H}\right\|^2\frac{2}{T^2}\sum_{t=1}^{T}\max_{1\leq t\leq T}\left\|\sum_{s=1}^{t}\left(\boldsymbol{H}^{-1}\widehat{\boldsymbol{B}}_s-\boldsymbol{f}_s^0\right)\right\|^2+\frac{2}{T^2}\sum_{t=1}^{T}\left\|\sum_{s=1}^{t}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_s^0\right\|^2\left\|\boldsymbol{H}\left(\boldsymbol{\rho}_0^0\right)^{T}\right\|^2\\
=\ &O_p\left(\mathcal{J}_B^2[\left(\rho_{01}^0\right)^{-2T}+\ldots+\left(\rho_{0R_0}^0\right)^{-2T}]\right)O_p\left(N^{-p}\right)+O_p\left(T^{-1}\kappa_T\right)O_p\left(\mathcal{J}_B^2\right)\\
=\ &O_p\left(\mathcal{J}_B^2T^{-1}\kappa_T\right),
\end{aligned}$$

where the first equality follows from (B.2.14),(B.2.10), and Lemma B.1.3, $\left\|\sum_{s=1}^{t}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_s^0\right\|^2\leq\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\kappa_T^{-1}\sum_{s=2}^{T}diag\left(c_1,\ldots,c_{R_0}\right)\boldsymbol{F}_{s-1}^0\right\|^2$ + the dominated term $=O_p\left(\kappa_T\right)$ by following proofs of Lemma B.3.2(b) as proofs of Lemma A5(b) in Horie and Yamamoto (2016), and Assumption A2.1(b) and (d) when $\boldsymbol{F}_t$ follows mildly explosive processes. Then

$$\begin{aligned}
\mathbb{Q}_{2,3}\ &=\ \frac{1}{N}\sum_{i=1}^{N}\left\|\boldsymbol{d}_i\right\|^2\frac{1}{T^2}\sum_{t=2}^{T}\left\|\hat{\boldsymbol{F}}_t\right\|^2\\
&=\ O_p\left(\mathcal{J}_B^{-2}\right)O_p\left(\mathcal{J}_B^2T^{-1}\kappa_T\right)=O_p(T^{-1}\kappa_T)=o_p(1).
\end{aligned}$$

Consequently, $\mathbb{Q}_2=o_p(1)$.

Step 2: We show (ii). $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}\widehat{e}_{it-1}\Delta\widehat{e}_{it}=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}e_{it-1}\Delta e_{it}+o_p(1)$

As in Bai and Ng (2010), we can readily show that

$$\frac{1}{T}\sum_{t=2}^{T}\left(\widehat{e}_{it-1}\Delta\widehat{e}_{it}-\frac{1}{T}\sum_{t=2}^{T}e_{it-1}\Delta e_{it}\right)=\frac{1}{2T}\left(e_{iT}^2-\widehat{e}_{iT}^2\right)-\frac{1}{2T}\left(e_{i1}^2-\widehat{e}_{i1}^2\right)-\frac{1}{2T}\sum_{t=2}^{T}\left(\Delta e_{it}\right)^2-\left(\Delta\widehat{e}_{it}\right)^2.$$

Then it suffices to prove (ii) by establishing the follow claims:

(ii1) $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left(\widehat{e}_{i1}^2-e_{i1}^2\right)=O_p\left(N^{1/2}T^{-1}\right)=o_p(1)$,

(ii2) $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left(\widehat{e}_{iT}^2-e_{iT}^2\right)=O_p(N^{1/2}T^{-1}+N^{-1/2}+T^{-1/2})+O_p\left(\kappa_T N^{1/2}T^{-1}\right)=o_p(1)$,

(ii3) $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}\left[\left(\Delta\widehat{e}_{it}\right)^2-\left(\Delta e_{it}\right)^2\right]=O_p(N^{1/2}T^{-1}+N^{-1/2})=o_p(1)$.

(ii1) follows immediately from the proof of Lemma 3 in Bai and Ng (2010). For (ii2), recalling that $\hat{e}_{iT}=e_{iT}+A_{iT}$, we have

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left(\widehat{e}_{iT}^2-e_{iT}^2\right)=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}A_{iT}^2+\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}A_{iT}e_{iT}\equiv B_1+2B_2.$$

For B_1, we have

$$\begin{aligned}B_1 &\leq \frac{3}{\sqrt{NT}}\sum_{i=1}^{N}e_{i1}^2+\frac{3}{\sqrt{NT}}\sum_{i=1}^{N}\left\|\boldsymbol{\lambda}_i^0\right\|^2\left\|\boldsymbol{V}_T\right\|^2+\frac{3}{\sqrt{NT}}\sum_{i=1}^{N}\left\|\boldsymbol{d}_i\right\|^2\left\|\hat{\boldsymbol{F}}_T\right\|^2\\ &\equiv 3\left(B_{1,1}+B_{1,2}+B_{1,3}\right).\end{aligned}$$

Note that $B_{1,1}=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}e_{i1}^2=O_p\left(N^{1/2}T^{-1}\right)$ by Markov inequality. By Assumption A2.2 and the fact that $\left\|\boldsymbol{V}_T\right\|^2/T=O_p\left(N^{-p}\right)$ due to Lemma B.1.3,

$$B_{1,2}=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\|\boldsymbol{\lambda}_i^0\right\|^2\frac{1}{T}\left\|\boldsymbol{V}_T\right\|^2=O_p\left(N^{p-1/2}\right)O_p\left(N^{-p}\right)=O_p\left(N^{-1/2}\right).$$

where the final equality holds under Assumption A2.5 and A2.7. In addition, $\sum_{i=1}^{N}\left\|\boldsymbol{d}_i\right\|^2=O_p\left(N\mathcal{J}_B^{-2}\right)$

by similar arguments as used in the proof of Lemma B.1.2, note that for $\hat{\boldsymbol{F}}_T$, we have

$$
\begin{aligned}
\left\|\hat{\boldsymbol{F}}_T\right\|^2 &\leq 2\left\|\sum_{s=1}^T\left(\widehat{\boldsymbol{B}}_s-\boldsymbol{H}\boldsymbol{f}_s^0\right)\right\|^2+2\left\|\sum_{s=1}^T\boldsymbol{f}_s^0\right\|^2\|\boldsymbol{H}\|^2 \\
&\leq 2\|\boldsymbol{H}\|^2\max_{1\leq t\leq T}\left\|\sum_{s=1}^t\left(\boldsymbol{H}^{-1}\widehat{\boldsymbol{B}}_s-\boldsymbol{f}_s^0\right)\right\|^2+2\left\|\sum_{s=1}^T\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_s^0\right\|^2\left\|\boldsymbol{H}\left(\boldsymbol{\rho}_0^0\right)^T\right\|^2 \\
&= O_p\left(N^{-p}T\mathcal{J}_B^2\left[\left(\rho_{01}^0\right)^{-2T}+\ldots+\left(\rho_{0R_0}^0\right)^{-2T}\right]\right)+O_p\left(\mathcal{J}_B^2\kappa_T\right)=O_p\left(\mathcal{J}_B^2\kappa_T\right)
\end{aligned}
$$

where the first inequality holds by the fact that $(a+b)^2\leq 2a^2+2b^2$, and first equality holds by Lemma B.1.3, the construction of $\boldsymbol{H}$ in (B.2.10) and Lemma B.3.2(b), and the final equality holds when $\boldsymbol{F}_t$ follows mildly explosive processes under Assumption A2.5.

$$
B_{1,3} = \frac{1}{\sqrt{N}}\sum_{i=1}^N\|\boldsymbol{d}_i\|^2\frac{1}{T}\left\|\hat{\boldsymbol{F}}_T\right\|^2=O_p\left(N^{1/2}\mathcal{J}_B^{-2}\right)O_p\left(T^{-1}\mathcal{J}_B^2\kappa_T\right)=O_p\left(N^{1/2}T^{-1}\kappa_T\right)
$$

Thus, $B_{1,3}=o_p(1)$ under Assumption A2.7(a). In sum, $B_1=o_p(1)$.

Next, we study B_2. We make the following decomposition:

$$
\begin{aligned}
B_2 &= \frac{-1}{\sqrt{NT}}\sum_{i=1}^N e_{iT}e_{i1}+\frac{1}{\sqrt{NT}}\sum_{i=1}^N e_{iT}\boldsymbol{\lambda}_i^{0\prime}\boldsymbol{V}_T-\frac{1}{\sqrt{NT}}\sum_{i=1}^N e_{iT}\boldsymbol{d}_i'\widehat{\boldsymbol{F}}_T \\
&\equiv -B_{2,1}+B_{2,2}-B_{2,3}.
\end{aligned}
$$

Following the proof of Lemma 3 in Bai and Ng (2010), we can readily show that $B_{2,1}=O_p\left(T^{-1/2}\right)$. For $B_{2,2}$, we have

$$
B_{2,2}\leq\frac{1}{\sqrt{NT}}\left\|\sum_{i=1}^N\sum_{t=1}^T\boldsymbol{\lambda}_i^0\epsilon_{it}\right\|\frac{\|\boldsymbol{V}_T\|}{\sqrt{T}}=O_p(N^{-(1-p)/2})O_p\left(N^{-p/2}\right)=O_p\left(N^{-1/2}\right),
$$

where we use Lemma B.1.3 and the fact that $\left\|\sum_{i=1}^N\sum_{t=1}^T\boldsymbol{\lambda}_i^0\epsilon_{it}\right\|=O_p\left(N^{p/2}T^{1/2}\right)$ by (B.2.4). We have shown $||\widehat{\boldsymbol{F}}_T|| = O_p(\mathcal{J}_B\kappa_T^{1/2})$. Below, we will show that

$$\begin{aligned}
&\left\|\frac{1}{N}\sum_{i=1}^N e_{iT}\boldsymbol{d}_i\right\| \\
=\ & O_p\left(\mathcal{J}_B^{-1}N^{-1-p/2}T^{3/2}\left[\left(\rho_{01}^0\right)^{-T}+\ldots+\left(\rho_{0R_0}^0\right)^{-T}\right]\right)+O_p\left(\mathcal{J}_B^{-1}\left(N^{-1}T^{1/2}+\kappa_T^{1/2}\right)\right) \\
=\ & O_p\left(\mathcal{J}_B^{-1}\left(N^{-1}T^{1/2}+\kappa_T^{1/2}\right)\right), \qquad \text{(B.1.2)}
\end{aligned}$$

under Assumptions A2.5 and A2.7. Then, under Assumptions A2.7(a), we have

$$\begin{aligned}
B_{2,3} \ &\leq\ \sqrt{N}\left\|\frac{1}{N}\sum_{i=1}^N e_{iT}\boldsymbol{d}_i'\right\|\frac{1}{T}\left\|\widehat{\boldsymbol{F}}_T\right\| = N^{1/2}O_p\left(\mathcal{J}_B^{-1}\left(N^{-1}T^{1/2}+\kappa_T^{1/2}\right)\right)O_p(T^{-1}\mathcal{J}_B\kappa_T^{1/2}) \\
&=\ O_p\left(N^{-1/2}T^{-1/2}\kappa_T^{1/2}+\kappa_T N^{1/2}T^{-1}\right) \\
&=\ o_p(1).
\end{aligned}$$

It follows that $B_2 = O_p\left(N^{-1/2}+T^{-1/2}+\kappa_T N^{1/2}T^{-1}\right) = o_p(1)$ and $\frac{1}{\sqrt{NT}}\sum_{i=1}^N(\widehat{e}_{iT}^2 - e_{iT}^2) = o_p(1)$ under Assumption A2.7(a).

Now, we show (B.1.2). Noting that

$$\begin{aligned}
\boldsymbol{d}_i \ &=\ \widehat{\boldsymbol{\lambda}}_i - \boldsymbol{H}'^{-1}\boldsymbol{\lambda}_i^0 \\
&=\ \mathcal{J}_B^{-2}\boldsymbol{H}\left(\widehat{\boldsymbol{B}}'\boldsymbol{H}^{-1}-\boldsymbol{f}^0\right)'\left(\boldsymbol{f}^0-\widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1}\right)\boldsymbol{\lambda}_i^0+\mathcal{J}_B^{-2}\boldsymbol{H}\boldsymbol{f}^{0\prime}\left(\boldsymbol{f}^0-\widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1}\right)\boldsymbol{\lambda}_i^0 \\
&\quad +\mathcal{J}_B^{-2}\boldsymbol{H}\boldsymbol{f}^{0\prime}\boldsymbol{\epsilon}_i+\mathcal{J}_B^{-2}\left(\widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1}-\boldsymbol{f}^0\right)'\boldsymbol{\epsilon}_i \\
&\equiv\ \boldsymbol{d}_{i,1}+\boldsymbol{d}_{i,2}+\boldsymbol{d}_{i,3}+\boldsymbol{d}_{i,4} \qquad \text{(B.1.3)}
\end{aligned}$$

we have

$$\frac{1}{N}\sum_{i=1}^N\boldsymbol{d}_i e_{iT} = \frac{1}{N}\sum_{i=1}^N\left(\boldsymbol{d}_{i,1}+\boldsymbol{d}_{i,2}+\boldsymbol{d}_{i,3}+\boldsymbol{d}_{i,4}\right)e_{iT} \equiv \mathcal{D}_{1NT}+\mathcal{D}_{2NT}+\mathcal{D}_{3NT}+\mathcal{D}_{4NT}.$$

We bound $\mathcal{D}_{lNT}$ for $l=1,2,3,4$ in turn by using $e_{it}=\sum_{s=1}^{t}\epsilon_{is}$ under the null. First,

$$\begin{aligned}
&\|\mathcal{D}_{1NT}\| \\
\leq\;& \mathcal{J}_B^{-2}N^{-1}\left\|\widehat{\boldsymbol{B}}-\boldsymbol{f}^0\boldsymbol{H}'\right\|\left\|\boldsymbol{f}^0-\widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1}\right\|\left\|\sum_{i=1}^{N}\boldsymbol{\lambda}_i^0 e_{iT}\right\| \\
=\;& \mathcal{J}_B^{-2}N^{-1}O_p\left(\mathcal{J}_B N^{-p/2}T^{1/2}\left[\left(\rho_{01}^0\right)^{-T}+\ldots+\left(\rho_{0R_0}^0\right)^{-T}\right]\right)O_p\left(N^{-p/2}T^{1/2}\right)O_p\left(N^{p/2}T^{1/2}\right) \\
=\;& O_p\left(\mathcal{J}_B^{-1}N^{-1-p/2}T^{3/2}\left[\left(\rho_{01}^0\right)^{-T}+\ldots+\left(\rho_{0R_0}^0\right)^{-T}\right]\right),
\end{aligned}$$

where the first equality follows by (B.2.15), Lemma B.1.2(a) and the fact that $E\left\|\sum_{i=1}^{N}\boldsymbol{\lambda}_i^0 e_{iT}\right\|^2=O\left(N^pT\right)$ by Assumptions A2.1(c), A2.2 and A2.4.

For $\mathcal{D}_{2NT}$, we have $\mathcal{D}_{2NT}=\bar{\mathcal{D}}_{2NT}\sum_{i=1}^{N}\boldsymbol{\lambda}_i^0e_{iT}$, where $\sum_{i=1}^{N}\boldsymbol{\lambda}_i^0e_{iT}=O_p\left(N^{p/2}T^{1/2}\right)$ as studied above, and $\bar{\mathcal{D}}_{2NT}=N^{-1}\mathcal{J}_B^{-2}\boldsymbol{H}\boldsymbol{f}^{0\prime}(\widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1}-\boldsymbol{f}^0)$. For $\bar{\mathcal{D}}_{2NT}$, we adopt the decomposition of $\widehat{\boldsymbol{B}}-\boldsymbol{f}^0\boldsymbol{H}'$ in (B.2.13) to obtain

$$\begin{aligned}
&\bar{\mathcal{D}}_{2NT} \\
=\;& \mathcal{J}_B^{-2}N^{-p-1}\left[\boldsymbol{H}\boldsymbol{f}^{0\prime}\left(\boldsymbol{\Lambda}^0\boldsymbol{u}'+\boldsymbol{\epsilon}\right)'\left(\boldsymbol{\Lambda}^0\boldsymbol{u}+\boldsymbol{\epsilon}\right)\widehat{\boldsymbol{B}}+\boldsymbol{H}\boldsymbol{f}^{0\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}+\boldsymbol{H}\boldsymbol{f}^{0\prime}\boldsymbol{\epsilon}'\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}\right] \\
\equiv\;& \mathcal{D}_{2NTa}+\mathcal{D}_{2NTb}+\mathcal{D}_{2NTc}.
\end{aligned}$$

Following the proof of Lemma B.1.2(c), it easy to argue that $\mathcal{D}_{2NTb}$ and $\mathcal{D}_{2Ntc}$ dominate $\mathcal{D}_{2NTa}$. So we focus on $\mathcal{D}_{2NTb}$ and $\mathcal{D}_{2NTc}$ below. Note that the leading term in $\mathcal{D}_{2NTb}$ is given by $\mathcal{D}_{2NTb1}=\mathcal{J}_B^{-2}N^{-p-1}\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}\boldsymbol{f}^0\boldsymbol{H}'\mathcal{J}_T^{-1}$ and

$$\begin{aligned}
\|\mathcal{D}_{2NTb1}\| =\;& \mathcal{J}_B^{-2}N^{-p-1}\left\|\boldsymbol{H}\boldsymbol{f}^{0\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}\boldsymbol{f}^0\boldsymbol{H}'\mathcal{J}_T^{-1}\right\| \\
\leq\;& \mathcal{J}_B^{-2}N^{-1}\left\|\mathcal{J}_T^{-1}N^{-p}\sum_{j=1}^{N}\sum_{s=2}^{T}\boldsymbol{H}\boldsymbol{f}_s^0\epsilon_{js}\boldsymbol{\lambda}_j^{0\prime}\right\|\left\|\sum_{t=2}^{T}\boldsymbol{B}_t^0\boldsymbol{B}_t^{0\prime}\boldsymbol{H}'\right\| \\
=\;& \mathcal{J}_B^{-2}N^{-1}O_p\left(N^{-p/2}\left[\left(\rho_{01}^0\right)^{-T}+\ldots+\left(\rho_{0R_0}^0\right)^{-T}\right]\right)O_p\left(\mathcal{J}_B\left[\left(\rho_{01}^0\right)^{T}+\ldots+\left(\rho_{0R_0}^0\right)^{T}\right]\right) \\
=\;& O_p\left(\mathcal{J}_B^{-1}N^{-1-p/2}\right),
\end{aligned}$$

where the first equality follows because we can show that $\left\| \mathcal{J}_T^{-1} N^{-p} \sum_{j=1}^N \sum_{s=2}^T \boldsymbol{H} \boldsymbol{f}_s^0 \epsilon_{js} \boldsymbol{\lambda}_j^{0\prime} \right\| = O_p\left(N^{-p/2}\left[\left(\rho_{01}^0\right)^{-T} + \ldots + \left(\rho_{0R_0}^0\right)^{-T}\right]\right)$ holds by the same arguments us used in the study of $\|\mathbb{A}_{2t1}\|$ in the proof of Lemma B.1.2(b), and the fact that $\left\| \sum_{t=2}^T \boldsymbol{B}_t^0 \boldsymbol{B}_t^{0\prime} \boldsymbol{H}' \right\| = O_p\left(\mathcal{J}_B\left[\left(\rho_{01}^0\right)^{T} + \ldots + \left(\rho_{0R_0}^0\right)^{T}\right]\right)$ by similar arguments as used to bound $\|\mathbb{A}_{3t1}\|$ in the proof of Lemma B.1.2(b). Similarly, the leading term in $\mathcal{D}_{2NTc}$ is given by $\mathcal{D}_{2NTc1} = \mathcal{J}_B^{-2} N^{-p-1} \boldsymbol{H} \boldsymbol{B}^{0\prime} \boldsymbol{\epsilon}' \boldsymbol{\Lambda}^0 \boldsymbol{B}^{0\prime} \boldsymbol{B}^0 \boldsymbol{H}' \mathcal{J}_T^{-1}$ and

$$
\begin{aligned}
\mathcal{D}_{2NTc1} &\leq \mathcal{J}_B^{-2} N^{-1} \left\| \mathcal{J}_T^{-1} N^{-p} \sum_{s=2}^T \boldsymbol{H} \boldsymbol{B}_s^0 \boldsymbol{B}_s^{0\prime} \right\| \left\| \sum_{j=1}^N \boldsymbol{\lambda}_j^0 \sum_{t=2}^T \epsilon_{jt} \boldsymbol{B}_t^{0\prime} \boldsymbol{H}' \right\| \\
&= \mathcal{J}_B^{-2} N^{-1} O_p(N^{-p}) O_p\left(\mathcal{J}_B N^{p/2}\right) \\
&= O_p\left(\mathcal{J}_B^{-1} N^{-1-p/2}\right),
\end{aligned}
$$

where the first equality follows by the fact that $\left\| \mathcal{J}_T^{-1} N^{-p} \sum_{s=2}^T \boldsymbol{H} \boldsymbol{B}_s^0 \boldsymbol{B}_s^{0\prime} \right\| = O_p(N^{-p})$ by following from the same arguments as used to bound $\|\mathbb{A}_{3t1}\|$ in the proof of Lemma A.2(b), and that $\left\| \sum_{j=1}^N \boldsymbol{\lambda}_j^0 \sum_t \epsilon_{jt} \boldsymbol{B}_t^{0\prime} \boldsymbol{H}' \right\| = O_p\left(\mathcal{J}_B N^{p/2}\right)$ by similar arguments to bound $\|\mathbb{A}_{1t3a}\|$ in the proof of Lemma B.1.2(b). Consequently, $\bar{\mathcal{D}}_{2NT} = O_p\left(\mathcal{J}_B^{-1} N^{-1-p/2}\right)$ abd

$$
\mathcal{D}_{2NT} = O_p\left(N^{p/2} T^{1/2}\right) O_p\left(\mathcal{J}_B^{-1} N^{-1-p/2}\right) = O_p\left(\mathcal{J}_B^{-1} N^{-1} T^{1/2}\right).
$$

For $\mathcal{D}_{3NT}$, we use $e_{iT} = \sum_{r=1}^T \epsilon_{ir}$ under the null to obtain

$$
\mathcal{D}_{3NT} = \mathcal{J}_B^{-2} \boldsymbol{H} \left(\boldsymbol{\rho}_0^0\right)^T \left(\frac{1}{N} \sum_{i=1}^N \sum_{r=2}^T \sum_{s=2}^T \left(\boldsymbol{\rho}_0^0\right)^{-T} \boldsymbol{f}_s^0 \epsilon_{is} \epsilon_{ir} \right) \equiv \mathcal{J}_B^{-2} \boldsymbol{H} \left(\boldsymbol{\rho}_0^0\right)^T \bar{\mathcal{D}}_{3NT}
$$

Under Assumption A2.1 and A2.4, we have $E(\bar{\mathcal{D}}_{3NT}) = 0$. In addition,

$$
\begin{aligned}
E\left\|\bar{\mathcal{D}}_{3NT}\right\|^2 &= E\left[\frac{1}{N^2} \sum_{i=1}^N \sum_{j \neq i}^N \sum_{r,s,q,v=2}^T \operatorname{tr}\left[\left(\boldsymbol{\rho}_0^0\right)^{-T} \boldsymbol{f}_s^0 \boldsymbol{f}_q^{0\prime} \left(\boldsymbol{\rho}_0^0\right)^{-T} \right] \epsilon_{is} \epsilon_{ir} \epsilon_{jq} \epsilon_{jv} \right] \\
&\quad + E\left[\frac{1}{N^2} \sum_{i=1}^N \sum_{r,s,q,v=2}^T \operatorname{tr}\left[\left(\boldsymbol{\rho}_0^0\right)^{-T} \boldsymbol{f}_s^0 \boldsymbol{f}_q^{0\prime} \left(\boldsymbol{\rho}_0^0\right)^{-T} \right] \epsilon_{is} \epsilon_{ir} \epsilon_{iq} \epsilon_{iv} \right] \\
&\equiv \mathbb{D}_{3NTa} + \mathbb{D}_{3NTb}.
\end{aligned}
$$

For $\mathbb{D}_{3NTa}$, we have

$$\begin{aligned}\mathbb{D}_{3NTa} &= E\left[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{s,q=2}^{T}\text{tr}\left[\left(\boldsymbol{\rho}_0^0\right)^{-T}E(\boldsymbol{f}_s^0\boldsymbol{f}_q^{0\prime})\left(\boldsymbol{\rho}_0^0\right)^{-T}\right]E\left(\epsilon_{is}^2\epsilon_{jq}^2\right)\right] \\ &= \left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sigma_i^2\sigma_j^2\right)\sum_{s,q=2}^{T}\text{tr}\left[\left(\boldsymbol{\rho}_0^0\right)^{-T}E(\boldsymbol{f}_s^0\boldsymbol{f}_q^{0\prime})\left(\boldsymbol{\rho}_0^0\right)^{-T}\right] = O\left(\kappa_T\right),\end{aligned}$$

where the first equality holds by Assumption A2.3(c) and the law of iterated expectations, the second equality holds by Assumption A2.1(c) and the law of iterated expectations, and the final equality holds because

$$\begin{aligned}\sum_{s,q=2}^{T}\text{tr}\left[\left(\boldsymbol{\rho}_0^0\right)^{-T}E(\boldsymbol{f}_s^0\boldsymbol{f}_q^{0\prime})\left(\boldsymbol{\rho}_0^0\right)^{-T}\right] &= E\left[\sum_{s,q=2}^{T}\boldsymbol{f}_q^{0\prime}\left(\boldsymbol{\rho}_0^0\right)^{-T}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_s^0\right] \\ &\leq E\left\|\sum_{s=2}^{T}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_s^0\right\|^2 = O\left(\kappa_T\right), \qquad \text{(B.1.4)}\end{aligned}$$

where the first equality holds by the definition of the trace operator, and the last equality holds because $\left\|\sum_{s=2}^{t}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_s^0\right\| \leq \left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\kappa_T^{-1}\sum_{s=2}^{T}diag\left(c_1,\ldots,c_{R_0}\right)\boldsymbol{F}_{s-1}^0\right\|$+the dominated term $= O_p\left(\kappa_T\right)$ by following proofs of Lemma B.3.2(b) as proofs of Lemma A5(b) in Horie and Yamamoto (2016).

For $\mathbb{D}_{3NTb}$, we have

$$\begin{aligned}\mathbb{D}_{3NTb} &= \frac{2}{N^2}\sum_{i=1}^{N}\sum_{s\neq q=2}^{T}\text{tr}\left[\left(\boldsymbol{\rho}_0^0\right)^{-T}E(\boldsymbol{f}_s^0\boldsymbol{f}_q^{0\prime})\left(\boldsymbol{\rho}_0^0\right)^{-T}\right]E\left(\epsilon_{is}^2\epsilon_{iq}^2\right) \\ &\quad +\frac{1}{N^2}\sum_{i=1}^{N}\sum_{s\neq r=2}^{T}\text{tr}\left[\left(\boldsymbol{\rho}_0^0\right)^{-T}E(\boldsymbol{f}_s\boldsymbol{f}_s^{\prime})\left(\boldsymbol{\rho}_0^0\right)^{-T}\right]E\left(\epsilon_{is}^2\epsilon_{ir}^2\right) \\ &\quad +\frac{1}{N^2}\sum_{i=1}^{N}\sum_{s=1}^{T}\text{tr}\left[\left(\boldsymbol{\rho}_0^0\right)^{-T}E(\boldsymbol{f}_s^0\boldsymbol{f}_s^{0\prime})\left(\boldsymbol{\rho}_0^0\right)^{-T}\right]E\left(\epsilon_{is}^4\right) \\ &\leq \left\{\frac{1}{N^2}\sum_{i=1}^{N}[3T\sigma_i^4+E\left(\epsilon_{is}^4\right)]\right\}E\sum_{s=1}^{T}\text{tr}\left[\left(\boldsymbol{\rho}_0^0\right)^{-T}E(\boldsymbol{f}_s^0\boldsymbol{f}_s^{0\prime})\left(\boldsymbol{\rho}_0^0\right)^{-T}\right] \\ &= O\left(N^{-1}T\right)O\left(1\right) = O\left(N^{-1}T\right),\end{aligned}$$

where the first equality holds by the m.d.s property of the sequence $\{\epsilon_{it}\}$ and the law of iterated expectations, the inequality holds by (B.1.4), and the last equality holds by the proof of Lemma B.3.2(e). It follows that $E\left\|\bar{\mathcal{D}}_{3NT}\right\|^2 = O\left(N^{-1}T+\kappa_T\right)$ and

$$
\begin{aligned}
\left\|\mathcal{D}_{3NT}\right\| &\leq \mathcal{J}_B^{-2}\left\|\boldsymbol{H}\left(\boldsymbol{\rho}_0^0\right)^T\right\|\left\|\bar{\mathcal{D}}_{3NT}\right\| \\
&= \mathcal{J}_B^{-2}O_p\left(\mathcal{J}_B\right)O_p\left(N^{-1/2}T^{1/2}+\kappa_T^{1/2}\right) = O_p\left(\mathcal{J}_B^{-1}\left(N^{-1/2}T^{1/2}+\kappa_T^{1/2}\right)\right) \qquad \text{(B.1.5)}
\end{aligned}
$$

where the first equality follows from (B.2.10). In addition, it is easy to show that $\mathcal{D}_{4NT}$ is dominated by $\mathcal{D}_{3NT}$.

In sum, we have

$$
\begin{aligned}
& \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{d}_i e_{iT} \\
=\ & O_p\left(\mathcal{J}_B^{-1}N^{-1-p/2}T^{3/2}\left[\left(\rho_{01}^0\right)^{-T}+\ldots+\left(\rho_{0R_0}^0\right)^{-T}\right]\right)+O_p\left(\mathcal{J}_B^{-1}N^{-1}T^{1/2}\right)+O_p\left(\mathcal{J}_B^{-1}\kappa_T^{1/2}\right) \\
=\ & O_p\left(\mathcal{J}_B^{-1}\kappa_T^{1/2}\right). \qquad \text{(B.1.6)}
\end{aligned}
$$

Now, we prove (ii3). Noting that $\Delta\widehat{e}_{it} = \Delta e_{it} - \left(\boldsymbol{\lambda}_i^{0\prime}\boldsymbol{v}_t + \boldsymbol{d}_i^{\prime}\hat{\boldsymbol{B}}_t\right) \equiv \Delta e_{it} - r_{it}$, we have

$$
\begin{aligned}
\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=2}^{T}\left[\left(\Delta\widehat{e}_{it}\right)^2-\left(\Delta e_{it}\right)^2\right] &= \frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=2}^{T}r_{it}^2-\frac{2}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\Delta e_{it}\right)r_{it} \\
&\equiv D_1 - 2D_2.
\end{aligned}
$$

By Lemma B.1.4, it is readily to see that $D_1 = N^{-1/2}T^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}r_{it}^2 = N^{-1/2}T^{-1}||\boldsymbol{\Lambda}^0\boldsymbol{f}^{0\prime} - \widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{B}}^{\prime}||^2 = N^{-1/2}T^{-1}O_p\left(N+T\right) = o_p\left(1\right)$ holds. For D_2, we have

$$
\begin{aligned}
D_2 &= \frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta e_{it}\boldsymbol{\lambda}_i^{0\prime}\boldsymbol{v}_t+\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta e_{it}\boldsymbol{d}_i^{\prime}\boldsymbol{H}\boldsymbol{f}_t^0+\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta e_{it}\boldsymbol{d}_i^{\prime}(\hat{\boldsymbol{B}}_t-\boldsymbol{H}\boldsymbol{f}_t^0) \\
&\equiv D_{2,1}+D_{2,2}+D_{2,3}.
\end{aligned}
$$

Noting that $\Delta e_{it} = \epsilon_{it}$ under the null, by the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
\|D_{2,1}\| &\leq \left(\frac{1}{T}\sum_{t=2}^{T}\left\|N^{-1/2}\sum_{i=1}^{N}\Delta e_{it}\boldsymbol{\lambda}_i^0\right\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=2}^{T}\|\boldsymbol{v}_t\|^2\right)^{1/2} \\
&= \left(\frac{1}{NT}\sum_{t=2}^{T}\left\|\sum_{i=1}^{N}\boldsymbol{\lambda}_i^0\epsilon_{it}\right\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=2}^{T}\left\|\boldsymbol{H}^{-1}\widehat{\boldsymbol{B}}_t - \boldsymbol{f}_t^0\right\|^2\right)^{1/2} \\
&= O_p\left(N^{(p-1)/2}\right)O_p\left(N^{p/2}\right) = O_p\left(N^{-1/2}\right),
\end{aligned}$$

where the second equality holds by (B.2.4), and Lemma B.1.2(b). Similarly,

$$\begin{aligned}
\|D_{2,2}\| &= \frac{1}{\sqrt{N}T}\left\|\sum_{i=1}^{N}\sum_{t=2}^{T}\boldsymbol{d}_i'\boldsymbol{H}\left(\boldsymbol{\rho}_0^0\right)\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_t^0\epsilon_{it}\right\| \\
&\leq T^{-1}\left(\sum_{i=1}^{N}\|\boldsymbol{d}_i\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\sum_{t=2}^{T}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_t^0\epsilon_{it}\right\|^2\right)^{1/2}\left\|\left(\boldsymbol{\rho}_0^0\right)^{T}\boldsymbol{H}\right\| \\
&= T^{-1}O_p\left(N^{1/2}\mathcal{J}_B^{-1}\right)O_p(1)\,O_p(\mathcal{J}_B) = O_p\left(N^{1/2}T^{-1}\right).
\end{aligned}$$

where the second equality holds by the fact that $\sum_{i=1}^{N}\|\boldsymbol{d}_i\|^2 = O_p\left(N\mathcal{J}_B^{-2}\right)$ by Lemma B.1.2, that $\frac{1}{N}\sum_{i=1}^{N}\left\|\sum_{t=2}^{T}\epsilon_{it}\boldsymbol{f}_t\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|^2 = O_p(1)$ by similar arguments as used to obtain Lemma B.3.2(c) (see Horie and Yamamoto (2016)), and that $\left\|\left(\boldsymbol{\rho}_0^0\right)^{T}\boldsymbol{H}\right\| = O_p(\mathcal{J}_B)$ as in (B.2.10). In addition, we can readily argue that $D_{2,3}$ is dominated by $D_{2,2}$. Then $D_2 = O_p\left(N^{-1/2} + N^{1/2}T^{-1}\right)$ and $\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=2}^{T}\left[(\Delta\widehat{e}_{it})^2 - (\Delta e_{it})^2\right] = o_p(1)$.

Step 3: We show (iii).

Let $\eta_{Nt} = \sum_{s=1}^{t-1}\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\epsilon_{is}\epsilon_{it}$. Then $Z_{NT} \equiv \frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\epsilon_{is}\epsilon_{it} = \sum_{t=2}^{T}\eta_{Nt}$. Note that $\Delta e_{it} = \epsilon_{it}$ and $e_{i,t-1} = \sum_{s=1}^{t-1}\epsilon_{is}$ under the null. Under Assumption A2.1(c), $E\left(\eta_{Nt} \mid \mathcal{F}_{NT,t-1}^{\epsilon}\right) = 0$ by the law of iterated expectations and we can resort to the martingale central limit theorem (CLT) by verifying the conditions:

(1) $s_T^2 \equiv E\left[\left(\sum_{t=2}^{T}\eta_{Nt}\right)^2\right] \to \frac{1}{2}\overline{\sigma_\epsilon^4}$ where $\overline{\sigma_\epsilon^4} = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sigma_i^4$;

(2) $\left(s_T^2\right)^{-1}\sum_{t=2}^{T}E\left[\eta_{Nt}^2\left(|\eta_{Nt}| > \delta s_T\right) \mid \mathcal{F}_{NT,t-1}^{\epsilon}\right] \to 0$ for all $\delta > 0$;

(3) $\left(s_T^2\right)^{-1} V_T \to 1$ where $V_T \equiv \sum_{t=2}^{T} E\left[\eta_{Nt}^2 \mid \mathcal{F}_{NT,t-1}^{\epsilon}\right]$.

Gvien the above conditions, the martingale CLT implies that $Z_{NT} = \sum_{t=2}^{T} \eta_{Nt} \xrightarrow{d} \mathcal{N}(0, \frac{1}{2}\overline{\sigma_\epsilon^4})$ as (N,T) go to infinity jointly. We now verify conditions (1)–(3) in order.

First, by the repeated use of Assumption A2.1(c) and the law of iterated expectations, we have

$$\begin{aligned}
s_T^2 &= E\left[\sum_{t=2}^{T}\sum_{r=2}^{T}\eta_{Nt}\eta_{Nr}\right] = \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=2}^{T}\sum_{r=2}^{T}\sum_{s=1}^{t-1}\sum_{q=1}^{r-1} E\left(\epsilon_{is}\epsilon_{it}\epsilon_{jr}\epsilon_{jq}\right) \\
&= \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\sum_{q=1}^{t-1} E\left(\epsilon_{is}\epsilon_{iq}\epsilon_{it}^2\right) = \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\sum_{q=1}^{t-1} E\left[E\left(\epsilon_{it}^2|\mathcal{F}_{NT,t-1}^{\epsilon}\right)\epsilon_{is}\epsilon_{iq}\right] \\
&= \frac{1}{NT^2}\sum_{i=1}^{N}\sigma_i^2\sum_{t=2}^{T}\sum_{s=1}^{t-1} E\left(\epsilon_{is}^2\right) = \frac{T-1}{2TN}\sum_{i=1}^{N}\sigma_i^4 \to \frac{1}{2}\overline{\sigma_\epsilon^4},
\end{aligned}$$

where $\overline{\sigma_\epsilon^4} \equiv \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sigma_i^4$. This verifies condition (1).

For condition (2) above, it suffices to verify that

$$\left(s_T^2\right)^{-1-\beta}\sum_{t=2}^{T} E\left|\eta_{Nt}\right|^{2+\beta} \to 0 \quad \text{as} \quad (N,T)\to\infty \text{ for some } \beta > 0.$$

Since $1/s_T^2 = O_p(1)$ as shown in condition (1), it is sufficient to show that $\sum_{t=2}^{T} E\left|\eta_{Nt}\right|^4 \to 0 \quad \text{as} \quad (N,T)\to\infty$. Note that

$$\begin{aligned}
\sum_{t=2}^{T} E\left|\eta_{Nt}\right|^4 &= \sum_{t=2}^{T} E\left[\sum_{s=1}^{t-1}\left(\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\epsilon_{is}\epsilon_{it}\right)\right]^4 \\
&= \frac{1}{N^2T^4}\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N}\sum_{1\le r,s,q,v\le t-1} E\left(\epsilon_{is}\epsilon_{it}\epsilon_{jr}\epsilon_{jt}\epsilon_{kq}\epsilon_{kt}\epsilon_{lv}\epsilon_{lt}\right) \\
&= \frac{1}{N^2T^4}\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{1\le r,s,q,v\le t-1} E\left(\epsilon_{is}\epsilon_{ir}\epsilon_{iq}\epsilon_{iv}\epsilon_{it}^4\right) \\
&\quad + \frac{3}{N^2T^4}\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{j=1,j\ne i}^{N}\sum_{1\le r,s,q,v\le t-1} E\left(\epsilon_{is}\epsilon_{ir}\epsilon_{it}^2\epsilon_{jq}\epsilon_{jv}\epsilon_{jt}^2\right) \\
&\equiv \mathcal{R}_1 + 3\mathcal{R}_2.
\end{aligned}$$

For $\mathcal{R}_1$, we have

$$
\begin{aligned}
\mathcal{R}_1 &= \frac{1}{N^2T^3}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=2}^{T}E\left[\left(\sum_{s=1}^{t-1}\epsilon_{is}\right)^4\epsilon_{it}^4\right] \\
&\leq \frac{1}{N^2T^3}\sum_{i=1}^{N}E\left\{\left[\frac{1}{T}\sum_{t=2}^{T}\left(\sum_{s=1}^{t-1}\epsilon_{is}\right)^8\right]^{1/2}\left[\frac{1}{T}\sum_{t=2}^{T}\epsilon_{it}^8\right]^{1/2}\right\} \\
&\leq \frac{1}{N^2T^3}\sum_{i=1}^{N}\left\{\frac{1}{T}\sum_{t=2}^{T}E\left[\left(\sum_{s=1}^{t-1}\epsilon_{is}\right)^8\right]\right\}^{1/2}\left[\frac{1}{T}\sum_{t=2}^{T}E\left(\epsilon_{it}^8\right)\right]^{1/2} \\
&\leq \frac{M}{NT^3}\left\{\frac{1}{T}\sum_{t=2}^{T}t^4\right\}^{1/2} = O\left(N^{-1}T^{-1}\right),
\end{aligned}
$$

where the first and second inequalities holds by Cauchy-Schwarz and Jensen inequalities, and the last inequality holds by the fact that $\left|\sum_{1\leq s\leq t-1}\epsilon_{is}\right|^8 \leq Mt^4$ (see, e.g., Shao et al. (1996)) under Assumption A2.1(a). Similarly,

$$
\begin{aligned}
\mathcal{R}_2 &= \frac{1}{N^2T^4}\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\sum_{1\leq r,s,q,v\leq t-1}E\left(\epsilon_{is}\epsilon_{ir}\epsilon_{jq}\epsilon_{jv}\epsilon_{it}^2\epsilon_{jt}^2\right) \\
&= \frac{1}{N^2T^4}\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}E\left[\left(\sum_{s=1}^{t-1}\epsilon_{is}\right)^2\left(\sum_{q=1}^{t-1}\epsilon_{jq}\right)^2\epsilon_{it}^2\epsilon_{jt}^2\right] \\
&\leq \frac{1}{N^2T^3}\sum_{i,j=1}^{N}\left[\frac{1}{T}\sum_{t=2}^{T}E\left[\left(\sum_{s=1}^{t-1}\epsilon_{is}\right)^4\left(\sum_{q=1}^{t-1}\epsilon_{jq}\right)^4\right]\right]^{1/2}\left[\frac{1}{T}\sum_{t=2}^{T}E\left(\epsilon_{it}^4\epsilon_{jt}^4\right)\right]^{1/2} \\
&\leq \frac{M}{N^2T^3}\sum_{i,j=1}^{N}\left[\frac{1}{T}\sum_{t=2}^{T}t^4\right]^{1/2} = O\left(T^{-1}\right).
\end{aligned}
$$

It follows that $\sum_{t=2}^{T}E\left|\eta_{Nt}\right|^4 = O\left(T^{-1}\right) = o(1)$.

To verify condition (3), note that

$$\begin{aligned} V_T &= \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}E\left[\sum_{s=1}^{t-1}\sum_{r=1}^{t-1}\epsilon_{is}\epsilon_{ir}\epsilon_{it}^2|\mathcal{F}_{NT,t-1}^{\epsilon}\right] \\ &= \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\sum_{r=1}^{t-1}\epsilon_{is}\epsilon_{ir}E\left[\epsilon_{it}^2|\mathcal{F}_{NT,t-1}^{\epsilon}\right] \\ &= \frac{1}{NT^2}\sum_{i=1}^{N}\sigma_i^2\sum_{t=2}^{T}\sum_{s=1}^{t-1}\sum_{r=1}^{t-1}\epsilon_{is}\epsilon_{ir}. \end{aligned}$$

where we use the m.d.s property of $\{\epsilon_{it}\}$, the law of iterated expectations, and Assumption A2.1(c). By the the law of iterated expectations and the arguments as used to verify condition (1), $E(V_T) = s_T^2$. By direct moment calculations as in the verification of condition (2), we can show that Var$(V_T) = o(1)$. Then $V_T = s_T^2 + o_p(1)$ and condition (3) follows.

Step 4: We show (iv).

Under the null, $e_{it} = \sum_{s=1}^{t}\epsilon_{is}$. Then

$$A_{NT} \equiv \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}e_{it}^2 = \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\sum_{r=1}^{t-1}\epsilon_{is}\epsilon_{ir}.$$

By direct moment calculations, we can readily show that

$$E(A_{NT}) = \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=1}^{t-1}E\left(\epsilon_{is}^2\right) = \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\sigma_i^2 = \frac{1}{2}\overline{\sigma_\epsilon^2} + o(1)$$

where $\overline{\sigma_\epsilon^2} = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2$. As in Step 3, we can also show that Var$(A_{NT}) = O\left(T^{-1}\right)$. Then $A_{NT} = \frac{1}{2}\overline{\sigma_\epsilon^2} + o_p(1)$.

The Case of I(1)-type of Factors

Under Assumptions A2.1, when the factors follow a unit root process, the proof is a combination of similar arguments of Steps 1–4 in the last appendix and the corresponding proof given in Bai and Ng (2010). We omit the details for brevity.

B.1.3 Proof of Corollary 3.2

Case 1: Factors follow mildly explosive processes

We are going to show $\widehat{\overline{\sigma_\epsilon^4}}$ and $\widehat{\overline{\sigma_\epsilon^2}}$ are the consistent estimates of $\overline{\sigma_\epsilon^4}$ and $\overline{\sigma_\epsilon^2}$, respectively. The desired result then follows by Theorem 2.3.1 and the Slutsky lemma. By the law of large numbers, it suffices to show,

$$
\begin{aligned}
(a) \quad & \frac{2}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\hat{\epsilon}_{is}^2 = \frac{2}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\epsilon_{is}^2 + o_p(1), \\
(b) \quad & \frac{2}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\hat{\epsilon}_{is}^2\hat{\epsilon}_{it}^2 = \frac{2}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\epsilon_{is}^2\epsilon_{it}^2 + o_p(1).
\end{aligned}
$$

Recall that $\hat{\epsilon}_{it} = \epsilon_{it} + \boldsymbol{\lambda}_i^{0\prime}\boldsymbol{f}_t^0 - \widehat{\lambda}_i^{\prime}\widehat{\boldsymbol{B}}_t$. Then, we show (a),

$$
\begin{aligned}
\frac{2}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\hat{\epsilon}_{is}^2 \quad \leq \quad & \frac{2}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\epsilon_{is}^2 + \frac{4}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\epsilon_{is}\left\|\boldsymbol{\lambda}_i^{0\prime}\boldsymbol{f}_s^0 - \widehat{\lambda}_i^{\prime}\widehat{\boldsymbol{B}}_s\right\| \\
& + \frac{2}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\left\|\boldsymbol{\lambda}_i^{0\prime}\boldsymbol{f}_s^0 - \widehat{\lambda}_i^{\prime}\widehat{\boldsymbol{B}}_s\right\|^2 \\
\equiv \quad & \frac{2}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\epsilon_{is}^2 + 4G_{a,1} + 2G_{a,2}.
\end{aligned}
$$

We use the fact that $G_{a,2} = o_p(1)$ and verify this claim below, then, by Cauchy-Schwarz inequality, $G_{a,1} = o_p(1)$. Thus, proofs for (a) are completed. To show $G_{a,2} = o_p(1)$, note that

$$
\begin{aligned}
G_{a,2} \quad \leq \quad & \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=1}^{T}\left\|\boldsymbol{\lambda}_i^{0\prime}\boldsymbol{f}_s^0 - \widehat{\lambda}_i^{\prime}\widehat{\boldsymbol{B}}_s\right\|^2 \\
= \quad & \frac{1}{T}\sum_{t=2}^{T}\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}\left\|\boldsymbol{\lambda}_i^{0\prime}\boldsymbol{f}_s^0 - \widehat{\lambda}_i^{\prime}\widehat{\boldsymbol{B}}_s\right\|^2\right\} \\
= \quad & \frac{1}{T}\sum_{t=2}^{T}\left\|\boldsymbol{\Lambda}^0\boldsymbol{f}^{0\prime} - \widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{B}}^{\prime}\right\|^2 = O_p\left(N^{-1} + T^{-1}\right),
\end{aligned}
$$

where the final result follows from Lemma B.1.4 directly.

Now, we prove (b). Recall that $\hat{\epsilon}_{it} = \epsilon_{it} + \boldsymbol{\lambda}_i^{0\prime}\boldsymbol{f}_t^0 - \widehat{\lambda}_i^{\prime}\widehat{\boldsymbol{B}}_t = \epsilon_{it} + \boldsymbol{\lambda}_i^{0\prime}\boldsymbol{v}_t - \boldsymbol{d}_i^{\prime}\widehat{\boldsymbol{f}}_t \equiv \epsilon_{it} + a_{it}$ where

$\boldsymbol{v}_t = \boldsymbol{H}^{-1}\widehat{\boldsymbol{B}}_t - \boldsymbol{f}_t^0$, $\boldsymbol{d}_i = \widehat{\boldsymbol{\lambda}}_i - \boldsymbol{H}'^{-1}\boldsymbol{\lambda}_i^0$, and $a_{it} = \boldsymbol{\lambda}_i^{0\prime}\boldsymbol{v}_t - \boldsymbol{d}_i'\widehat{\boldsymbol{B}}_t$, or equivalently, $a_{it} = \boldsymbol{\lambda}_i^{0\prime}\boldsymbol{f}_t^0 - \widehat{\boldsymbol{\lambda}}_i'\widehat{\boldsymbol{B}}_t$. We have

$$\begin{aligned}
\frac{2}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\hat{\epsilon}_{is}^2\hat{\epsilon}_{it}^2 &\leq \frac{2}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\epsilon_{is}^2\epsilon_{it}^2 + \frac{4}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\epsilon_{is}^2 a_{it}^2 + \frac{4}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=1}^{t-1}a_{is}^2\epsilon_{it}^2 \\
&\quad + \frac{2}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=1}^{t-1}a_{is}^2 a_{it}^2 \\
&\equiv \frac{2}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\epsilon_{is}^2 + 4G_{b,1} + 4G_{b,2} + 2G_{b,3}.
\end{aligned}$$

For $G_{b,1}$, note that

$$\begin{aligned}
G_{b,1} &\leq \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=1}^{T}\epsilon_{is}^2 a_{it}^2 \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\frac{1}{T}\sum_{s=1}^{T}\epsilon_{is}^2\right)a_{it}^2 \\
&\leq (M + o_p(1))\frac{1}{NT}\left\|\boldsymbol{\Lambda}^0\boldsymbol{f}^{0\prime} - \widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{B}}'\right\|^2 = O_p\left(N^{-1} + T^{-1}\right),
\end{aligned}$$

where the second inequality follows by the Law of Large Numbers and Assumption A2.1, the final equality follows from Lemma B.1.4 directly. Similarly, we can readily show $G_{b,2} = o_p(1)$. To show $G_{b,3} = o_p(1)$, following arguments involving $T^{-1}\sum_{t=2}^{T}a_{it}^2$ in proofs of Lemma A.4 in Westerlund (2015), we have

$$\begin{aligned}
G_{b,3} &\leq \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=1}^{T}a_{is}^2 a_{it}^2 \\
&= \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{T}\sum_{t=1}^{T}a_{it}^2\right]^2 \\
&= \frac{1}{N}\sum_{i=1}^{N}\left\{\|\boldsymbol{\lambda}_i^{0\prime}\|^2\frac{1}{T}\sum_{t=2}^{T}\left\|\hat{\boldsymbol{B}}_t H^{-1} - \boldsymbol{f}_t^0\right\|^2 + \left\|\hat{\boldsymbol{\lambda}}_i - H^{-1}\boldsymbol{\lambda}_i\right\|^2\frac{1}{T}\sum_{t=2}^{T}\left\|\widehat{\boldsymbol{B}}_t\right\|^2\right\}^2 \\
&= \frac{1}{N}\sum_{i=1}^{N}\left\{O_p\left(N^{-p}\right)\left\|\boldsymbol{\lambda}_i^{0\prime}\right\|^2 + O_p\left(T^{-1}\right)\right\}^2 = O_p\left(N^{-2p} + N^{-p}T^{-1} + T^{-2}\right),
\end{aligned}$$

where the third equality follows from Lemma B.1.2 directly, and the normalization condition $\mathcal{J}_B^{-2}\widehat{\boldsymbol{B}}'\widehat{\boldsymbol{B}} = I_{R_0}$. Then, $G_{b,3} = o_p(1)$ holds for sure under Assumption A2.7(b) and user-specified $\mathcal{J}_B$ like $\sqrt{T}$, T and so on.

Case 2: Factors follow unit-root-type processes

For this case, we can combine above proofs for the mildly explosive factors with arguments in proofs of Lemma B.5 in Bai and Ng (2004) and Lemma A.4 in Westerlund (2015). We omit details here for brevity.

B.1.4 Proof of Theorem 3.3

Case 1: Factors follow mildly explosive processes

We first present some important immediate results under the local alternative:

(1) $\frac{1}{T}\sum_{t=2}^{T}\left\|\hat{\boldsymbol{B}}_t H^{-1} - \boldsymbol{f}_t\right\|^2 = O_p(N^{-p})$.

(2) $\left(\boldsymbol{H}^{-1}\widehat{\boldsymbol{B}}_t - \boldsymbol{f}_t^0\right) = O_p(N^{-p/2})$, each given t.

(3) $\left(\hat{\boldsymbol{\lambda}}_i - H^{-1}\boldsymbol{\lambda}_i\right) = O_p(\mathcal{J}_B^{-1})$, for each given i

(4) $\max\limits_{1\le t\le T}\frac{1}{\sqrt{T}}\left\|\sum_{s=2}^{t}\boldsymbol{H}^{-1}\hat{\boldsymbol{B}}_s - \boldsymbol{f}_s\right\| = O_p\left(N^{-\frac{p}{2}}\right)$

In order to prove above immediate results, one can follow the proofs for Lemmas 1–3 in Westerlund (2015) and modify the proofs of Lemmas B.1.2–B.1.3 by incorporating some addtional terms involving e_{it-1} under $H_{1NT}: \rho_i = 1 + \frac{c_i}{TN^{1/2}}\ \forall i$. It is easy to argue that such terms are dominated by the leading terms in the proofs of Lemma B.1.2 and B.1.3 under the null. This explains why the above immediate results remain the same as those under the null. We omit the details here for brevity.

Given (B.1.1), it suffices to prove the theorem by establishing the following results:

(i) $\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\hat{e}_{it}^2 = \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}e_{it}^2 + O_p(N^{1/2-p}) + O_p(\kappa_T N^{1/2}T^{-1}) = \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}e_{it}^2 + o_p(1)$;

(ii) $\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=2}^{T}\widehat{e}_{it-1}\Delta\widehat{e}_{it} = \frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=2}^{T}e_{it-1}\Delta e_{it} + O_p(N^{1/2}T^{-1} + N^{-1/2} + T^{-1/2}) + O_p(\kappa_T N^{1/2}T^{-1}) = \frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=2}^{T}e_{it-1}\Delta e_{it} + o_p(1)$;

(iii) $\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} e_{it-1} \Delta e_{it} \xrightarrow{d} \mathcal{N}(\mu, \frac{1}{2}\overline{\sigma_\epsilon^4})$ as $(N,T) \to \infty$ where $\mu = \lim_{N\to\infty} \frac{1}{2N} \sum_{i=1}^{N} c_i \sigma_{\epsilon,i}^2 >$ 0 and $\sigma_{\epsilon,i}^2 = E\left(\epsilon_{it}^2\right)$;

(iv) $\frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=2}^{T} e_{it}^2 = \frac{1}{2}\overline{\sigma_\epsilon^2} + o_p(1)$.

Given (i)–(iv), we have

$$
\begin{aligned}
\sqrt{N}T\left(\hat{\rho}-1\right) &= \left[\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T} e_{it}^2 + o_p\left(1\right)\right]^{-1} \left[\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=2}^{T} e_{it-1}\Delta e_{it} + o_p\left(1\right)\right] \\
&\xrightarrow{d} \mathcal{N}\left(\mu, 2\left(\overline{\sigma_\epsilon^2}\right)^{-2}\overline{\sigma_\epsilon^4}\right).
\end{aligned}
$$

Step 1: We show (i).

Recall that $r_{it} \equiv \boldsymbol{\lambda}_i^{0\prime}\boldsymbol{f}_t^0 - \widehat{\boldsymbol{\lambda}}_i'\widehat{\boldsymbol{B}}_t$. Note that $\Delta e_{it} = e_{it} - e_{it-1} = \left(\rho_i - 1\right)e_{it-1} + \epsilon_{it}$ and $\Delta\widehat{e}_{it} = Z_{it} - \widehat{\boldsymbol{\lambda}}_i'\widehat{\boldsymbol{B}}_t = r_{it} + \Delta e_{it} = r_{it} + \left(\rho_i - 1\right)e_{it-1} + \epsilon_{it}$. Without loss of generality, define $\widehat{e}_{i1} = 0$ to simplify proofs below as in Bai and Ng (2010). Then

$$
\widehat{e}_{it} = \sum_{s=2}^{t}\Delta\widehat{e}_{is} = \sum_{s=2}^{t}\Delta e_{is} + \sum_{s=2}^{t} r_{is} = e_{it} - e_{i1} + R_{it}
$$

where $R_{it} = \boldsymbol{\lambda}_i^{0\prime}\boldsymbol{V}_t - \boldsymbol{d}_i'\widehat{\boldsymbol{F}}_t$, $\boldsymbol{V}_t = \sum_{s=2}^{t}\boldsymbol{v}_s$, $\boldsymbol{v}_t = \boldsymbol{H}^{-1}\widehat{\boldsymbol{B}}_t - \boldsymbol{f}_t^0$, and $\boldsymbol{d}_i = \widehat{\boldsymbol{\lambda}}_i - \boldsymbol{H}'^{-1}\boldsymbol{\lambda}_i^0$. Consequently, we have

$$
\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\widehat{e}_{it}^2 - e_{it}^2\right) = \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(e_{i1}^2 + R_{it}^2 - 2e_{it}e_{i1} - 2e_{i1}R_{it} + 2e_{it}R_{it}\right).
$$

Under H_{1NT}, we can use the above stated immediate results and follow the proof of Lemma A.4 in

Westerlund (2015) to obtain the following results:

$$
\begin{aligned}
&(1a) \quad \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}e_{i1}^2 = O_p\left(T^{-1}\right),\\
&(1b) \quad \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}e_{i1}e_{it} = \frac{1}{\sqrt{NT}}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}e_{i1}\frac{e_{it}}{\sqrt{T}} = O_p\left(N^{-1/2}T^{-1/2}\right),\\
&(1c) \quad \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}e_{i1}R_{it} = \frac{1}{NT}\sum_{i=1}^{N}e_{i1}\frac{1}{T}\sum_{t=2}^{T}R_{it} = O_p\left(T^{-1/2}N^{-p/2} + \kappa_T^{1/2}T^{-1}\right),\\
&(1d) \quad \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}R_{it}^2 = O_p\left(N^{-1} + \kappa_T T^{-1}\right),\\
&(1e) \quad \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}e_{it}R_{it} = O_p\left(N^{-1/2} + \kappa_T^{1/2}T^{-1/2}\right),
\end{aligned} \tag{B.1.7}
$$

where the first line follows from Markov inequality and the fact that $E\left(e_{i1}^2\right) \leq M$ under Assumption A2.3(c), the second line follows moment calculations under the m.d.s property of $\{\epsilon_{it}\}$, the third and fourth lines implicitly use the fact that $\|R_{it}\| \leq \|\boldsymbol{\lambda}_i^0\| \|\boldsymbol{V}_t\| + \|\boldsymbol{d}_i\| \left\|\widehat{\boldsymbol{F}}_t\right\|$ and the above stated immediate results, and the last line holds by using Cauchy–Schwarz inequality based on the result of (1d) above and the fact that $N^{-1}T^{-2}\sum_{i=1}^{N}\sum_{t=2}^{T}e_{it}^2 = O_p(1)$ under Assumption A2.8. Consequently, we have $\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\widehat{e}_{it}^2 - e_{it}^2\right) = o_p(1)$.

Step 2: We show (ii)

$$
\widehat{e}_{it} = \sum_{s=2}^{t}\Delta\widehat{e}_{is} = \sum_{s=2}^{t}\Delta e_{is} + \sum_{s=2}^{t}r_{is} \equiv (\rho_i - 1)\sum_{s=2}^{t}e_{is-1} + \sum_{s=2}^{t}\epsilon_{is} + R_{it}
$$

Noting that $\widehat{e}_{it}^2 = \left(\widehat{e}_{it-1} + \Delta\widehat{e}_{it}\right)^2 = \widehat{e}_{it-1}^2 + 2\widehat{e}_{it-1}\Delta\widehat{e}_{it} + \left(\Delta\widehat{e}_{it}\right)^2$, we have

$$
\begin{aligned}
\sum_{t=2}^{T}\widehat{e}_{it-1}\Delta\widehat{e}_{it} &= \frac{1}{2}\sum_{t=2}^{T}\widehat{e}_{it}^2 - \frac{1}{2}\sum_{t=2}^{T}\widehat{e}_{it-1}^2 - \frac{1}{2}\sum_{t=2}^{T}\left(\Delta\widehat{e}_{it}\right)^2\\
&= \frac{1}{2}\widehat{e}_{iT}^2 - \frac{1}{2}\widehat{e}_{i1}^2 - \frac{1}{2}\sum_{t=2}^{T}\left(\Delta\widehat{e}_{it}\right)^2.
\end{aligned}
$$

By the same token, $\sum_{t=2}^{T}e_{it-1}\Delta e_{it} = \frac{1}{2}\sum_{t=2}^{T}e_{it}^2 - \frac{1}{2}\sum_{t=2}^{T}e_{it-1}^2 - \frac{1}{2}\sum_{t=2}^{T}\left(\Delta e_{it}\right)^2 = \frac{1}{2}e_{iT}^2 - \frac{1}{2}e_{i1}^2 -$

$\frac{1}{2}\sum_{t=2}^{T}\left(\Delta e_{it}\right)^{2}$. It follows that

$$\begin{aligned}
& \frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\widehat{e}_{it-1}\Delta\widehat{e}_{it}-e_{it-1}\Delta e_{it}\right) \\
= & \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left(\widehat{e}_{iT}^{2}-e_{iT}^{2}\right)-\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left(\widehat{e}_{i1}^{2}-e_{i1}^{2}\right)-\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}\left[\left(\Delta\widehat{e}_{it}\right)^{2}-\left(\Delta e_{it}\right)^{2}\right] \\
\equiv & A_{1}-A_{2}-A_{3}.
\end{aligned}$$

Since $\widehat{e}_{i1}=0$ by construction, it is easy to see $A_{2}=\frac{-1}{\sqrt{NT}}\sum_{i=1}^{N}e_{i1}^{2}=O_{p}\left(N^{1/2}T^{-1}\right)$ by Markov inequality and Assumption A2.1. It remains to bound A_1 and A_3.

We first bound A_1. Noting that $\widehat{e}_{it}=\sum_{s=2}^{t}\Delta e_{is}+R_{it}=e_{it}-e_{i1}+R_{it}\equiv e_{it}+A_{it}$ with $A_{it}=-e_{i1}+R_{it}$, we have

$$A_{1}=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left[\left(e_{iT}+A_{iT}\right)^{2}-e_{iT}^{2}\right]=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}A_{iT}^{2}-\frac{2}{\sqrt{NT}}\sum_{i=1}^{N}e_{iT}A_{iT}\equiv A_{1,1}-2A_{1,2}.$$

To bound $A_{1,1}$ and $A_{1,2}$, most arguments are the same as those used in Step 1 of the proof of Theorem 2.3.1 by using the above stated immediate results for the local alternative case. Below we focus on the tho terms that involve e_{it-1} and arise only under the local alternative. For $A_{1,2}$, we write $A_{1,2}=-\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}e_{iT}e_{i1}+\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}e_{iT}R_{iT}\equiv -A_{1,2a}+A_{1,2b}$. Note that time-series observations for each individual i for available from period 1 to period T. Therefore, we use $e_{iT}=\sum_{s=1}^{T}\epsilon_{is}$ and $\epsilon_{i1}=e_{i1}$ again as Bai and Ng (2010) and Westerlund (2015). Then, for $A_{1,2a}$, we have

$$\begin{aligned}
A_{1,2a} & = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left(1+\frac{c_{i}}{TN^{1/2}}\right)^{T}\epsilon_{i1}^{2}+\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=2}^{T}\epsilon_{is}\epsilon_{i1} \\
& \equiv A_{1,2a1}+A_{1,2a2}.
\end{aligned}$$

Note that the second term $A_{1,2a2}$ is $O_p\left(T^{-1/2}\right)$ by the m.d.s property of $\{\epsilon_{is}\}$ under Assumption A2.1(c). The first term $A_{1,2a1}$ is $O_p\left(N^{1/2}T^{-1}\right)$ by Markov equality and the fact that it is nonnegative and has expectation

$$\begin{aligned}\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\left(1+\frac{c_i}{TN^{1/2}}\right)^{T}E\left(\epsilon_{i1}^{2}\right) &\asymp \frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\exp\left(c_i/N^{1/2}\right)E\left(\epsilon_{i1}^{2}\right)\\ &\asymp \frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\left(1+c_i/N^{1/2}\right)E\left(\epsilon_{i1}^{2}\right)=O\left(N^{1/2}T^{-1}\right).\end{aligned}$$

Then $A_{1,2a}=O_p\left(T^{-1/2}\right)+O_p\left(N^{1/2}T^{-1}\right)$. For $A_{1,2b}$, we use the fact that $\Delta e_{it}=\left(\rho_i-1\right)e_{i,t-1}+\epsilon_{it}=\frac{c_i}{TN^{1/2}}e_{i,t-1}+\epsilon_{it}$ to obtain

$$A_{1,2b}=\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta e_{it}R_{iT}=\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}c_i e_{i,t-1}R_{iT}+\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=2}^{T}\epsilon_{it}R_{iT}.$$

The second term in the last displayed equation was bounded by $o(1)$ in Step 2 of the proof for Theorem 2.3.1 and the result continues to hold by using the above stated immediate results for the case of local alternatives. That is,

$$\begin{aligned}\left\|\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=2}^{T}\epsilon_{it}R_{iT}\right\| &= O_p\left(N^{-1/2}+T^{-1/2}+\kappa_T N^{1/2}T^{-1}\right)\\ &= o_p\left(1\right),\end{aligned}$$

where the final equality holds under Assumption A2.7(a).

For the first term, we use $R_{iT}=\boldsymbol{\lambda}_i^{0\prime}\boldsymbol{V}_T-\boldsymbol{d}_i'\widehat{\boldsymbol{F}}_T$ to obtain

$$\left\|\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}c_i e_{it-1}R_{iT}\right\|\leq\left\|\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}c_i e_{it-1}\boldsymbol{\lambda}_i^{0\prime}\boldsymbol{V}_T\right\|+\left\|\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}c_i e_{it-1}\boldsymbol{d}_i'\widehat{\boldsymbol{F}}_T\right\|. \tag{B.1.8}$$

By similar arguments as used in the proof of Lemma A.4 in Westerlund (2015) and the immediate results derived in the proof of Theorem 2.3.1, we have

$$\begin{aligned}\left\|\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}c_i e_{it-1}\boldsymbol{\lambda}_i^{0\prime}\boldsymbol{V}_T\right\| &\leq \left\|\frac{1}{NT^{3/2}}\sum_{i=1}^{N}\sum_{t=2}^{T}c_i e_{i,t-1}\boldsymbol{\lambda}_i^{0\prime}\right\|\frac{1}{T^{1/2}}\left\|\boldsymbol{V}_T\right\|\\ &= O_p\left(N^{-1+p/2}\right)O_p\left(N^{-p/2}\right)=O_p\left(N^{-1}\right),\end{aligned}$$

where we use the fact that $\left\|\sum_{i=1}^{N}\sum_{t=2}^{T} c_i e_{it-1}\boldsymbol{\lambda}_i^{0\prime}\right\| = O_p\left(N^{p/2}T^{3/2}\right)$ under Assumptions A2.1(c), A2.2, A2.4, and A2.8 by writing that $e_{it} = \sum_{s=2}^{t}\epsilon_{is} + \left(1+\frac{c_i}{TN^{1/2}}\right)^t \epsilon_{i1} = O_p\left(T^{1/2}\right)$, and that $\left\|\boldsymbol{V}_T/T^{1/2}\right\| = O_p\left(N^{-p/2}\right)$ by the above immediate result (4) given at the beginning of this proof. Similarly,

$$\begin{aligned}
\left\|\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T} c_i e_{i,t-1}\boldsymbol{d}_i'\widehat{\boldsymbol{F}}_T\right\| &\leq \left\|\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T} c_i e_{i,t-1}\boldsymbol{d}_i'\right\|\left\|\widehat{\boldsymbol{F}}_T\right\| \\
&\leq \left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T^{3/2}}\sum_{t=2}^{T} c_i e_{i,t-1}\right\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{i=1}^{N}\|\boldsymbol{d}_i\|^2\right)^{1/2}\frac{1}{\sqrt{T}}\left\|\widehat{\boldsymbol{F}}_T\right\| \\
&= O_p(1)O_p\left(\mathcal{J}_B^{-1}\right)O_p\left(\mathcal{J}_B T^{-1/2}\kappa_T^{1/2}\right) = O_p\left(T^{-1/2}\kappa_T^{1/2}\right),
\end{aligned}$$

where the second inequality follows by Cauchy-Schwarz inequality, and the first equality holds by the facts that $\left\|\widehat{\boldsymbol{F}}_T\right\| = O_p(\mathcal{J}_B\kappa_T^{1/2})$, $\|\boldsymbol{d}_i\| = O_p\left(\mathcal{J}^{-1}\right)$ as obtained in Step 2 of the proof of Theorem 2.3.1, and $T^{-1/2}e_{it} = O_p(1)$ under Assumption A2.8 by similar arguments in Westerlund (2015). Then, we have

$$\left\|\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T} c_i e_{i,t-1}R_{iT}\right\| = O_p\left(N^{-1}+T^{-1/2}\kappa_T^{1/2}\right).$$

In sum, we have $A_{1,2} = O_p(\kappa_T^{1/2}T^{-1/2}+N^{1/2}T^{-1}+N^{-1}) = o_p(1)$.

For $A_{1,1}$, we have $A_{1,1} = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}A_{iT}^2 = O_p(N^{1/2}T^{-1}+N^{-1/2}+\kappa_T N^{1/2}T^{-1})$ as obtained in Step 2 of the proof for Theorem 2.3.1. Then, we obtain that

$$A_1 = A_{1,1}-2A_{1,2} = O_p(N^{1/2}T^{-1}+N^{-1/2}+\kappa_T N^{1/2}T^{-1}) = o_p(1),$$

where the final equality holds under Assumption A2.7(a).

Now, we study A_3. Noting that $\Delta\widehat{e}_{it} = \Delta e_{it}+r_{it}$, we have

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}\left[(\Delta\widehat{e}_{it})^2-(\Delta e_{it})^2\right] = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T} r_{it}^2 + \frac{2}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta e_{it}r_{it} \equiv A_{3,1}+2A_{3,2}.$$

Following the proof of Lemma B.1.4 but with the above immediate results (2)-(3) in place, we can show that $A_{3,1} = O_p(N^{1/2}T^{-1} + N^{-1/2})$. For $A_{3,2}$, we make the decomposition

$$A_{3,2} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \Delta e_{it} \boldsymbol{\lambda}_i^{0\prime} \boldsymbol{v}_t + \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \Delta e_{it} \boldsymbol{d}_i' \widehat{\boldsymbol{f}}_t \equiv A_{3,21} + A_{3,22}.$$

For $A_{3,21}$, we have

$$\begin{aligned} \|A_{3,21}\| &\leq \frac{1}{\sqrt{N}} \left[\frac{1}{T} \sum_{t=2}^{T} \left\| \sum_{i=1}^{N} \Delta e_{it} \boldsymbol{\lambda}_i^{0\prime} \right\|^2 \right]^{1/2} \left[\frac{1}{T} \sum_{t=2}^{T} \|\boldsymbol{v}_t\|^2 \right]^{1/2} \\ &= N^{1/2} O_p\left(N^{p/2}\right) O_p\left(N^{-p/2}\right) = O_p\left(N^{-1/2}\right) \end{aligned}$$

where we use the fact that $\frac{1}{T}\sum_{t=2}^{T} \|\boldsymbol{v}_t\|^2 = O_p\left(N^{-p}\right)$ by the above immediate result (1) and that

$$\begin{aligned} \frac{1}{T} \sum_{t=2}^{T} \left\| \sum_{i=1}^{N} \Delta e_{it} \boldsymbol{\lambda}_i^{0\prime} \right\|^2 &\leq \frac{2}{T} \sum_{t=2}^{T} \left\| \sum_{i=1}^{N} \frac{c_i}{TN^{1/2}} \Delta e_{i,t-1} \boldsymbol{\lambda}_i^{0\prime} \right\|^2 + \frac{2}{T} \sum_{t=2}^{T} \left\| \sum_{i=1}^{N} \epsilon_{it} \boldsymbol{\lambda}_i^{0\prime} \right\|^2 \\ &= O_p\left(N^{p-1}T^{-1}\right) + O_p\left(N^p\right) = O_p\left(N^p\right) \end{aligned}$$

under Assumptions A2.8, A2.2 and the m.d.s property of $\{\epsilon_{it}\}$ under Assumption A2.1(c). For $A_{3,22}$, we have

$$A_{3,22} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \Delta e_{it} \boldsymbol{d}_i' \boldsymbol{H} \boldsymbol{f}_t^0 + \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \Delta e_{it} \boldsymbol{d}_i' \left(\widehat{\boldsymbol{f}}_t - \boldsymbol{H} \boldsymbol{f}_t^0 \right).$$

It is easy to argue that the leading term of $A_{3,22}$ is given by $\bar{A}_{3,22} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \Delta e_{it} \boldsymbol{d}_i' \boldsymbol{H} \boldsymbol{f}_t^0$. In addition,

$$\begin{aligned} \left\|\bar{A}_{3,22}\right\| &\leq \frac{\sqrt{N}}{T} \left[\frac{1}{N} \sum_{i=1}^{N} \left\| \sum_{t=2}^{T} \Delta e_{it} \boldsymbol{f}_t^{0\prime} \boldsymbol{H}' \right\|^2 \right]^{1/2} \left[\frac{1}{N} \sum_{i=1}^{N} \|\boldsymbol{d}_i\|^2 \right]^{1/2} \\ &= N^{1/2}T^{-1} O_p\left(\mathcal{J}_B\right) O_p\left(\mathcal{J}_B^{-1}\right) = O_p\left(N^{1/2}T^{-1}\right), \end{aligned}$$

where we use the fact that $\frac{1}{N}\sum_{i=1}^{N} \|\boldsymbol{d}_i\|^2 = O_p\left(\mathcal{J}_B^{-1}\right)$ by following the proof of Lemma 3 in Bai and

Ng (2010), and by (B.2.10) and similar arguments as used to obtain Lemma B.3.2(c), $\|\sum_t \Delta e_{it} \boldsymbol{f}_t^{0\prime} \boldsymbol{H}'\| \leq \left\|\sum_t \Delta e_{it} \boldsymbol{f}_t^{0\prime} (\boldsymbol{\rho}_0^0)^{-T}\right\| \left\|(\boldsymbol{\rho}_0^0)^T \boldsymbol{H}'\right\| = \left\|\sum_t \epsilon_{it} \boldsymbol{f}_t^{0\prime} (\boldsymbol{\rho}_0^0)^{-T}\right\| \left\|(\boldsymbol{\rho}_0^0)^T \boldsymbol{H}'\right\| \{1 + o_p(1)\} = O_p(1) O_p(\mathcal{J}_B)$. In sum, $A_{3,2} = O_p\left(N^{-1/2} + N^{1/2}T^{-1}\right)$ and $A_3 = O_p(N^{1/2}T^{-1} + N^{-1/2})$

Then, $\frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=2}^T (\widehat{e}_{i,t-1}\Delta\widehat{e}_{it} - e_{i,t-1}\Delta e_{it}) = O_p(N^{1/2}T^{-1} + N^{-1/2} + \kappa_T N^{1/2}T^{-1}) = o_p(1)$ under Assumption A2.7(a).

Step 3: We show (iii).

Noting that $\Delta e_{it} = e_{it} - e_{i,t-1} = (\rho_i - 1) e_{i,t-1} + \epsilon_{it}$, by direct and simple calculations, we can show that

$$
\begin{aligned}
\frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=2}^T e_{i,t-1}\Delta e_{it} &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=2}^T \sum_{s=2}^{t-1} \epsilon_{it}\epsilon_{is} + \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=2}^T \sum_{s=2}^{t-1} (\rho_i - 1)(e_{i,s-1}\epsilon_{it} + e_{i,t-1}\epsilon_{is}) \\
&\quad + \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=2}^T \sum_{s=2}^{t-1} (\rho_i - 1)^2 e_{i,t-1}e_{i,s-1} \\
&\equiv \mathbb{A}_1 + \mathbb{A}_2 + \mathbb{A}_3.
\end{aligned}
$$

By Theorem 2.3.1, $\mathbb{A}_1 \xrightarrow{d} \mathcal{N}(0, \frac{1}{2}\overline{\sigma_\epsilon^4})$. It remains to study $\mathbb{A}_2$ and $\mathbb{A}_3$.

Noting that $\rho_i - 1 = \frac{c_i}{TN^{1/2}}$ under H_{1NT},

$$
\begin{aligned}
\mathbb{A}_2 &= \frac{1}{N} \sum_{i=1}^N c_i \left(\frac{1}{T^2} \sum_{t=2}^T \sum_{s=2}^{t-1} e_{i,t-1}\epsilon_{is} \right) + \frac{1}{N} \sum_{i=1}^N c_i \left(\frac{1}{T^2} \sum_{t=2}^T \sum_{s=2}^{t-1} e_{i,s-1}\epsilon_{it} \right) \\
&\equiv \mathbb{A}_{2,1} + \mathbb{A}_{2,2}.
\end{aligned}
$$

Noting that

$$
e_{is} = \rho_i e_{i,s-1} + \epsilon_{is} = \sum_{j=0}^{s-1} \rho_i^j \epsilon_{i,s-j} = \sum_{r=1}^{s} \rho_i^{s-r} \epsilon_{ir}, \qquad \text{(B.1.9)}
$$

we can make the following decomposition for $\mathbb{A}_{2,1}$:

$$
\begin{aligned}
\mathbb{A}_{2,1} &= \frac{1}{N} \sum_{i=1}^N c_i \frac{1}{T^2} \sum_{t=2}^T \sum_{s=2}^{t-1} \rho_i^{t-1-s} \epsilon_{is}^2 + \frac{1}{N} \sum_{i=1}^N c_i \frac{1}{T^2} \sum_{t=2}^T \sum_{s=2}^{t-1} \sum_{r\neq s, r=1}^{t-1} \rho_i^{t-1-r} \epsilon_{ir}\epsilon_{is} \\
&\equiv \mathbb{A}_{2,11} + \mathbb{A}_{2,12}
\end{aligned}
$$

It is easy to show that $\mathbb{A}_{2,12} = o_p(1)$ by Chebyshev inequality with the repeated uses of the m.d.s property of the sequence $\{\epsilon_{it}\}$ under Assumption A2.1(c) and the mixing equality (see, e.g., Shao et al. (1996)) under Assumption A2.1(a). To study $\mathbb{A}_{2,11}$, without loss of generality, we assume that $c_i > 0$ for all i. Then

$$\begin{aligned}
\mathbb{A}_{2,11} &= \frac{1}{N}\sum_{i=1}^{N} c_i \frac{1}{T^2}\sum_{t=2}^{T}\sum_{s=2}^{t-1}\rho_i^{t-1-s}E\left(\epsilon_{is}^2\right) + o_p(1) = \frac{1}{N}\sum_{i=1}^{N} c_i\sigma_{\epsilon,i}^2 \frac{1}{T^2}\sum_{t=2}^{T}\sum_{s=2}^{t-1}\rho_i^{t-1-s} + o_p(1) \\
&= \frac{1}{N}\sum_{i=1}^{N}\frac{c_i\sigma_{\epsilon,i}^2}{\rho_i - 1}\frac{1}{T^2}\sum_{t=2}^{T}\left(\rho_i^{t-2} - 1\right) + o_p(1) \\
&= \frac{1}{N}\sum_{i=1}^{N}\frac{c_i\sigma_{\epsilon,i}^2}{(\rho_i - 1)^2}\frac{1}{T^2}\left[\rho_i^{T-1} - 1 - (T-1)(\rho_i - 1)\right] + o_p(1) \\
&= \frac{1}{N}\sum_{i=1}^{N}\frac{c_i\sigma_{\epsilon,i}^2}{\left(\frac{c_i}{TN^{1/2}}\right)^2}\frac{1}{T^2}\left\{\left[\left(1 + \frac{c_i}{TN^{1/2}}\right)^{\frac{TN^{1/2}}{c_i}}\right]^{\frac{(T-1)c_i}{TN^{1/2}}} - 1 - (T-1)\frac{c_i}{TN^{1/2}}\right\} + o_p(1) \\
&= \frac{1}{N}\sum_{i=1}^{N}\frac{N\sigma_{\epsilon,i}^2}{c_i}\left[\exp\left(\frac{(T-1)c_i}{TN^{1/2}}\right) - 1 - \frac{c_i}{N^{1/2}}\right] + o_p(1) \\
&= \frac{1}{N}\sum_{i=1}^{N}\frac{N\sigma_{\epsilon,i}^2}{c_i}\left[\frac{1}{2}\left(\frac{(T-1)\,c_i}{TN^{1/2}}\right)^2\right] + o_p(1) \\
&= \frac{1}{2N}\sum_{i=1}^{N} c_i\sigma_{\epsilon,i}^2 + o_p(1) \xrightarrow{p} \mu.
\end{aligned}$$

Consequently, we have $\mathbb{A}_{2,1} \xrightarrow{p} \mu$. By (B.1.9),

$$\mathbb{A}_{2,2} = \frac{1}{N}\sum_{i=1}^{N} c_i\left(\frac{1}{T^2}\sum_{t=2}^{T}\sum_{s=2}^{t-1}\sum_{r=1}^{s-1}\rho_i^{s-1-r}\epsilon_{ir}\epsilon_{it}\right)$$

By straightforward moment calculations, we can show that $E(\mathbb{A}_{2,2}) = 0$ and Var$(\mathbb{A}_{2,2}) = o(1)$ with the repeated uses of the m.d.s property of the sequence $\{\epsilon_{it}\}$ under Assumption A2.1(c) and the mixing equality (see, e.g., Shao et al. (1996)) under Assumption A1(a). Then $\mathbb{A}_{2,2} = o_p(1)$ and $\mathbb{A}_2 = \mathbb{A}_{2,1} + \mathbb{A}_{2,2} = \mu + o_p(1)$.

For $\mathbb{A}_3$, it is easy to verify that $E(\mathbb{A}_3) = o(1)$ and Var$(\mathbb{A}_3) = o(1)$ as above for $\mathbb{A}_{2,11}$ and $\mathbb{A}_{2,2}$ by using the m.d.s property of the sequence $\{\epsilon_{it}\}$ under Assumption A2.1(c) and the mixing equality (see, e.g., Shao et al. (1996)) under Assumption A2.1(a). Then $\mathbb{A}_3 = o_p(1)$.

In sum, we have shown that $\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} e_{i,t-1} \Delta e_{it} \xrightarrow{d} \mathcal{N}(\mu, \frac{1}{2}\overline{\sigma_\epsilon^4})$ as $(N,T) \to \infty$.

Step 4: We show (iv).

Noting that $\Delta e_{is} = (\rho_i - 1) e_{i,s-1} + \epsilon_{is}$, we have

$$e_{i,t-1} - e_{i1} = \sum_{s=2}^{t-1} \Delta e_{is} = \sum_{s=2}^{t-1} (\rho_i - 1) e_{i,s-1} + \sum_{s=2}^{t-1} \epsilon_{is}$$

or equivalently when $t \geq 3$, $e_{i,t-1} = \sum_{s=2}^{t-1} \epsilon_{is} + \sum_{s=2}^{t-1} (\rho_i - 1) e_{i,s-1} + e_{i1}.$, and $e_{it-1} = e_{i1}$ when $t = 2$. Then, we have

$$\begin{aligned}
\frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=3}^{T} e_{i,t-1}^2 &= \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=3}^{T} \sum_{s=2}^{t-1} \sum_{k=2}^{t-1} \epsilon_{is}\epsilon_{ik} + \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=3}^{T} \sum_{s=2}^{t-1} \sum_{k=2}^{t-1} (\rho_i - 1)^2 e_{i,s-1} e_{i,k-1} \\
&+ \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=2}^{T} e_{i1}^2 + \frac{2}{NT^2} \sum_{i=1}^{N} \sum_{t=3}^{T} \sum_{s=2}^{t-1} \sum_{k=2}^{t-1} \epsilon_{is} (\rho_i - 1) e_{i,k-1} \\
&+ \frac{2}{NT^2} \sum_{i=1}^{N} \sum_{t=3}^{T} \sum_{s=2}^{t-1} \epsilon_{is} e_{i1} + \frac{2}{NT^2} \sum_{i=1}^{N} \sum_{t=3}^{T} \sum_{s=2}^{t-1} (\rho_i - 1) e_{i,s-1} e_{i1} \\
&\equiv \mathbb{A}_4 + \mathbb{A}_5 + \mathbb{A}_6 + 2\mathbb{A}_7 + 2\mathbb{A}_8 + 2\mathbb{A}_9.
\end{aligned}$$

By Step 4 in the proof of Theorem 2.3.1, $\mathbb{A}_4 \xrightarrow{p} \frac{1}{2}\overline{\sigma_\epsilon^2}$. As in Step 3, we can easily show that $\mathbb{A}_l = o_p(1)$ for $l = 5, 6$ by straightforward expectation calculations and Markov inequality and by noting that they are nonnegative. Then by the Cauchy-Schwarz inequality, we have $\mathbb{A}_7 \leq \{\mathbb{A}_4\mathbb{A}_5\}^{1/2} = o_p(1)$, $\mathbb{A}_8 \leq \{\mathbb{A}_4\mathbb{A}_6\}^{1/2} = o_p(1)$, and $\mathbb{A}_9 \leq \{\mathbb{A}_5\mathbb{A}_6\}^{1/2} = o_p(1)$. It follows that $\frac{1}{NT^2} \sum_i \sum_t e_{it-1}^2 \xrightarrow{p} \frac{1}{2}\overline{\sigma_\epsilon^2}$ as $(N,T) \to \infty$.

Case 2: Factors follow a unit root process

When the factors follow a unit root process, Westerlund (2015) consider the local power property in a similar framework. The proof in this case follows essentially that in Westerlund (2015). The detailed arguments are the combination of proofs for Case 1 and corresponding arguments in Westerlund (2015). We omitted them for the brevity.

B.1.5 Proofs of Theorem 3.4

Let P^* denote the probability measure induced by the wild bootstrap conditional on $\boldsymbol{Z}$. Let E^* and Var^* denote the expectation and variance under P^* and O_{p^*} and o_{p^*} the probability order under P^*. In view of the fact that (1) the null hypothesis, that is $\rho_i = 1$ for all i, is satisfied in the bootstrap world, (2) ϵ^*_{it} are independent over i and t conditional on $\boldsymbol{Z}$ and (3) both $\hat{\boldsymbol{\lambda}}_i$ and $\hat{\boldsymbol{B}}_t$ are fixed given $\boldsymbol{Z}$. Henceforth, the proofs can be adopted straightforwardly but much plain than that of Theorem 2.3.1.

The outline of proofs are presented here for brevity. Recall $Z^*_{it} = \hat{\boldsymbol{\lambda}}_i' \hat{\boldsymbol{B}}_t + \epsilon^*_{it}$ with $\epsilon^*_{it} = \hat{\epsilon}_{it}\varsigma_{it}$, and denote $\hat{\boldsymbol{\lambda}}^*_i$ and $\hat{\boldsymbol{B}}^*_t$ as the principal component estimates of this equation in the bootstrap world under the same normalization condition used for obtain $\hat{\boldsymbol{\lambda}}_i$ and $\hat{\boldsymbol{B}}_t$.

(1) Those immediate results stated in in Lemma B.1.2 to Lemma B.1.6 still hold for $\hat{\boldsymbol{\lambda}}^*_i$ and $\hat{\boldsymbol{B}}^*_t$. In short, $\hat{\boldsymbol{\lambda}}^*_i$ and $\hat{\boldsymbol{B}}^*_t$ can estimate $\hat{\boldsymbol{\lambda}}_i$ and $\hat{\boldsymbol{B}}_t$ consistently upon to a well-defined rotation matrix $\boldsymbol{H}^*$ in bootstrap world, detailed proofs are very similar to proofs of Lemma B.1.2 to Lemma B.1.6.

(2) Let $\hat{\epsilon}^*_{it} = Z^*_{it} - \hat{\boldsymbol{\lambda}}^{*\prime}_i \hat{\boldsymbol{B}}^*_t = \epsilon^*_{it} + \left(\hat{\boldsymbol{\lambda}}_i' \hat{\boldsymbol{B}}_t - \hat{\boldsymbol{\lambda}}^{*\prime}_i \hat{\boldsymbol{B}}^*_t\right) \equiv \epsilon^*_{it} + r^*_{it}$, where r^*_{it} is defined implicitly. Because $\hat{e}_{i1} = 0$ implied by (2.2 .12), it follows $\hat{e}^*_{it} = \sum_{s=2}^t \hat{\epsilon}^*_{is} = \sum_{s=2}^t \epsilon^*_{is} + \sum_{s=2}^t r^*_{is} \equiv \sum_{s=2}^t \epsilon^*_{is} + R^*_{it}$, and thus $\Delta\widehat{e}^*_{it} = \hat{e}^*_{it} - \hat{e}^*_{it-1}$.

Let $e^*_{it} = \sum_{s=2}^t \epsilon^*_{is}$, and $\Delta e^*_{it} = e^*_{it} - e^*_{it-1}$ with $\epsilon^*_{it} = \hat{\epsilon}_{it}\varsigma_{it}$, ς_{it} are i.i.d random variables such that ς_{it} has a 0.5 chance of being 1 and a 0.5 chance of being -1, therefore, ϵ^*_{it} has the same mean and variance as those of $\hat{\epsilon}_{it}$.

(3) Then, base on above facts in (1) and (2), following arguments in Step 1 and Step 2 of proofs for Theorem 2.3.1 in above section, it can be shown

$$\frac{1}{NT^2} \sum_{i=1}^N \sum_{t=2}^T \left((\widehat{e}^*_{it})^2 - (e^*_{it})^2 \right) = o_p(1),$$
$$\frac{1}{\sqrt{N}T} \sum_{i=1}^N \sum_{t=2}^T \left(\widehat{e}^*_{it-1} \Delta\widehat{e}^*_{it} - e^*_{it-1} \Delta e^*_{it} \right). = o_p(1)$$

(4) We have proved that under the null hypothesis with Assumption A2.1–A2.7, $\widetilde{PPT}$ will convergence to the standard normal Distribution as $N, T \to \infty$ in Corollary 2.3.2, by Polya-Cantelli lemma, it follows that $\sup_{x\in\mathbb{R}} \left| P\left(\widetilde{PPT} \leq x\right) - \Psi(x) \right| \xrightarrow{P} 0$, where $\Psi(x)$ denotes the standard normal distribution with mean 0 and variance 1. Then the results follows if following statement holds

$$\sup_{x\in\mathbb{R}} \left| P^*\left(PPT^* \leq x\right) - \Psi(x) \right| \xrightarrow{P} 0.$$

Then, based on above arguments, the validity of bootstrap version of proposed test statistic can be concluded because

$$\sup_{\boldsymbol{x}\in\mathbb{R}^p} \left| P\left\{\widetilde{PPT} \leq \boldsymbol{x}\right\} - P^*\left\{PPT^* \leq \boldsymbol{x}\right\} \right| \to 0,$$

holds. To this end, we show the asymptotic normality of PPT^* by imitating the Step 3 and 4 of the proofs for Theorem 2.3.1 in the previous section. Now, based on results in (3) above, again, we have

$$\begin{aligned}
\sqrt{N}T\left(\hat{\rho}^* - 1\right) &= \sqrt{N}T \frac{\text{tr}\left(\hat{e}^{*\prime}_{-1}\widehat{\Delta e}^*\right)}{\text{tr}\left(\hat{e}^{*\prime}_{-1}\hat{e}^*_{-1}\right)} \\
&= \frac{\sqrt{N}\,\text{tr}\left(\frac{1}{NT}e^{*\prime}_{-1}\Delta e^*\right)}{\frac{1}{NT^2}\,\text{tr}\left(e^{*\prime}_{-1}e^*_{-1}\right)} + o_p(1) \\
&= \left[\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(e^*_{it}\right)^2\right]^{-1} \frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=2}^{T} e^*_{it-1}\Delta e^*_{it} + o_p(1) \\
&= K_1^{*-1}K_2^* + o_p(1),
\end{aligned}$$

where $e^*_{it} = \sum_{s=2}^{t} \epsilon^*_{is} = \sum_{s=s}^{t} \hat{\epsilon}_{is}\varsigma_{is}$ by construction in (2) above.

Based on above facts, we claim that

$$\frac{\sqrt{N}\,\text{tr}\left(\frac{1}{NT}e^{*\prime}_{-1}\Delta e^*\right)}{\frac{1}{NT^2}\,\text{tr}\left(e^{*\prime}_{-1}e^*_{-1}\right)} \xrightarrow{d} \mathcal{N}\left(0, 2\left(\overline{\sigma^2_\epsilon}\right)^{-2}\overline{\sigma^4_\epsilon}\right) \quad \text{as} \quad (N,T)\to\infty.$$

Similar to Step 4 of proofs for Theorem 2.3.1 in previous section, by straightforward calculations,

$\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(e_{it}^{*}\right)^2 = \frac{1}{2}\widehat{\sigma_\epsilon^2}$ because $e_{it}^{*} = \sum_{s=2}^{t}\epsilon_{is}^{*}$ and ϵ_{is}^{*}'s are $i.i.d$ sequences with the same mean and variance as the original sequence $\{\hat{\epsilon}_{is}\}$.

As for the asymptotic normality of K_2^* above, it can be proved by employing central limit theorem for m.d.s as done in proofs of Theorem 2.3.1 in the previous section. To see this,

$$
\begin{aligned}
K_2^* &= \frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=2}^{T} e_{it-1}^{*}\Delta e_{it}^{*} \\
&= \frac{1}{\sqrt{N}T}\sum_{i=1}^{N} e_{i,1}^{*}\Delta e_{i,2}^{*} + \frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=3}^{T}\sum_{s=2}^{t-1}\epsilon_{is}^{*}\epsilon_{it}^{*} \\
&= \sum_{t=3}^{T}\left[\sum_{s=2}^{t-1}\left(\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\epsilon_{is}^{*}\epsilon_{it}^{*}\right)\right] \equiv \sum_{t=3}^{T}\eta_{Nt}^{*},
\end{aligned}
$$

where the third equality holds because of facts that$e_{i1}^{*} = \hat{e}_{i1}\varsigma_{i1}$ and $\hat{e}_{i1} = 0$. Then, for $\{\epsilon_{it}^{*}\}$ in bootstrap world and Assumption A2.1(c), we can also construct σ-algebra $\mathcal{F}_{NT,t-1}^{\epsilon,*}$ as the analogy of $\mathcal{F}_{NT,t-1}^{\epsilon}$, it can be seen that $E\left(\eta_{Nt}^{*} \mid \mathcal{F}_{NT,t-1}^{\epsilon,*}\right) = 0$ by the usage of law of iterated expectations for the sequence $\{\epsilon_{it}^{*}\}$ and the properties of i.i.d sequence ς_{it}. Therefore, $\{\eta_{Nt}^{*}\}$ are also martingale difference sequences with respect to $\mathcal{F}_{NT,t-1}^{\epsilon,*}$.

By imitating arguments in Step 3 of proofs for Theorem 2.3.1 in the previous section, and the usage of properties of i.i.d sequence ς_{it}, those conditions required by CLT for m.d.s also can be verified. Then, $K_2^* \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{2}\overline{\sigma_\epsilon^4}\right)$ follows immediately.

To this end, we are going to verify following conditions to apply the central limit theorem for m.d.s, namely,

(1) $s_T^2 \to \frac{1}{2}\widehat{\sigma_\epsilon^4}$ with $s_T^2 = E^*\left[\left(\sum_{t=3}^{T}\eta_{Nt}^{*}\right)^2\right]$, and $\frac{1}{2}\widehat{\sigma_\epsilon^4} = \lim_{N,T\to\infty}\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=3}^{T}\sum_{s=2}^{t-1}\hat{\epsilon}_{is}^2\hat{\epsilon}_{it}^2$,

(2) $\left(s_T^2\right)^{-1}\sum_{t=3}^{T}E^*\left[\left(\eta_{Nt}^{*}\right)^2\mathbf{1}\left(|\eta_{Nt}^{*}| > \delta s_T\right) \mid \mathcal{F}_{NT,t-1}^{\epsilon,*}\right] \to 0$ for all $\delta > 0$,

(3) $\left(s_T^2\right)^{-1}V_T \to 1$ where $V_T = \sum_{t=3}^{T}E^*\left[\left(\eta_{Nt}^{*}\right)^2 \mid \mathcal{F}_{NT,t-1}^{\epsilon,*}\right]$.

Then, we can conclude that

$$K_2^* = \sum_{t=3}^{T} \eta_{Nt}^* \xrightarrow{d} \mathcal{N}(0, \frac{1}{2}\overline{\sigma_\epsilon^4}) \quad \text{as} \quad (N,T) \to \infty.$$

For condition (1) above, we have

$$\begin{aligned}
s_T^2 &= E^*\left[\sum_{t=3}^{T}\sum_{r=1}^{T}\eta_{Nt}^*\eta_{Nr}^*\right] \\
&= \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=3}^{T}\sum_{r=1}^{T}\sum_{s=2}^{t-1}\sum_{q=1}^{r-1} E^*\left(\epsilon_{is}^*\epsilon_{it}^*\epsilon_{jr}^*\epsilon_{jq}^*\right) \\
&= \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=3}^{T}\sum_{s=2}^{t-1}\sum_{q=1}^{t-1} E^*\left(\epsilon_{is}^*\epsilon_{iq}^*\left(\epsilon_{it}^*\right)^2\right) \\
&= \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=3}^{T}\sum_{s=2}^{t-1}\sum_{q=1}^{t-1} E^*\left[E^*\left(\left(\epsilon_{it}^*\right)^2 \mid \mathcal{F}_{NT,t-1}^{\epsilon,*}\right)\epsilon_{is}^*\epsilon_{iq}^*\right] \\
&= \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=3}^{T}\epsilon_{it}^2\sum_{s=2}^{t-1}\sum_{q=1}^{t-1} E^*\left[\hat{\epsilon}_{is}^*\epsilon_{iq}^*\right] \\
&= \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=3}^{T}\hat{\epsilon}_{it}^2\sum_{s=2}^{t-1} E^*\left[\left(\epsilon_{is}^*\right)^2\right] \\
&= \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=3}^{T}\sum_{s=2}^{t-1}\hat{\epsilon}_{is}^2\hat{\epsilon}_{it}^2 \\
&\to \lim_{N,T\to\infty}\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=3}^{T}\sum_{s=2}^{t-1}\hat{\epsilon}_{is}^2\hat{\epsilon}_{it}^2 = \frac{1}{2}\widehat{\overline{\sigma_\epsilon^4}},
\end{aligned}$$

where the third to seventh equality hold by the independence among $\{\epsilon_{it}^*\}$ over $\forall(i,t)$ with $\epsilon_{it}^* = \hat{\epsilon}_{it}\varsigma_{it}$, where $\{\varsigma_{it}\}$ are i.i.d sequences with mean 0 and variance 1 by construction.

To verify condition (2) above, it suffices to show

$$\left(s_T^2\right)^{-1-\beta}\sum_{t=3}^{T} E^*\left|\eta_{Nt}^*\right|^{2+\beta} \to 0 \quad \text{as} \quad (N,T)\to\infty \text{ for some } \beta > 0.$$

Since $s_T^2 = O_p^*(1)$ shown above for condition (1), it is sufficient to show that $\sum_{t=2}^{T} E\left|\eta_{Nt}^*\right|^4 \to$

0 as $(N,T)\to\infty$. Note that

$$\begin{aligned}\sum_{t=3}^{T} E^{*}\left|\eta_{Nt}^{*}\right|^{4} &= \sum_{t=3}^{T} E^{*}\left[\sum_{s=2}^{t-1}\left(\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\epsilon_{is}^{*}\epsilon_{it}^{*}\right)\right]^{4} \\ &= \frac{1}{N^{2}T^{4}}\sum_{t=3}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N}\sum_{2\leq r,s,q,v\leq t-1} E\left(\epsilon_{is}^{*}\epsilon_{it}^{*}\epsilon_{jr}^{*}\epsilon_{jt}^{*}\epsilon_{kq}^{*}\epsilon_{kt}^{*}\epsilon_{lv}^{*}\epsilon_{lt}^{*}\right) \\ &= 3\mathcal{R}_{1}+\mathcal{R}_{2},\end{aligned}$$

where

$$\begin{aligned}\mathcal{R}_{1} &= \frac{1}{N^{2}T^{4}}\sum_{t=3}^{T}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\sum_{2\leq r,s,q,v\leq t-1} E^{*}\left[\epsilon_{is}^{*}\epsilon_{ir}^{*}\left(\epsilon_{it}\right)^{2}\epsilon_{jq}^{*}\epsilon_{jv}^{*}\left(\epsilon_{jt}^{*}\right)^{2}\right] \\ \mathcal{R}_{2} &= \frac{1}{N^{2}T^{4}}\sum_{t=3}^{T}\sum_{i=1}^{N}\sum_{2\leq r,s,q,v\leq t-1} E^{*}\left[\epsilon_{is}^{*}\epsilon_{ir}^{*}\epsilon_{iq}^{*}\epsilon_{iv}^{*}\left(\epsilon_{it}^{*}\right)^{4}\right].\end{aligned}$$

For $\mathcal{R}_2$, we have the following decomposition by direct calculations,

$$\begin{aligned}\mathcal{R}_{2} &= \frac{6}{N^{2}T^{4}}\sum_{t=3}^{T}\sum_{i=1}^{N}\sum_{2\leq r\neq s\leq t-1} E^{*}\left[\left(\epsilon_{is}^{*}\right)^{2}\left(\epsilon_{ir}^{*}\right)^{2}\left(\epsilon_{it}^{*}\right)^{4}\right]+\frac{1}{N^{2}T^{4}}\sum_{t=3}^{T}\sum_{i=1}^{N}\sum_{2\leq r\leq t-1} E^{*}\left[\left(\epsilon_{ir}^{*}\right)^{4}\left(\epsilon_{it}^{*}\right)^{4}\right] \\ &\leq \frac{6}{N^{2}T^{4}}\sum_{t=3}^{T}\sum_{i=1}^{N}\sum_{2\leq r\neq s\leq t-1}\hat{\epsilon}_{is}^{2}\hat{\epsilon}_{ir}^{2}\hat{\epsilon}_{it}^{4}+\frac{1}{N^{2}T^{4}}\sum_{t=3}^{T}\sum_{i=1}^{N}\sum_{2\leq r\leq t-1}\hat{\epsilon}_{ir}^{4}\hat{\epsilon}_{it}^{4} \\ &= O_{p}\left(N^{-1}T^{-1}\right)+O_{p}\left(N^{-1}T^{-2}\right)\end{aligned}$$

where the final inequality holds by the independence among $\{\epsilon_{it}^{*}\}$ over $\forall(i,t)$ with $\epsilon_{it}^{*}=\hat{\epsilon}_{it}\varsigma_{it}$, where $\{\varsigma_{it}\}$ are i.i.d sequences with mean 0 and variance 1 by construction. For $\mathcal{R}_1$,

$$\begin{aligned}
\mathcal{R}_1 &= \frac{1}{N^2T^4}\sum_{t=3}^{T}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\sum_{2\leq r,s,q,v\leq t-1} E^*\left[\epsilon_{is}^*\epsilon_{ir}^*\left(\epsilon_{it}\right)^2\epsilon_{jq}^*\epsilon_{jv}^*\left(\epsilon_{jt}^*\right)^2\right] \\
&= \frac{1}{N^2T^4}\sum_{t=3}^{T}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\sum_{2\leq s,q\leq t-1} E^*\left[\left(\epsilon_{is}^*\right)^2\left(\epsilon_{it}\right)^2\left(\epsilon_{jq}^*\right)^2\left(\epsilon_{jt}^*\right)^2\right] \\
&= \frac{1}{N^2T^4}\sum_{t=3}^{T}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\sum_{2\leq s,q\leq t-1} E^*\left(\epsilon_{is}^*\right)^2 E^*\left(\epsilon_{it}\right)^2 E^*\left(\epsilon_{jq}^*\right)^2 E^*\left(\epsilon_{jt}^*\right)^2 \\
&= \frac{1}{N^2T^4}\sum_{t=3}^{T}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\sum_{2\leq s,q\leq t-1} \hat{\epsilon}_{is}^2\hat{\epsilon}_{it}^2\hat{\epsilon}_{jq}^2\hat{\epsilon}_{jt}^2 \\
&\leq \frac{1}{T}\left\{\frac{1}{T^3}\sum_{t=3}^{T}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\sum_{2\leq s\leq t}\hat{\epsilon}_{is}^2\right)\hat{\epsilon}_{it}^2\right]^2\right\} = O_p\left(T^{-1}\right),
\end{aligned}$$

where the second to fourth equality hold by the independence among $\{\epsilon_{it}^*\}$ over $\forall(i,t)$ with $\epsilon_{it}^* = \hat{\epsilon}_{it}\varsigma_{it}$, where $\{\varsigma_{it}\}$ are i.i.d sequences with mean 0 and variance 1 by construction, the first inequality holds by Cauchy-Schwarz inequality.

Collecting the bounds of $\mathcal{R}_1$ and $\mathcal{R}_2$ above, it follows $\sum_{t=3}^{T} E^*\left|\eta_{Nt}^*\right|^4 = o_p(1)$ directly, which completes the verification of condition (2) stated above.

To verify the condition (3) given above, note that

$$\begin{aligned}
V_T &= \sum_{t=3}^{T} E^*\left[\left(\eta_{Nt}^*\right)^2 \mid \mathcal{F}_{NT,t-1}^{\epsilon,*}\right] \\
&= \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=3}^{T} E^*\left[\sum_{s=2}^{t-1}\sum_{r=2}^{t-1}\epsilon_{is}^*\epsilon_{ir}^*\left(\epsilon_{it}^*\right)^2 \mid \mathcal{F}_{NT,t-1}^{\epsilon,*}\right] \\
&= \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=3}^{T}\sum_{s=2}^{t-1}\hat{\epsilon}_{is}^2 E^*\left[\left(\epsilon_{it}^*\right)^2 \mid \mathcal{F}_{NT,t-1}^{\epsilon,*}\right],
\end{aligned}$$

where the second and third equality hold by the independence among $\{\epsilon_{it}^*\}$ over $\forall(i,t)$ with $\epsilon_{it}^* = \hat{\epsilon}_{it}\varsigma_{it}$, where $\{\varsigma_{it}\}$ are i.i.d sequences with mean 0 and variance 1 by construction, and the law of iterated expectations. Obviously, $E^*\left(V_T\right) = s_T^2$ by the law of iterated expectations.

By direct calculations of moments, it is straightforward to show $E^*(V_T^2) = s_T^2 + o_p^*(1)$. Thus, condition (3) is satisfied.

Recall that, $e_{it}^* = \sum_{s=2}^{t} \epsilon_{is}^*$, then $\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=3}^{T}\left(e_{it}^*\right)^2 = \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=3}^{T}\left(\sum_{s=2}^{t-1}\epsilon_{is}^*\right)^2$, by direct calculations, we have

$$\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=3}^{T}\left(e_{it}^*\right)^2 = \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=3}^{T}\sum_{s=2}^{t-1}\sum_{r=2}^{t-1}\epsilon_{is}^*\epsilon_{ir}^*,$$

then, by using the law of large numbers for the sequence $\{\epsilon_{it}\}$, as $(N,T)\to\infty$, it follows that

$$\begin{aligned}
\lim_{N,T\to\infty}\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=3}^{T}\left(e_{it}^*\right)^2 &= \lim_{N,T\to\infty}\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=3}^{T}\sum_{s=2}^{t-1}\sum_{r=2}^{t-1}E^*\left(\epsilon_{is}^*\epsilon_{ir}^*\right)\\
&= \lim_{N,T\to\infty}\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=3}^{T}\sum_{s=2}^{t-1}E^*\left(\epsilon_{is}^*\right)^2\\
&= \lim_{N,T\to\infty}\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=3}^{T}\sum_{s=2}^{t-1}\hat{\epsilon}_{is}^2 = \frac{1}{2}\widehat{\sigma_\epsilon^2}
\end{aligned}$$

Note that $\frac{1}{2}\widehat{\sigma_\epsilon^2}$ and $\frac{1}{2}\widehat{\sigma_\epsilon^4}$ are used in Corollary 2.3.2. By the central limit theorem for m.d.s, we then conclude that

$$\left(2\left(\widehat{\sigma_\epsilon^2}\right)^{-2}\widehat{\sigma_\epsilon^4}\right)^{-1/2}\frac{\sqrt{N}\,\mathrm{tr}\left(\frac{1}{NT}e_{-1}^{*\prime}\Delta e^*\right)}{\frac{1}{NT^2}\,\mathrm{tr}\left(e_{-1}^{*\prime}e_{-1}^*\right)} \xrightarrow{d} \mathcal{N}(0,1) \quad \text{as} \quad (N,T)\to\infty$$

Based on the construction of PPT^*, the desired result follows directly.

Collecting all above arguments, the validity of the bootstrap version of test statistic is justified.

B.2 Proofs for Technical Lemmas

This section is composed of 3 parts. Section B.2.1 contains the proofs of Lemmas B.1-B.5 in the above paper. Section B.2.2 contains some technical lemmas that are used in the proofs of the main results and Lemmas B.1-B.5. Section S3 provides some additional simulation results.

B.2.1 Proofs of Lemma B.1.1

The proof is similar to that of Proposition 1 in Bai (2003). The major difference is that one needs to replace the scaling constant T^{-1} in Bai (2003) by $\mathcal{J}_T^{-1}$ in suitable places.

B.2.2 Proofs of Lemma B.1.2

Proofs of Lemma B.1.2(a)

Note that (2.2 .10) can be written in matrix form as follows

$$\boldsymbol{Z} = \boldsymbol{\Lambda}^0 \boldsymbol{B}^{0\prime} + \boldsymbol{\Lambda}^0 \boldsymbol{u}' + \boldsymbol{\epsilon} \tag{B.2.1}$$

Recall that $\Sigma_{Z,N} = \frac{1}{N}\boldsymbol{Z}'\boldsymbol{Z}$ and $\mathcal{J}_T = (\boldsymbol{\rho}_0^0)^T \mathcal{J}_B$, where $(\boldsymbol{\rho}_0^0)^T \equiv diag((\rho_{01}^0)^T, \ldots, (\rho_{0R_0}^0)^T)$, and $\mathcal{J}_B$ is user-specified scalar in the normalization conditions for the PC estimation. By the eigenvalue problem, we have the identity

$$\Sigma_{Z,N}\widehat{\boldsymbol{B}} = \widehat{\boldsymbol{B}}\boldsymbol{V_{Z,N}}, \tag{B.2.2}$$

where $\boldsymbol{V_{Z,N}}$ is a diagonal matrix that consists of the first R_0 eigenvalues of $\Sigma_{Z,N}$ arranged in descending order along its main diagonal line. Postmultiplying both sides of (B.2.2) by the recalling factor $\mathcal{J}_T^{-1}$ and substituting (B.2.1) into (B.2.2), we have

$$\begin{aligned}
& \widehat{\boldsymbol{B}}\boldsymbol{V}_{Z,N}\mathcal{J}_T^{-1} - N^{-1+p}\boldsymbol{B}^0\frac{1}{N^p}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0(\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}) \\
= \; & N^{-1}\left(\boldsymbol{\Lambda}^0\boldsymbol{u}' + \boldsymbol{\epsilon}\right)'\left(\boldsymbol{\Lambda}^0\boldsymbol{u}' + \boldsymbol{\epsilon}\right)\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1} \\
& + N^{-1}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\left(\boldsymbol{\Lambda}^0\boldsymbol{u}' + \boldsymbol{\epsilon}\right)\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1} + N^{-1}\left(\boldsymbol{\Lambda}^0\boldsymbol{u}' + \boldsymbol{\epsilon}\right)'\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1} \\
\equiv \; & A_1 + A_2 + A_3.
\end{aligned} \tag{B.2.3}$$

Below, we bound each term on the last line. First, for A_1 we have

$$\begin{aligned} A_1 &= N^{-1}\boldsymbol{u}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\boldsymbol{u}'\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1} + N^{-1}\boldsymbol{u}\left(\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}\right)\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1} + N^{-1}\left(\boldsymbol{\epsilon}'\boldsymbol{\Lambda}^0\right)\boldsymbol{u}'\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1} \\ &\quad + N^{-1}\boldsymbol{\epsilon}'\boldsymbol{\epsilon}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1} \\ &\equiv A_{11} + A_{12} + A_{13} + A_{14}. \end{aligned}$$

For A_{11}, we have by the sub-multiplicity of Frobenius norm,

$$\begin{aligned} \|A_{11}\| &\leq N^{-1+p}\left\|\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0}{N^p}\right\|\|\boldsymbol{u}\|^2\left\|\mathcal{J}_B^{-1}\widehat{\boldsymbol{B}}\right\|\left\|\mathcal{J}_B\mathcal{J}_T^{-1}\right\| \\ &\leq N^{-1+p}O_p(T)\left\|\mathcal{J}_B\mathcal{J}_T^{-1}\right\| = O_p\left(N^{-1+p}T\left[\left(\rho_{01}^0\right)^{-T} + \ldots + \left(\rho_{0R_0}^0\right)^{-T}\right]\right), \end{aligned}$$

where the second inequality by Assumptions A2.1(b) and A2.2 and the normalization condition, and the final equality holds by the fact that $\mathcal{J}_T \equiv \mathcal{J}_B^{-1}\left(\boldsymbol{\rho}_0^0\right)^T$. Similarly,

$$\begin{aligned} \|A_{12}\| &\leq N^{-1+p/2}T\left\|\frac{\boldsymbol{u}}{T^{1/2}}\right\|\left\|\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}}{N^{p/2}T^{1/2}}\right\|\left\|\mathcal{J}_B^{-1}\widehat{\boldsymbol{B}}\right\|\left\|\mathcal{J}_B\mathcal{J}_T^{-1}\right\| \\ &\leq O_p\left(N^{-1+p/2}T\right)\|\left\|\mathcal{J}_B\mathcal{J}_T^{-1}\right\| = O_p\left(N^{-1+\frac{p}{2}}T\left[\left(\rho_{01}^0\right)^{-T} + \ldots + \left(\rho_{0R_0}^0\right)^{-T}\right]\right), \end{aligned}$$

where the second inequality holds by Assumptions A2.1(b)-(c) and A2.2, and the normalization condition. In particular,

$$E\left\|\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}\right\|^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=1}^{T}E\left(\epsilon_{is}\boldsymbol{\lambda}_i^{0\prime}\boldsymbol{\lambda}_j\epsilon_{js}\right) = \sum_{i=1}^{N}\sum_{s=1}^{T}E\left(\epsilon_{is}^2\boldsymbol{\lambda}_i^{0\prime}\boldsymbol{\lambda}_i^0\right) = O_p\left(N^pT\right) \qquad \text{(B.2.4)}$$

where the second equality holds because $E\left(\epsilon_{is} \mid \{\epsilon_{js}\}_{j=1,j\neq i}^N, \{\boldsymbol{\lambda}_j\}_{j=1}^N\right) = 0$ by Assumption A2.1(c), and the last holds by Assumptions A2.2, A2.3 and A2.4. By the same token,

$$\begin{aligned} \|A_{13}\| &\leq N^{-1+p/2}T\left\|\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}}{N^{p/2}T^{1/2}}\right\|\left\|\frac{\boldsymbol{u}}{T^{1/2}}\right\|\left\|\mathcal{J}_B^{-1}\widehat{\boldsymbol{B}}\right\|\left\|\mathcal{J}_B\mathcal{J}_T^{-1}\right\| \\ &= N^{-1+p/2}TO_p\left(\left(\rho_{01}^0\right)^{-T} + \ldots + \left(\rho_{0R_0}^0\right)^{-T}\right), \end{aligned}$$

and

$$\|A_{14}\| \leq N^{-1}\|\boldsymbol{\epsilon}\|_{sp}^2 \left\|\mathcal{J}_B^{-1}\widehat{\boldsymbol{B}}\right\| \left\|\mathcal{J}_B\mathcal{J}_T^{-1}\right\| = N^{-1}(N+T)\,O_p\left(\left[\left(\rho_{01}^0\right)^{-T}+\ldots+\left(\rho_{0R_0}^0\right)^{-T}\right]\right)$$

where we also use the that that $\|\boldsymbol{\epsilon}\|_{sp}^2 = O_p(N+T)$ under Assumption A2.3(e). In sum, we have shown that $\|A_1\| = O_p\left(\left(N^{-1+p}T+1\right)\left[\left(\rho_{01}^0\right)^{-T}+\ldots+\left(\rho_{0R_0}^0\right)^{-T}\right]\right)$.

Next, we study A_2. Note that $A_2 = N^{-1}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\boldsymbol{u}'\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1} + N^{-1}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1} \equiv A_{21}+A_{22}$. For A_{21}, we have

$$\begin{aligned}
A_{21} &\leq N^{-1+p}T^{1/2}\left\|\boldsymbol{B}^0\left(\boldsymbol{\rho}_0^0\right)^{-T}\left(\boldsymbol{\rho}_0^0\right)^{T}\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0}{N^p}\frac{\boldsymbol{u}'}{T^{1/2}}\mathcal{J}_B^{-1}\widehat{\boldsymbol{B}}\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\| \\
&\leq N^{-1+p}T^{1/2}\left\|\boldsymbol{B}^0\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|\left\|\left(\boldsymbol{\rho}_0^0\right)^{T}\right\|_{sp}\left\|\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0}{N^p}\right\|\left\|\frac{\boldsymbol{u}'}{T^{1/2}}\right\|\left\|\mathcal{J}_B^{-1}\widehat{\boldsymbol{B}}\right\|\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|_{sp} \\
&\leq N^{-1+p}T^{1/2}O_p(1)\left\|\left(\boldsymbol{\rho}_0^0\right)^{T}\right\|_{sp}\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|_{sp} = O_p\left(N^{-1+p}T^{1/2}\right),
\end{aligned}$$

where the third inequality above follows by Assumptions A2.1(b) and A2.2, Lemma B.3.2(e), and the normalization condition, and the equality follows because $\left\|\left(\boldsymbol{\rho}_0^0\right)^{T}\right\|_{sp}\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|_{sp} = O_p(1)$ under Assumption A2.5(a). To see this, note that as $T\to\infty$, $\left\|\left(\boldsymbol{\rho}_0^0\right)^{T}\right\|_{sp} = \left(1+\frac{c_{r,max}}{\kappa_T}\right)^T \asymp \exp\left(c_{r,max}T\kappa_T^{-1}\right)$ and $\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|_{sp} = \left(1+\frac{c_{r,min}}{\kappa_T}\right)^{-T} \asymp \exp\left(-c_{r,min}T\kappa_T^{-1}\right)$, which in conjunction with the requirement that $|c_{r,max}-c_{r,min}| = O_p\left(\frac{\kappa_T}{T}\right)$, implies that $\left\|\left(\boldsymbol{\rho}_0^0\right)^{T}\right\|_{sp}\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|_{sp} = O_p(\exp\left(c_{r,max}-c_{r,min})T\kappa_T^{-1}\right)) = O_p(1)$. Similarly,

$$\begin{aligned}
\|A_{22}\| &\leq N^{-1+p/2}T^{1/2}\left\|\boldsymbol{B}^0\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|\left\|\left(\boldsymbol{\rho}_0^0\right)^{T}\right\|_{sp}\left\|\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}}{N^{p/2}T^{1/2}}\right\|\left\|\left(\mathcal{J}_B^{-1}\widehat{\boldsymbol{B}}\right)\right\|\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|_{sp} \\
&= O_p(N^{-1+\frac{p}{2}}T^{1/2}),
\end{aligned}$$

where we also use (B.2.4). In sum, we have $\|A_2\| = O_p\left(N^{-1+p}T^{1/2}\right)$.

Next, note that $A_3 = N^{-1}\boldsymbol{u}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1} + N^{-1}\boldsymbol{\epsilon}'\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1} \equiv A_{31}+A_{32}$. Under

Assumptions A2.1(b) and A2.2, and by Lemma B.1.1 and (B.2.4), we have

$$\|A_{31}\| \leq N^{-1+p}T^{1/2}\left\|\frac{\boldsymbol{u}}{T^{1/2}}\right\|\left\|\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^{0}}{N^{p}}\right\|\left\|\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_{T}^{-1}\right\| = O_{p}\left(N^{-1+p}T^{1/2}\right)$$

and

$$\|A_{32}\| \leq N^{-1+p/2}T^{1/2}\left\|\frac{\boldsymbol{\epsilon}'\boldsymbol{\Lambda}^{0}}{N^{p/2}T^{1/2}}\right\|\left\|\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_{T}^{-1}\right\| = O_{p}\left(N^{-1+p/2}T^{1/2}\right).$$

So $\|A_3\| = O_p\left(N^{-1+p}T^{1/2}\right)$ and

$$\|A_1 + A_2 + A_3\| = O_p\left(N^{-1+p}T^{1/2}\right) + O_p\left(\left(N^{-1+p}T + 1\right)\left[\left(\rho_{01}^{0}\right)^{-T} + \ldots + \left(\rho_{0R_0}^{0}\right)^{-T}\right]\right).$$

That is,

$$\begin{aligned} & \left\|\widehat{\boldsymbol{B}}\left(N^{1-p}\mathbf{V}_{Z,N}\mathcal{J}_{T}^{-1}\right) - \boldsymbol{B}^{0}\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^{0}}{N^{p}}\left(\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_{T}^{-1}\right)\right\| \\ = \;& O_p\left(T^{1/2}\right) + O_p\left(\left(T + N^{-p+1}\right)\left[\left(\rho_{01}^{0}\right)^{-T} + \ldots + \left(\rho_{0R_0}^{0}\right)^{-T}\right]\right) = O_p\left(T^{1/2}\right), \end{aligned}$$

where the last equality follows because we can readily see that

$$\left(T + N^{-p+1}\right)\max\{\left(\rho_{01}^{0}\right)^{-T}, \ldots, \left(\rho_{0R_0}^{0}\right)^{-T}\} \asymp \left(T + N^{-p+1}\right)\exp(-c_{r,min}T\kappa_T^{-1}) = o\left(1\right)$$

under Assumption A2.5. Here, $c_{r,min} = \min\{c_1, \ldots, c_{R_0}\}$. Then

$$\begin{aligned} & \frac{1}{T}\left\|\widehat{\boldsymbol{B}}\left(N^{1-p}\mathbf{V}_{Z,N}\mathcal{J}_{T}^{-1}\right) - \boldsymbol{B}^{0}\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^{0}}{N^{p}}\left(\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_{T}^{-1}\right)\right\|^{2} \\ = \;& \frac{1}{T}\sum_{t=2}^{T}\left\|\left(N^{1-p}\mathcal{J}_{T}^{-1}\mathbf{V}_{Z,N}\right)\widehat{\boldsymbol{B}}_{t} - \left(\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_{T}^{-1}\right)'\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^{0}}{N^{p}}\boldsymbol{B}_{t}^{0}\right\|^{2} = O_p\left(1\right). \end{aligned}$$

In addition, from above analyses of A_1 to A_3, we can see that A_{21} and A_{31} are the leading term in the above MSE bound, where $A_{21} = N^{-1+p}\boldsymbol{B}^{0}\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^{0}}{N^{p}}\boldsymbol{u}'\widehat{\boldsymbol{B}}\mathcal{J}_{T}^{-1}$ and $A_{31} = N^{-1+p}\boldsymbol{u}\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^{0}}{N^{p}}\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_{T}^{-1}$. Moving A_{21} and A_{31} from the right group of terms to the left group of terms with regard to

(B.2.3), squaring both sides and then dividing by T, we can readily show that

$$\frac{1}{T}\left\|\widehat{\boldsymbol{B}}\left(N^{1-p}\mathcal{J}_T^{-1}\boldsymbol{V}_{Z,N}\right)-\boldsymbol{B}^0\left[\tfrac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0}{N^p}(\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1})+\tfrac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0}{N^p}\boldsymbol{u}'\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}\right]-\boldsymbol{u}\left[\tfrac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0}{N^p}(\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1})\right]\right\|^2 = O_p\left(N^{-p}\right) \tag{B.2.5}$$

Now we are in a position to define the rotation matrix $\boldsymbol{H}$. Let

$$\boldsymbol{H}_u=\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0}{N^p}(\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}) \text{ and } \boldsymbol{H}_B=\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0}{N^p}\left(\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}\right)+\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0}{N^p}\boldsymbol{u}'\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}.$$

Note that $\boldsymbol{H}_u$ is nonsingular, $\|\boldsymbol{H}_u\|=O_p(1)$ and $\|\boldsymbol{H}_u^{-1}\|=O_p(1)$ by Assumption A2.2 and B.1.1. In addition,

$$\begin{aligned}\|\boldsymbol{H}_B-\boldsymbol{H}_u\| &= \left\|\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0}{N^p}\frac{\boldsymbol{u}'\widehat{\boldsymbol{B}}}{\mathcal{J}_B\sqrt{T}}\mathcal{J}_T^{-1}\mathcal{J}_B T^{1/2}\right\| \\ &\le \left\|\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0}{N^p}\right\|\left\|\frac{\boldsymbol{u}'}{\sqrt{T}}\right\|\left\|\mathcal{J}_B^{-1}\widehat{\boldsymbol{B}}\right\|\left\|\mathcal{J}_T^{-1}\mathcal{J}_B T^{1/2}\right\| \\ &= O_p\left(T^{1/2}\left[\left(\rho_{01}^0\right)^{-T}+\ldots+\left(\rho_{0R_0}^0\right)^{-T}\right]\right),\end{aligned}$$

where the second equality holds by Assumption A2.1(b), the normalization condition, and the fact that $\mathcal{J}_T\equiv\mathcal{J}_B^{-1}\left(\boldsymbol{\rho}_0^0\right)^T$. Note that $\|\boldsymbol{H}_B-\boldsymbol{H}_u\|=o_p(1)$ when all factors are mildly explosive. It follows that $\|\boldsymbol{H}_B\|\le\|\boldsymbol{H}_u\|+\|\boldsymbol{H}_B-\boldsymbol{H}_u\|=O_p(1)$ and

$$\begin{aligned}\left\|\boldsymbol{H}_B^{-1}\right\| &\le \left\|\boldsymbol{H}_u^{-1}\right\|+\|\boldsymbol{H}_B-\boldsymbol{H}_u\|\left\|\boldsymbol{H}_u^{-1}\right\|^2+O_p\left(\|\boldsymbol{H}_B-\boldsymbol{H}_u\|^2\right) \\ &= O_p(1)+O_p\left(T^{1/2}\left[\left(\rho_{01}^0\right)^{-T}+\ldots+\left(\rho_{0R_0}^0\right)^{-T}\right]\right)=O_p(1),\end{aligned} \tag{B.2.6}$$

where we use the fact that $\boldsymbol{H}_B^{-1}=\boldsymbol{H}_u^{-1}-\boldsymbol{H}_u^{-1}\left(\boldsymbol{H}_B-\boldsymbol{H}_u\right)\boldsymbol{H}_u^{-1}+O_p(\|\boldsymbol{H}_B-\boldsymbol{H}_u\|^2)$ by

Fact 9.9.60 in Bernstein (2009). Then, by (B.2.5) ,

$$\begin{aligned}
& \left\| \widehat{\boldsymbol{B}} N^{1-p} \boldsymbol{V}_{Z,N} \mathcal{J}_T^{-1} \boldsymbol{H}_B^{-1} - \boldsymbol{B}^0 - \boldsymbol{u} \boldsymbol{H}_u \boldsymbol{H}_B^{-1} \right\| \\
\leq & \left\| \widehat{\boldsymbol{B}} N^{1-p} \boldsymbol{V}_{Z,N} \mathcal{J}_T^{-1} - \boldsymbol{B}^0 \boldsymbol{H}_B - \boldsymbol{u} \boldsymbol{H}_u \right\| \left\| \boldsymbol{H}_B^{-1} \right\| \\
= & O_p \left(N^{-p/2} T^{1/2} \right) O_p \left(\left\| \boldsymbol{H}_B^{-1} \right\| \right) = O_p \left(N^{-p/2} T^{1/2} \right). \qquad \text{(B.2.7)}
\end{aligned}$$

In addition, noting that $\boldsymbol{H}_u \boldsymbol{H}_B^{-1} = \boldsymbol{I} - (\boldsymbol{H}_B - \boldsymbol{H}_u) \boldsymbol{H}_u^{-1} + O_p \left(\|\boldsymbol{H}_B - \boldsymbol{H}_u\|^2 \|\boldsymbol{H}_u\| \right)$,

$$\begin{aligned}
\left\| \boldsymbol{H}_u \boldsymbol{H}_B^{-1} - \boldsymbol{I} \right\| &= O_p \left(\|\boldsymbol{H}_B - \boldsymbol{H}_u\| \left\| \boldsymbol{H}_u^{-1} \right\| \right) + O_p \left(\|\boldsymbol{H}_B - \boldsymbol{H}_u\|^2 \|\boldsymbol{H}_u\| \right) \\
&= O_p \left(T^{1/2} \left[\left(\rho_{01}^0 \right)^{-T} + \ldots + \left(\rho_{0R_0}^0 \right)^{-T} \right] \right). \qquad \text{(B.2.8)}
\end{aligned}$$

Note that when $\frac{1}{T} \boldsymbol{B}^{0\prime} \boldsymbol{B}^0 = O_p(1)$, we have because $\|\boldsymbol{H}_B - \boldsymbol{H}_u\| = O_p(1)$ instead of $o_p(1)$. But this is the case where all factors exhibit a unit root process that is considered in Bai and Ng (2004) and corresponds to the case stated in Lemmas B.1.5 and B.1.6 in Section B.1.1.

Now, define

$$\left(\boldsymbol{H}^{-1} \right)' = N^{1-p} \boldsymbol{V}_{Z,N} \mathcal{J}_T^{-1} \boldsymbol{H}_B^{-1} \qquad \text{(B.2.9)}$$

Note that

$$\begin{aligned}
\left(\boldsymbol{H}^{-1} \right)' &= [N^{1-p} \left(\boldsymbol{\rho}_0^0 \right)^{-T} \boldsymbol{V}_{Z,N} \left(\boldsymbol{\rho}_0^0 \right)^{-T}] \left(\boldsymbol{\rho}_0^0 \right)^{2T} \mathcal{J}_T^{-1} \boldsymbol{H}_B^{-1} \\
&= [N^{1-p} \left(\boldsymbol{\rho}_0^0 \right)^{-T} \boldsymbol{V}_{Z,N} \left(\boldsymbol{\rho}_0^0 \right)^{-T}] \left(\boldsymbol{\rho}_0^0 \right)^{T} \mathcal{J}_B^{-1} \boldsymbol{H}_B^{-1}, \qquad \text{(B.2.10)}
\end{aligned}$$

where equality holds by the fact that $\boldsymbol{V}_{Z,N}$ and $\mathcal{J}_T^{-1}$ are both diagonal matrix and $\mathcal{J}_T = \left(\boldsymbol{\rho}_0^0 \right)^T \mathcal{J}_B$. According to Lemma B.3.3 and the fact shown above that

$$\left\| \boldsymbol{H}_B^{-1} \right\| = O_p(1) + O_p \left(T^{1/2} \left[\left(\rho_{01}^0 \right)^{-T} + \ldots + \left(\rho_{0R_0}^0 \right)^{-T} \right] \right),$$

it follows directly

$$\left\|\boldsymbol{H}^{-1}\right\| = O_p\left(\mathcal{J}_B^{-1}\left[\left(\rho_{01}^0\right)^T + \ldots + \left(\rho_{0R_0}^0\right)^T\right]\right) + O_p\left(\mathcal{J}_B^{-1}T^{1/2}\right) \tag{B.2.11}$$

where the latter term in above expression are of small order compared with the former terms in above expression in the case that factors follow mildly explosive processes.

Then by (B.2.9), (B.2.8), and (B.2.7) and the fact that $\|\boldsymbol{u}\|/\sqrt{T} = O_p(1)$, we have

$$\begin{aligned}
\left\|\widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1} - \left(\boldsymbol{B}^0 + \boldsymbol{u}\right)\right\| &\leq \left\|\widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1} - \boldsymbol{B}^0 - \boldsymbol{u}\boldsymbol{H}_u\boldsymbol{H}_B^{-1}\right\| + \left\|\boldsymbol{u}\left(\boldsymbol{H}_u\boldsymbol{H}_B^{-1} - \boldsymbol{I}\right)\right\| \\
&= O_p\left(N^{-p/2}T^{1/2}\right) + O_p\left(T\left[\left(\rho_{01}^0\right)^{-T} + \ldots + \left(\rho_{0R_0}^0\right)^{-T}\right]\right) \\
&= O_p\left(N^{-p/2}T^{1/2}\right). \qquad\qquad (\text{B.2.12})
\end{aligned}$$

That is, $\frac{1}{T}\left\|\widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1} - \left(\boldsymbol{B}^0 + \boldsymbol{u}\right)\right\|^2 = O_p\left(N^{-p} + T[\left(\rho_{01}^0\right)^{-2T} + \ldots + \left(\rho_{0R_0}^0\right)^{-2T}]\right)$. The desired result follows by the fact that $\boldsymbol{f}_t^0 = \boldsymbol{B}_t^0 + \boldsymbol{u}_t$.

Proofs of Lemma B.1.2(b)

By (B.2.2), (B.2.5), and Lemma B.3.3, we have that for each t,

$$\begin{aligned}
&\left\{\left[N^{1-p}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{V}_{Z,N}\left(\boldsymbol{\rho}_0^0\right)^{-T}\right]\left[\left(\boldsymbol{\rho}_0^0\right)^{2T}\mathcal{J}_T^{-1}\right]\right\}'\widehat{\boldsymbol{B}}_t \\
&-\left[\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0}{N^p}\left(\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}\right) + \frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0}{N^p}\boldsymbol{u}'\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}\right]'\boldsymbol{B}_t^0 + \left[\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0}{N^p}\left(\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}\right)\right]'\boldsymbol{u}_t \\
\equiv\ &\left\{\left[N^{1-p}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{V}_{Z,N}\left(\boldsymbol{\rho}_0^0\right)^{-T}\right]\left[\left(\boldsymbol{\rho}_0^0\right)^{2T}\mathcal{J}_T^{-1}\right]\right\}'\widehat{\boldsymbol{B}}_t - \boldsymbol{H}_B'\boldsymbol{B}_t^0 - \boldsymbol{H}_u'\boldsymbol{u}_t \\
=\ &\left(N^{-p}\mathcal{J}_T^{-1}\right)\widehat{\boldsymbol{B}}'\left(\boldsymbol{\Lambda}^0\boldsymbol{u}' + \boldsymbol{\epsilon}\right)'\left(\boldsymbol{\Lambda}^0\boldsymbol{u}_t + \boldsymbol{\epsilon}_t\right) + \left(N^{-p}\mathcal{J}_T^{-1}\right)\widehat{\boldsymbol{B}}'\boldsymbol{\epsilon}'\boldsymbol{\Lambda}^0\boldsymbol{B}_t^0 + \left(N^{-p}\mathcal{J}_T^{-1}\right)\widehat{\boldsymbol{B}}'\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}_t \\
\equiv\ &\left(N^{-p}\mathcal{J}_T^{-1}\right)\left(\mathbb{A}_{1t} + \mathbb{A}_{2t} + \mathbb{A}_{3t}\right), \qquad\qquad (\text{B.2.13})
\end{aligned}$$

where we use the fact that $\mathcal{J}_T^{-1} = \left(\boldsymbol{\rho}_0^0\right)^T\mathcal{J}_B$ and $\boldsymbol{V}_{Z,N}$ are both diagonal matrix, implying that $\boldsymbol{V}_{Z,N}\mathcal{J}_T^{-1} = \mathcal{J}_T^{-1}\boldsymbol{V}_{Z,N}$.

We first study $\mathbb{A}_{1t}$ by making the following decomposition:

$$\mathbb{A}_{1t} = \widehat{\boldsymbol{B}}'\boldsymbol{\epsilon}'\boldsymbol{\epsilon}_t + \widehat{\boldsymbol{B}}'\boldsymbol{u}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}_t + \widehat{\boldsymbol{B}}'\boldsymbol{\epsilon}'\boldsymbol{\Lambda}^0\boldsymbol{u}_t + \widehat{\boldsymbol{B}}'\boldsymbol{u}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\boldsymbol{u}_t \equiv \mathbb{A}_{1t1} + \mathbb{A}_{1t2} + \mathbb{A}_{1t3} + \mathbb{A}_{1t4}.$$

Noting that $\widehat{\boldsymbol{B}} = (\widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1} - \boldsymbol{f}^0)\boldsymbol{H}' + \boldsymbol{f}^0\boldsymbol{H}'$, we make further decomposition for $\mathbb{A}_{1t1}$:

$$\mathbb{A}_{1t1} = \boldsymbol{H}\boldsymbol{f}^{0\prime}\boldsymbol{\epsilon}'\boldsymbol{\epsilon}_t + \boldsymbol{H}\left(\widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1} - \boldsymbol{f}^0\right)'\boldsymbol{\epsilon}'\boldsymbol{\epsilon}_t = \mathbb{A}_{1t1a} + \mathbb{A}_{1t1b}.$$

Noting that $\boldsymbol{H} = (N^{1-p})^{-1}\boldsymbol{V}_{Z,N}^{-1}\mathcal{J}_T\boldsymbol{H}_B'$ by (B.2.9), and by Lemma B.3.3, we can show that

$$\begin{aligned}
\|\boldsymbol{H}\| &= \left\|\left[N^{1-p}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{V}_{Z,N}\left(\boldsymbol{\rho}_0^0\right)^{-T}\right]^{-1}\left(\boldsymbol{\rho}_0^0\right)^{-T}\mathcal{J}_B\boldsymbol{H}_B'\right\| \\
&\leq \left\|\left[N^{1-p}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{V}_{Z,N}\left(\boldsymbol{\rho}_0^0\right)^{-T}\right]^{-1}\right\|\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|\|\mathcal{J}_B\|\|\boldsymbol{H}_B\| \\
&= O_p\left(\mathcal{J}_B\left[\left(\rho_{01}^0\right)^{-T} + \ldots + \left(\rho_{0R_0}^0\right)^{-T}\right]\right) \qquad \text{(B.2.14)}
\end{aligned}$$

where we use the fact that $\|\boldsymbol{H}_B\| = O_p(1) + O_p\left(T^{1/2}\left[\left(\rho_{01}^0\right)^{-T} + \ldots + \left(\rho_{0R_0}^0\right)^{-T}\right]\right) = O_p(1)$. Recalling that $\gamma_N(s,t) = \sum_{i=1}^N E(\epsilon_{is}\epsilon_{it})$, we make further decomposition for $\mathbb{A}_{1t1a}$:

$$\mathbb{A}_{1t1a} = \sum_{s=1}^T \boldsymbol{H}\boldsymbol{f}_s^0\gamma_N(s,t) + \sum_{s=1}^T \boldsymbol{H}\boldsymbol{f}_s^0\left(\sum_{i=1}^N \epsilon_{is}\epsilon_{it} - \gamma_N(s,t)\right) \equiv \mathbb{A}_{1t1a1} + \mathbb{A}_{1t1a2}.$$

Note that $\sum_{s=1}^T |\gamma_N(s,t)|^2 \leq (\max_{s,t}|\gamma_N(s,t)|)\sum_{s=1}^T |\gamma_N(s,t)| = O(N^2)$ by Assumption A2.3(b). In addition,

$$\begin{aligned}
\sum_{s=1}^T \left\|\boldsymbol{H}\boldsymbol{f}_s^0\right\|^2 &= \sum_{s=1}^T \boldsymbol{f}_s^{0\prime}\boldsymbol{H}'\boldsymbol{H}\boldsymbol{f}_s^0 = \sum_{s=1}^T \boldsymbol{f}_s^{0\prime}\left(\boldsymbol{\rho}_0^0\right)^{-T}\left(\left(\boldsymbol{\rho}_0^0\right)^T\boldsymbol{H}'\right)\left(\left(\boldsymbol{\rho}_0^0\right)^T\boldsymbol{H}'\right)'\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_s^0 \\
&\leq \left\|\left(\boldsymbol{\rho}_0^0\right)^T\boldsymbol{H}'\right\|_{sp}^2\sum_{s=1}^T \boldsymbol{f}_s^{0\prime}\left(\boldsymbol{\rho}_0^0\right)^{-T}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_s^0 \\
&\leq O_p\left(\mathcal{J}_B^2\right)\sum_{s=1}^T \boldsymbol{f}_s^{0\prime}\left(\boldsymbol{\rho}_0^0\right)^{-T}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_s^0 = O_p\left(J_B^2\right)
\end{aligned}$$

where the second inequality holds by (B.2.14) and Assumption Assumption A2.5(a), the final equality follows by Lemma B.3.2(e). Then by the Cauchy-Schwarz inequality,

$$\|\mathbb{A}_{1t1a1}\| = \left\|\sum_{s=1}^{T} \boldsymbol{H}\boldsymbol{f}_s^0 \gamma_N(s,t)\right\| \leq \left[\sum_{s=1}^{T} \|\boldsymbol{H}\boldsymbol{f}_s^0\|^2\right]^{1/2} \left[\sum_{s=1}^{T} |\gamma_N(s,t)|^2\right]^{1/2} = O_p(J_B N).$$

For $\mathbb{A}_{1t1a2}$, we have by the Cauchy-Schwarz inequality

$$\|\mathbb{A}_{1t1a2}\| \leq \left[\sum_{s=1}^{T} \|\boldsymbol{H}\boldsymbol{f}_s^0\|^2\right]^{1/2} \left[\sum_{s=1}^{T} \left|\sum_{i=1}^{N} [\epsilon_{is}\epsilon_{it} - E(\epsilon_{is}\epsilon_{it})]\right|^2\right]^{1/2} = O_p\left(\mathcal{J}_B N^{1/2} T^{1/2}\right)$$

where we use the fact that $\sum_{s=1}^{T} \|\boldsymbol{H}\boldsymbol{f}_s^0\| = O_p(J_B)$ and Assumption A2.3(c).

Now, write $\mathbb{A}_{1t1b} = \sum_{s=1}^{T}(\widehat{\boldsymbol{B}}_s - \boldsymbol{H}\boldsymbol{f}_s^0)\gamma_N(s,t) + \sum_{s=1}^{T}(\widehat{\boldsymbol{B}}_s - \boldsymbol{H}\boldsymbol{f}_s^0)[\sum_{i=1}^{N} \epsilon_{is}\epsilon_{it} - \gamma_N(s,t)]$ $\equiv \mathbb{A}_{1t1b1} + \mathbb{A}_{1t1b2}$. By (B.2.12) and (B.2.14),

$$\begin{aligned} \left\|\widehat{\boldsymbol{B}} - \boldsymbol{f}^0\boldsymbol{H}'\right\| &\leq \left\|\widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1} - \boldsymbol{f}^0\right\| \|\boldsymbol{H}\| \\ &= O_p\left(\mathcal{J}_B N^{-p/2} T^{1/2} \left[\left(\rho_{01}^0\right)^{-T} + \ldots + \left(\rho_{0R_0}^0\right)^{-T}\right]\right). \end{aligned} \quad \text{(B.2.15)}$$

This, in conjunction with Cauchy-Schwarz inequality and Assumption A2.3(b), implies that

$$\begin{aligned} \|\mathbb{A}_{1t1b1}\| &\leq \left\|\widehat{\boldsymbol{B}} - \boldsymbol{f}^0\boldsymbol{H}'\right\| \left[\sum_{s=1}^{T} |\gamma_N(s,t)|^2\right]^{1/2} \\ &= O_p\left(\mathcal{J}_B N^{-p/2} T^{1/2} \left[\left(\rho_{01}^0\right)^{-T} + \ldots + \left(\rho_{0R_0}^0\right)^{-T}\right]\right) O_p(N). \end{aligned} \quad \text{(B.2.16)}$$

Similarly, by Assumption A2.3(c) we have

$$\begin{aligned} \|\mathbb{A}_{1t1b2}\| &\leq \left\|\widehat{\boldsymbol{B}} - \boldsymbol{f}^0\boldsymbol{H}'\right\| \left[\sum_{s=1}^{T} \left|\sum_{i=1}^{N} [\epsilon_{is}\epsilon_{it} - E(\epsilon_{is}\epsilon_{it})]\right|^2\right]^{1/2} \\ &= O_p\left(\mathcal{J}_B N^{-p/2} T^{1/2} \left[\left(\rho_{01}^0\right)^{-T} + \ldots + \left(\rho_{0R_0}^0\right)^{-T}\right]\right) O_p\left((NT)^{1/2}\right). \end{aligned}$$

In sum, we have

$$\begin{aligned}\|\mathbb{A}_{1t1}\| &= O_p\left(J_B N + J_B N^{1/2}T^{1/2}\right) \qquad \text{(B.2.17)}\\ &+O_p\left(\mathcal{J}_B N^{1/2-p/2}T^{1/2}(N^{1/2}+T^{1/2})\left[\left(\rho^0_{01}\right)^{-T}+\ldots+\left(\rho^0_{0R_0}\right)^{-T}\right]\right).\end{aligned}$$

For $\mathbb{A}_{1t2}$, we have $\mathbb{A}_{1t2} = \boldsymbol{H}\boldsymbol{f}^{0\prime}\boldsymbol{u}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}_t + \boldsymbol{H}(\widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1} - \boldsymbol{f}^0)'\boldsymbol{u}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}_t \equiv \mathbb{A}_{1t2a} + \mathbb{A}_{1t2b}$. Note that

$$\mathbb{A}_{1t2a} = \boldsymbol{H}\left(\boldsymbol{\rho}^0_0\right)^T \sum_{s=1}^{T}\left(\boldsymbol{\rho}^0_0\right)^{-T}\boldsymbol{f}^0_s\boldsymbol{u}'_s\sum_{i=1}^{N}\boldsymbol{\lambda}^0_i\epsilon_{it} \equiv \boldsymbol{H}\left(\boldsymbol{\rho}^0_0\right)^T\bar{\mathbb{A}}_{1t2a}.$$

Note that $E\left(\bar{\mathbb{A}}_{1t2a}\right) = \mathbf{0}$ under Assumptions A2.1(c) and A2.4 and

$$\begin{aligned}E\left\|\bar{\mathbb{A}}_{1t2a}\right\|^2 &= Etr\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\epsilon_{it}\boldsymbol{\lambda}^{0\prime}_i\sum_{s=1}^{T}\sum_{q=1}^{T}\boldsymbol{u}_q\boldsymbol{f}^{0\prime}_q\left(\boldsymbol{\rho}^0_0\right)^{-T}\left(\boldsymbol{\rho}^0_0\right)^{-T}\boldsymbol{f}^0_s\boldsymbol{u}'_s\boldsymbol{\lambda}^0_j\epsilon_{jt}\right]\\ &= Etr\left[\sum_{i=1}^{N}\epsilon^2_{it}\boldsymbol{\lambda}^0_i\boldsymbol{\lambda}^{0\prime}_i\sum_{s=1}^{T}\sum_{q=1}^{T}\boldsymbol{u}_q\boldsymbol{f}^{0\prime}_q\left(\boldsymbol{\rho}^0_0\right)^{-T}\left(\boldsymbol{\rho}^0_0\right)^{-T}\boldsymbol{f}^0_s\boldsymbol{u}'_s\right]\\ &= Etr\left[\sum_{i=1}^{N}\sigma^2_i\boldsymbol{\lambda}^0_i\boldsymbol{\lambda}^{0\prime}_i\sum_{s=1}^{T}\sum_{q=1}^{T}\boldsymbol{u}_q\boldsymbol{f}^{0\prime}_q\left(\boldsymbol{\rho}^0_0\right)^{-T}\left(\boldsymbol{\rho}^0_0\right)^{-T}\boldsymbol{f}^0_s\boldsymbol{u}'_s\right]\\ &\le M\cdot N^pE\left\|\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0}{N^p}\right\|E\left\|\sum_{q=1}^{T}\boldsymbol{u}_q\boldsymbol{f}^{0\prime}_q\left(\boldsymbol{\rho}^0_0\right)^{-T}\right\|^2\\ &\le O\left(N^p\right)O\left(1\right) = O\left(N^p\right),\end{aligned}$$

where the second equality holds by Assumption A2.1(c) and the law of iterated expectations, the third equality holds by Assumptions A2.3(a) and A2.4, the first inequality holds by Assumption A2.3(a), the second inequality holds by Assumption A2.2, and and Lemma B.3.2(c). In addition, $\left\|\boldsymbol{H}'\left(\boldsymbol{\rho}^0_0\right)^T\right\| \le \|\boldsymbol{H}\|\left\|\left(\boldsymbol{\rho}^0_0\right)^T\right\| = O_p\left(\mathcal{J}_B\right)$, where by (B.2.14) and Assumption A2.5(a). It follows that

$$\|\mathbb{A}_{1t2a}\| \le \left\|\boldsymbol{H}\left(\boldsymbol{\rho}^0_0\right)^T\right\|\left\|\bar{\mathbb{A}}_{1t2a}\right\| = O_p\left(\mathcal{J}_B N^{p/2}\right).$$

For $\mathbb{A}_{1t2b}$, we have

$$\begin{aligned}
\|\mathbb{A}_{1t2b}\| &= \left\|\sum_{s=1}^{T}\left(\widehat{\boldsymbol{B}}_s - \boldsymbol{H}\boldsymbol{f}_s^0\right)\sum_{i=1}^{N}\boldsymbol{u}_s\boldsymbol{\lambda}_i^0\epsilon_{it}\right\| \leq \|\boldsymbol{B}-\boldsymbol{f}^0\boldsymbol{H}'\|\left[\sum_{s=1}^{T}\left\|\sum_{i=1}^{N}\boldsymbol{u}_s\boldsymbol{\lambda}_i^0\epsilon_{it}\right\|^2\right]^{1/2} \\
&= O_p\left(\mathcal{J}_B N^{-p/2}T^{1/2}\left[\left(\rho_{01}^0\right)^{-T}+\ldots+\left(\rho_{0R_0}^0\right)^{-T}\right]\right)O_p\left(N^{p/2}T^{1/2}\right) \\
&= O_p\left(\mathcal{J}_B T\left[\left(\rho_{01}^0\right)^{-T}+\ldots+\left(\rho_{0R_0}^0\right)^{-T}\right]\right),
\end{aligned}$$

where the second equality follows from (B.2.15) and the fact that $\sum_{s=1}^{T}\left\|\sum_{i=1}^{N}\boldsymbol{u}_s\boldsymbol{\lambda}_i^0\epsilon_{it}\right\|^2 \leq \sum_{s=1}^{T}\|\boldsymbol{u}_s\|^2\times\left\|\sum_{i=1}^{N}\boldsymbol{\lambda}_i^0\epsilon_{it}\right\|^2 = O_p\left(N^pT\right)$ by Assumption A2.1(b) and similar arguments as used to obtain (B.2.4). In sum, we have

$$\|\mathbb{A}_{1t2}\| = O_p\left(\mathcal{J}_B N^{p/2}\right) + O_p\left(\mathcal{J}_B T\left[\left(\rho_{01}^0\right)^{-T}+\ldots+\left(\rho_{0R_0}^0\right)^{-T}\right]\right).$$

Next, we study $\mathbb{A}_{1t3}$. Note that $\mathbb{A}_{1t3} = \boldsymbol{H}\boldsymbol{f}^{0\prime}\boldsymbol{\epsilon}'\boldsymbol{\Lambda}^0\boldsymbol{u}_t + \boldsymbol{H}(\widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1}-\boldsymbol{f}^0)'\boldsymbol{\epsilon}'\boldsymbol{\Lambda}^0\boldsymbol{u}_t \equiv \mathbb{A}_{1t3a} + \mathbb{A}_{1t3b}$. Write $\mathbb{A}_{1t3a} = \boldsymbol{H}\left(\boldsymbol{\rho}_0^0\right)^T\sum_{s=1}^{T}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_s^0\sum_{i=1}^{N}\epsilon_{is}\boldsymbol{\lambda}_i^{0\prime}\boldsymbol{u}_t \equiv \boldsymbol{H}\left(\boldsymbol{\rho}_0^0\right)^T\bar{\mathbb{A}}_{1t3a}$. As in the analysis of $\bar{\mathbb{A}}_{1t2a}$, we have

$$\begin{aligned}
E\left\|\bar{\mathbb{A}}_{1t3a}\right\|^2 &= Etr\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{u}_t'\boldsymbol{\lambda}_i^0\left(\sum_{s=1}^{T}\sum_{q=1}^{T}\epsilon_{iq}\boldsymbol{f}_q^{0\prime}\left(\boldsymbol{\rho}_0^0\right)^{-T}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_s^0\epsilon_{js}\right)\boldsymbol{\lambda}_j^{0\prime}\boldsymbol{u}_t\right] \\
&= Etr\left[\sum_{i=1}^{N}\boldsymbol{u}_t'\boldsymbol{\lambda}_i^0\boldsymbol{\lambda}_i^{0\prime}\boldsymbol{u}_t\left(\sum_{s=1}^{T}\epsilon_{is}^2\boldsymbol{f}_s^{0\prime}\left(\boldsymbol{\rho}_0^0\right)^{-T}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_s^0\right)\right] \\
&= Etr\left[\boldsymbol{u}_t\boldsymbol{u}_t'\sum_{i=1}^{N}\sigma_i^2\boldsymbol{\lambda}_i^0\boldsymbol{\lambda}_i^{0\prime}\left(\sum_{s=1}^{T}\boldsymbol{f}_s^{0\prime}\left(\boldsymbol{\rho}_0^0\right)^{-T}\left(\boldsymbol{\rho}_0^0\right)^{-T}\right)\right] \\
&\leq M\cdot N^p\left\|\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0}{N^p}\right\|E\left\{\|\boldsymbol{u}_t\|^2\left\|\sum_{s=1}^{T}\boldsymbol{f}_s^{0\prime}\left(\boldsymbol{\rho}_0^0\right)^{-T}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_s^0\right\|\right\} \\
&\leq O_p\left(N^p\right)\left\|\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0}{N^p}\right\|\left[E\|\boldsymbol{u}_t\|^4\right]^{1/2}\left\{E\left\|\sum_{s=1}^{T}\boldsymbol{f}_s^{0\prime}\left(\boldsymbol{\rho}_0^0\right)^{-T}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_s^0\right\|^2\right\}^{1/2} \\
&= O_p\left(N^p\right),
\end{aligned}$$

where the second equality holds by Assumption A2.1(c) and the law of iterated expectationsand the fact that $\sum_{s=1}^{T} \epsilon_{is}^2 \boldsymbol{f}_s^{0\prime} \left(\boldsymbol{\rho}_0^0\right)^{-T} \left(\boldsymbol{\rho}_0^0\right)^{-T} \boldsymbol{f}_s^0$ is a scalar, the first inequality holds by Assumption A2.3(a) and the last equality holds by Assumptions A2.2 and A2.1(a), and the fact that $E\left\|\sum_{s=1}^{T} \boldsymbol{f}_s^{0\prime} \left(\boldsymbol{\rho}_0^0\right)^{-T} \left(\boldsymbol{\rho}_0^0\right)^{-T} \boldsymbol{f}_s^0\right\|^2 = O(1)$ by the proof of Lemma B.3.2(e). This result, along with the probability bound for $\left\|\boldsymbol{H}' \left(\boldsymbol{\rho}_0^0\right)^T\right\|$, implies that

$$\|\mathbb{A}_{1t3a}\| \leq \left\|\boldsymbol{H} \left(\boldsymbol{\rho}_0^0\right)^T\right\| \|\bar{\mathbb{A}}_{1t3a}\| = O_p\left(\mathcal{J}_B N^{p/2}\right).$$

For $\mathbb{A}_{1t3b}$, we have

$$\begin{aligned}
\|\mathbb{A}_{1t3b}\| &\leq \left\|\widehat{\boldsymbol{B}} - \boldsymbol{f}^0 \boldsymbol{H}'\right\| \left[\sum_{s=1}^{T} \left\|\sum_{i=1}^{N} \epsilon_{is} \boldsymbol{\lambda}_i^{0\prime} \boldsymbol{u}_t\right\|^2\right]^{1/2} \\
&= O_p\left(\mathcal{J}_B N^{-p/2} T^{1/2} \left[\left(\rho_{01}^0\right)^{-T} + \ldots + \left(\rho_{0R_0}^0\right)^{-T}\right]\right) O_p\left(N^{p/2} T^{1/2}\right) \\
&= O_p\left(\mathcal{J}_B T \left[\left(\rho_{01}^0\right)^{-T} + \ldots + \left(\rho_{0R_0}^0\right)^{-T}\right]\right),
\end{aligned}$$

where we use (B.2.15) and the fact that

$$\begin{aligned}
E\left(\sum_{s=1}^{T} \left\|\sum_{i=1}^{N} \epsilon_{is} \boldsymbol{\lambda}_i^{0\prime} \boldsymbol{u}_t\right\|^2\right) &= \sum_{s=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} E\left(\epsilon_{is} \boldsymbol{\lambda}_i^{0\prime} \boldsymbol{u}_t \boldsymbol{u}_t' \boldsymbol{\lambda}_j^0 \epsilon_{js}\right) = \sum_{s=1}^{T} \sum_{i=1}^{N} E\left(\epsilon_{is} \boldsymbol{\lambda}_i^{0\prime} \boldsymbol{u}_t \boldsymbol{u}_t' \boldsymbol{\lambda}_i^0 \epsilon_{is}\right) \\
&= \sum_{s=1}^{T} \sum_{i=1}^{N} \sigma_i^2 E\left(\boldsymbol{\lambda}_i^{0\prime} \boldsymbol{u}_t \boldsymbol{u}_t' \boldsymbol{\lambda}_i^0\right) \\
&\leq M \sum_{s=1}^{T} E\|\boldsymbol{u}_t\|^2 \sum_{i=1}^{N} E\left(\boldsymbol{\lambda}_i^{0\prime} \boldsymbol{\lambda}_i^0\right) = O_p\left(N^p T\right),
\end{aligned}$$

where the second equality holds by Assumption A2.1(c) and the law of iterated expectations, the third equality holds by Assumptions A2.3–A2.4, and the final equality holds by Assumptions A2.1(b), A2.2, and A2.4. In sum, we have shown that

$$\|\mathbb{A}_{1t3}\| = O_p\left(\mathcal{J}_B N^{p/2}\right) + O_p\left(\mathcal{J}_B T \left[\left(\rho_{01}^0\right)^{-T} + \ldots + \left(\rho_{0R_0}^0\right)^{-T}\right]\right).$$

Now, we write $\mathbb{A}_{1t4} = \boldsymbol{H} \boldsymbol{f}^{0\prime} \boldsymbol{u} \boldsymbol{\Lambda}^{0\prime} \boldsymbol{\Lambda}^0 \boldsymbol{u}_t + \boldsymbol{H}(\widehat{\boldsymbol{B}} \boldsymbol{H}'^{-1} - \boldsymbol{f}^0)' \boldsymbol{u} \boldsymbol{\Lambda}^{0\prime} \boldsymbol{\Lambda}^0 \boldsymbol{u}_t \equiv \mathbb{A}_{1t4a} + \mathbb{A}_{1t4b}$. For

$\mathbb{A}_{1t4a}$, we have

$$\begin{aligned}\|\mathbb{A}_{1t4a}\| &= \left\|\sum_{s=1}^{T}\boldsymbol{H}\boldsymbol{f}_s^0\boldsymbol{u}_s'\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\boldsymbol{u}_t\right\| \leq \left\|\sum_{s=1}^{T}\boldsymbol{H}\boldsymbol{f}_s^0\boldsymbol{u}_s'\right\| \left\|\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\right\| \|\boldsymbol{u}_t\| \\ &\leq N^p \left\|\boldsymbol{H}\left(\boldsymbol{\rho}_0^0\right)^T\right\| \left\|\sum_{s=1}^{T}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_s^0\boldsymbol{u}_s'\right\| \left\|\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0}{N^p}\right\| \|\boldsymbol{u}_t\| = O_p\left(\mathcal{J}_B N^p\right),\end{aligned}$$

where we use Assumptions A2.1(b) and A2.2, the probability bound $\left\|\boldsymbol{H}\left(\boldsymbol{\rho}_0^0\right)^T\right\| = O_p\left(\mathcal{J}_B\right)$, and the fact that $\left\|\sum_{s=1}^{T}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_s^0\boldsymbol{u}_s'\right\| = O_p\left(1\right)$ by Lemma B.3.2(c). For $\mathbb{A}_{1t4b}$, we have

$$\|\mathbb{A}_{1t4b}\| \leq \left\|\widehat{\boldsymbol{B}} - \boldsymbol{f}^0\boldsymbol{H}'\right\| \left[\sum_{s=1}^{T}\left\|\boldsymbol{u}_s'\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\boldsymbol{u}_t\right\|^2\right]^{1/2} = O_p\left(\mathcal{J}_B N^{p/2}T\left[\left(\rho_{01}^0\right)^{-T} + \ldots + \left(\rho_{0R_0}^0\right)^{-T}\right]\right)$$

where we use (B.2.15) and the fact that

$$\begin{aligned}\sum_{s=1}^{T}\left\|\boldsymbol{u}_s'\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\boldsymbol{u}_t\right\|^2 &= N^{2p}tr\left(\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0}{N^p}\boldsymbol{u}_t\boldsymbol{u}_t'\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0}{N^p}\sum_{s=1}^{T}\boldsymbol{u}_s\boldsymbol{u}_s'\right) \\ &\leq N^{2p}\left\|\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0}{N^p}\right\|^2 \|\boldsymbol{u}_t\|^2\|\boldsymbol{u}\|^2 = O_p\left(N^{2p}T\right),\end{aligned}$$

where the final equality holds by Assumption A2.1(b) and A2.2(b). In sum, we have

$$\|\mathbb{A}_{1t4}\| = O_p\left(\mathcal{J}_B N^p\right) + O_p\left(\mathcal{J}_B N^{p/2}T\left[\left(\rho_{01}^0\right)^{-T} + \ldots + \left(\rho_{0R_0}^0\right)^{-T}\right]\right).$$

Collecting the above results for $\mathbb{A}_{1tl}$, $l = 1, 2, 3, 4$, we obtain

$$\|\mathbb{A}_{1t}\| = O_p\left(J_B(N + N^{1/2}T^{1/2})\right) + O_p\left(\mathcal{J}_B\left[N^{1-p/2}T^{1/2} + N^{p/2}T\right]\left[\left(\rho_{01}^0\right)^{-T} + \ldots + \left(\rho_{0R_0}^0\right)^{-T}\right]\right).$$

Now, we study $\mathbb{A}_{2t}$. Note that $\mathbb{A}_{2t} = \boldsymbol{H}\boldsymbol{f}^{0\prime}\boldsymbol{\epsilon}'\boldsymbol{\Lambda}^0\boldsymbol{B}_t^0 + (\widehat{\boldsymbol{B}} - \boldsymbol{f}^0\boldsymbol{H}')'\boldsymbol{\epsilon}'\boldsymbol{\Lambda}^0\boldsymbol{B}_t^0 \equiv \mathbb{A}_{2t1} + \mathbb{A}_{2t2}$. Note $\mathbb{A}_{2t1} = \boldsymbol{H}\left(\boldsymbol{\rho}_0^0\right)^T\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}^{0\prime}\boldsymbol{\epsilon}'\boldsymbol{\Lambda}^0\boldsymbol{B}_t^0 = \boldsymbol{H}\left(\boldsymbol{\rho}_0^0\right)^T\bar{\mathbb{A}}_{2t1}\boldsymbol{B}_t^0$, where $\bar{\mathbb{A}}_{2t1} = \left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}^{0\prime}\boldsymbol{\epsilon}'\boldsymbol{\Lambda}^0 =$

$\sum_{i=1}^{N}\sum_{s=1}^{T}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_s^0\epsilon_{is}\boldsymbol{\lambda}_i^{0\prime}$.

$$
\begin{aligned}
E\left\|\bar{\mathbb{A}}_{2t1}\right\|^2 &= Etr\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=1}^{T}\sum_{q=1}^{T}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_s^0\epsilon_{is}\boldsymbol{\lambda}_i^{0\prime}\boldsymbol{\lambda}_j^0\epsilon_{jq}\boldsymbol{f}_q^{0\prime}\left(\boldsymbol{\rho}_0^0\right)^{-T}\right)\\
&= Etr\left(\sum_{i=1}^{N}\sigma_i^2\boldsymbol{\lambda}_i^{0\prime}\boldsymbol{\lambda}_i^0\sum_{s=1}^{T}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_s^0\boldsymbol{f}_s^{0\prime}\left(\boldsymbol{\rho}_0^0\right)^{-T}\right)\\
&\leq MN^p\cdot E\left\|\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0}{N^p}\right\|E\left\|\sum_{s=1}^{T}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_s^0\boldsymbol{f}_s^{0\prime}\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\| = O_p\left(N^p\right),
\end{aligned}
$$

where the second equality holds by A2.1(c) and the law of iterated expectations, and the equality follows from Assumption A2.2, and the proof of Lemma B.3.2(e). As a result, we have

$$
\begin{aligned}
\left\|\mathbb{A}_{2t1}\right\| &\leq \left\|\boldsymbol{H}\left(\boldsymbol{\rho}_0^0\right)^T\right\|\left\|\bar{\mathbb{A}}_{2t1}\right\|\left\|\boldsymbol{B}_t^0\right\| = O_p\left(\mathcal{J}_B\right)O_p\left(N^{p/2}\right)O_p\left(\left[\left(\rho_{01}^0\right)^T+\ldots+\left(\rho_{0R_0}^0\right)^T\right]\right)\\
&= O_p\left(\mathcal{J}_B N^{p/2}\left[\left(\rho_{01}^0\right)^T+\ldots+\left(\rho_{0R_0}^0\right)^T\right]\right),
\end{aligned}
$$

where we use the fact that $\left\|\boldsymbol{B}_t^0\right\| \leq O(\left\|\boldsymbol{B}_T^0\right\|) = O_p\left(\left(\rho_{01}^0\right)^T+\ldots+\left(\rho_{0R_0}^0\right)^T\right)$ by direct calculations for $\boldsymbol{B}_T^0$ under Assumption A2.1 and A2.5. Similarly,

$$
\begin{aligned}
\left\|\mathbb{A}_{2t2}\right\| &\leq \left\|\widehat{\boldsymbol{B}}-\boldsymbol{f}^0\boldsymbol{H}'\right\|\left[\sum_{s=1}^{T}\left|\sum_{i=1}^{N}\epsilon_{is}\boldsymbol{\lambda}_i\right|^2\right]^{1/2}\left\|\boldsymbol{B}_t^0\right\|\\
&= O_p\left(\mathcal{J}_B T[\left(\rho_{01}^0\right)^{-T}+\ldots+\left(\rho_{0R_0}^0\right)^{-T}]\right)O_p\left(\left(\rho_{01}^0\right)^T+\ldots+\left(\rho_{0R_0}^0\right)^T\right),
\end{aligned}
$$

where we also use (B.2.15). Then

$$
\begin{aligned}
\max_{1\leq t\leq T}\left\|\mathbb{A}_{2t}\right\| &= O_p\left(\mathcal{J}_B N^{p/2}\left[\left(\rho_{01}^0\right)^T+\ldots+\left(\rho_{0R_0}^0\right)^T\right]\right)\\
&\quad+\left\{O_p\left(\mathcal{J}_B T\left[\left(\rho_{01}^0\right)^{-T}+\ldots+\left(\rho_{0R_0}^0\right)^{-T}\right]\right)\right\}O_p\left(\left(\rho_{01}^0\right)^T+\ldots+\left(\rho_{0R_0}^0\right)^T\right).
\end{aligned}
$$

Next, we study $\mathbb{A}_{3t}$. Note that $\mathbb{A}_{3t} = \boldsymbol{H}\boldsymbol{f}^{0\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^0\boldsymbol{\epsilon}_t + (\widehat{\boldsymbol{B}}-\boldsymbol{f}^0\boldsymbol{H}')'\boldsymbol{B}^0\boldsymbol{\Lambda}^0\boldsymbol{\epsilon}_t \equiv \mathbb{A}_{3t1}+\mathbb{A}_{3t2}$. $\mathbb{A}_{3t1} = \sum_{s=1}^{T}\boldsymbol{H}'\boldsymbol{f}_s^0\boldsymbol{B}_s'\sum_{i=1}^{N}\boldsymbol{\lambda}_i^0\epsilon_{it} = \sum_{s=1}^{T}\boldsymbol{H}'\boldsymbol{B}_s^0\boldsymbol{B}_s^{0\prime}\sum_{i=1}^{N}\boldsymbol{\lambda}_i^0\epsilon_{it}+\sum_{s=1}^{T}\boldsymbol{H}'\boldsymbol{u}_s\boldsymbol{B}_s'\sum_{i=1}^{N}\boldsymbol{\lambda}_i^0\epsilon_{it} \equiv$

$\mathbb{A}_{3t1a}+\mathbb{A}_{3t1b}$ as $\boldsymbol{f}_t^0=\boldsymbol{B}_t^0+\boldsymbol{u}_t$.

$$\begin{aligned}\left\|\mathbb{A}_{3t1a}\right\| &\leq \left\|\boldsymbol{H}\left(\boldsymbol{\rho}_0^0\right)^T\right\|\left\|\sum_{s=1}^T\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{B}_s^0\boldsymbol{B}_s^{0\prime}\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|\left\|\left(\boldsymbol{\rho}_0^0\right)^T\right\|\left\|\sum_{i=1}^N\boldsymbol{\lambda}_i^0\epsilon_{it}\right\| \\ &= O_p\left(\mathcal{J}_B\right)O_p\left(1\right)O_p\left(\left(\rho_{01}^0\right)^T+\ldots+\left(\rho_{0R_0}^0\right)^T\right)O_p\left(N^{p/2}\right) \\ &= O_p\left(\mathcal{J}_BN^{p/2}\left[\left(\rho_{01}^0\right)^T+\ldots+\left(\rho_{0R_0}^0\right)^T\right]\right),\end{aligned}$$

where the first equality holds by the fact that $\left\|\boldsymbol{H}\left(\boldsymbol{\rho}_0^0\right)^T\right\|=O_p\left(\mathcal{J}_B\right)$, Lemma B.3.2(e), and the fact that $\left\|\sum_{i=1}^N\boldsymbol{\lambda}_i^0\epsilon_{it}\right\|=O_p\left(N^{p/2}\right)$ under Assumptions A2.1 and A2.2. It is easy to see that $\mathbb{A}_{3t1b}$ is dominated by $\mathbb{A}_{3t1a}$ in terms of probability order. So we can conclude that $\left\|\mathbb{A}_{3t1}\right\|=O_p\left(\mathcal{J}_BN^{p/2}\left[\left(\rho_{01}^0\right)^T+\ldots+\left(\rho_{0R_0}^0\right)^T\right]\right)$. Similarly,

$$\begin{aligned}\left\|\mathbb{A}_{3t2}\right\| &\leq \left\|\widehat{\boldsymbol{B}}-\boldsymbol{f}^0\boldsymbol{H}'\right\|\left[\sum_{s=1}^T\left|\sum_{i=1}^N\boldsymbol{B}_s^{0\prime}\boldsymbol{\lambda}_i\epsilon_{it}\right|^2\right]^{1/2} \\ &= O_p\left(\mathcal{J}_BT^{1/2}\left[\left(\rho_{01}^0\right)^{-T}+\ldots+\left(\rho_{0R_0}^0\right)^{-T}\right]\right)O_p\left(\left(\rho_{01}^0\right)^T+\ldots+\left(\rho_{0R_0}^0\right)^T\right),\end{aligned}$$

where we use (B.2.15) and the fact that

$$\begin{aligned}& E\left(\sum_{s=1}^T\left|\sum_{i=1}^N\boldsymbol{B}_s^{0\prime}\boldsymbol{\lambda}_i\epsilon_{it}\right|^2\right) \\ =& \sum_{s=1}^T\sum_{i=1}^N\sum_{j=1}^N E\left(\boldsymbol{B}_s^{0\prime}\boldsymbol{\lambda}_i\epsilon_{it}\epsilon_{jt}\boldsymbol{\lambda}_j'\boldsymbol{B}_s^0\right)=\sum_{s=1}^T\sum_{i=1}^N E\left(\boldsymbol{B}_s^{0\prime}\boldsymbol{\lambda}_i\epsilon_{it}\epsilon_{it}\boldsymbol{\lambda}_i'\boldsymbol{B}_s^0\right) \\ =& \sum_{s=1}^T E\left[\boldsymbol{B}_s'\left(\sum_{i=1}^N\epsilon_{it}^2\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i'\right)\boldsymbol{B}_s\right]\leq M\cdot E\left(\left\|\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\right\|\right)E\left\|\sum_{s=1}^T\boldsymbol{B}_s^{0\prime}\boldsymbol{B}_s^0\right\| \\ =& O\left(N^p\right)O_p\left(\left(\rho_{01}^0\right)^{2T}+\ldots+\left(\rho_{0R_0}^0\right)^{2T}\right),\end{aligned}$$

where the second equality holds by A2.1(c) and the law of iterated expectations, the first inequality holds by Assumptions A2.3–A2.4, and the last line follows by Assumption A2.2 and

Lemma B.3.2(e). In sum, we have

$$\begin{aligned}\|\mathbb{A}_{3t}\| &= O_p\left(\mathcal{J}_B N^{p/2}\left[\left(\rho_{01}^0\right)^T + \ldots + \left(\rho_{0R_0}^0\right)^T\right]\right) \\ &\quad + O_p\left(\mathcal{J}_B T^{1/2}\left[\left(\rho_{01}^0\right)^{-T} + \ldots + \left(\rho_{0R_0}^0\right)^{-T}\right]\right) O_p\left(\left(\rho_{01}^0\right)^T + \ldots + \left(\rho_{0R_0}^0\right)^T\right),\end{aligned}$$

and

$$\|\mathbb{A}_{1t} + \mathbb{A}_{2t} + \mathbb{A}_{3t}\| \leq \|\mathbb{A}_{1t}\| + \|\mathbb{A}_{2t}\| + \|\mathbb{A}_{3t}\| \leq O_p\left\{\mathcal{J}_B N^{p/2}[\left(\rho_{01}^0\right)^T + \ldots + \left(\rho_{0R_0}^0\right)^T]\right\}, \tag{B.2.18}$$

where $\mathbb{A}_{2t1}$ and $\mathbb{A}_{3t1}$ are leading terms. Then by (B.2.13),

$$\begin{aligned} & \left\|\boldsymbol{H}^{-1}\widehat{\boldsymbol{B}}_t - \boldsymbol{B}_t^0 - \boldsymbol{u}_t\right\| \\ = & \left\|\boldsymbol{H}_B'^{-1}\left\{\left[N^{1-p}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{V}_{Z,N}\left(\boldsymbol{\rho}_0^0\right)^{-T}\right]\left[\left(\boldsymbol{\rho}_0^0\right)^{2T}\mathcal{J}_T^{-1}\right]'\widehat{\boldsymbol{B}}_t - \boldsymbol{H}_B'\boldsymbol{B}_t^0 - \boldsymbol{H}_B'\boldsymbol{u}_t\right\}\right\| \\ \leq & \left\|\boldsymbol{H}_B'^{-1}\right\|\left\|\left\{\left[N^{1-p}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{V}_{Z,N}\left(\boldsymbol{\rho}_0^0\right)^{-T}\right]\left[\left(\boldsymbol{\rho}_0^0\right)^{2T}\mathcal{J}_T^{-1}\right]\right\}'\widehat{\boldsymbol{B}}_t - \boldsymbol{H}_B'\boldsymbol{B}_t^0 - \boldsymbol{H}_u'\boldsymbol{u}_t\right\| \\ & + \left\|\boldsymbol{H}_u\boldsymbol{H}_B^{-1} - \boldsymbol{I}\right\|\|\boldsymbol{u}_t\| \\ = & \left\|\boldsymbol{H}_B'^{-1}\right\|\left\{N^{-p}\mathcal{J}_T^{-1}\left(\mathbb{A}_{1t} + \mathbb{A}_{2t} + \mathbb{A}_{3t}\right)\right\} + \left\|\boldsymbol{H}_u\boldsymbol{H}_B^{-1} - \boldsymbol{I}\right\|\|\boldsymbol{u}_t\| \\ \leq & \left\|\boldsymbol{H}_B^{-1}\right\| O_p\left(\mathcal{J}_B^{-1}N^{-p}\left[\left(\rho_{01}^0\right)^{-T} + \ldots + \left(\rho_{0R_0}^0\right)^{-T}\right]\right) O_p\left\{\mathcal{J}_B N^{p/2}[\left(\rho_{01}^0\right)^T + \ldots + \left(\rho_{0R_0}^0\right)^T]\right\} \\ & + \left\|\boldsymbol{H}_u\boldsymbol{H}_B^{-1} - \boldsymbol{I}\right\|\|\boldsymbol{u}_t\| \\ = & O_p\left(N^{-p/2}\right), \end{aligned} \tag{B.2.19}$$

where the first equality follows by the construction of $\boldsymbol{H}'^{-1}$ in (B.2.10), the second inequality follows from (B.2.18) and the fact that $\left\|\mathcal{J}_T^{-1}N^{-p}\right\| = O_p\left(\mathcal{J}_B^{-1}N^{-p}\left[\left(\rho_{01}^0\right)^{-T} + \ldots + \left(\rho_{0R_0}^0\right)^{-T}\right]\right)$, and the last equality follows from the bounds of $\left\|\boldsymbol{H}_B^{-1}\right\|$ and $\left\|\boldsymbol{H}_u\boldsymbol{H}_B^{-1} - \boldsymbol{I}\right\|$ derived in the proofs of Lemma B.1.2(a).

Proofs of Lemma B.1.2(c)

Recall that $\widehat{\boldsymbol{\lambda}}_i = (\widehat{\boldsymbol{B}}'\widehat{\boldsymbol{B}})^{-1}\widehat{\boldsymbol{B}}'\boldsymbol{Z}_i$ where $\boldsymbol{Z}_i = \boldsymbol{f}^0\boldsymbol{\lambda}_i^0 + \boldsymbol{\epsilon}_i$. By the identity that $\boldsymbol{f}^0 = (\boldsymbol{f}^0 - \widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1}) + \widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1}$ and $\widehat{\boldsymbol{B}} = (\widehat{\boldsymbol{B}} - \boldsymbol{f}^0\boldsymbol{H}') + \boldsymbol{f}^0\boldsymbol{H}'$, and the normalization condition that $\mathcal{J}_B^{-2}\widehat{\boldsymbol{B}}'\widehat{\boldsymbol{B}} = \boldsymbol{I}_{R_0}$, we have

$$\begin{aligned}
\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{H}'^{-1}\boldsymbol{\lambda}_i^0 &= \mathcal{J}_B^{-2}\boldsymbol{H}(\widehat{\boldsymbol{B}}'\boldsymbol{H}^{-1} - \boldsymbol{f}^0)'(\boldsymbol{f}^0 - \widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1})\boldsymbol{\lambda}_i^0 + \mathcal{J}_B^{-2}\boldsymbol{H}\boldsymbol{f}^{0\prime}(\boldsymbol{f}^0 - \widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1})\boldsymbol{\lambda}_i^0 \\
&\quad + \mathcal{J}_B^{-2}\boldsymbol{H}\boldsymbol{f}^{0\prime}\boldsymbol{\epsilon}_i + \mathcal{J}_B^{-2}(\widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1} - \boldsymbol{f}^0)'\boldsymbol{\epsilon}_i \\
&\equiv \mathbb{D}_{1i} + \mathbb{D}_{2i} + \mathbb{D}_{3i} + \mathbb{D}_{4i}. \qquad \text{(B.2.20)}
\end{aligned}$$

We bound each term in the last display in turn.

First, for $\mathbb{D}_{1i}$ we have

$$\begin{aligned}
\|\mathbb{D}_{1i}\| &\leq \mathcal{J}_B^{-2}\left\|\widehat{\boldsymbol{B}} - \boldsymbol{f}^0\boldsymbol{H}'\right\|\left\|\widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1} - \boldsymbol{f}^0\right\|\|\boldsymbol{\lambda}_i^0\| \\
&= \mathcal{J}_B^{-2}O_p\left(\mathcal{J}_B N^{-p/2}T^{1/2}\left[\left(\rho_{01}^0\right)^{-T} + \ldots + \left(\rho_{0R_0}^0\right)^{-T}\right]\right)O_p\left(N^{-p/2}T^{1/2}\right) \\
&= O_p\left(\mathcal{J}_B^{-1}N^{-p}T\left[\left(\rho_{01}^0\right)^{-T} + \ldots + \left(\rho_{0R_0}^0\right)^{-T}\right]\right),
\end{aligned}$$

where the first equality holds by (B.2.15) and (B.2.12). As we shall see, $\|\mathbb{D}_{1i}\|$ is dominated by the other terms due to the presence of $O_p[(\rho_{01}^0)^{-T} + \ldots + (\rho_{0R_0}^0)^{-T}]$ in the case of mildly explosive factors.

For $\mathbb{D}_{3i}$, we have $\mathbb{D}_{3i} = \mathcal{J}_B^{-2}\boldsymbol{H}\sum_{s=1}^T \boldsymbol{f}_s^0\epsilon_{is} = \boldsymbol{H}\left(\boldsymbol{\rho}_0^0\right)^T\bar{\mathbb{D}}_{3i}$, where $\bar{\mathbb{D}}_{3i} = \mathcal{J}_B^{-2}\sum_{s=1}^T\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_s^0\epsilon_{is}$. Since $E\left(\bar{\mathbb{D}}_{3i}\right) = \boldsymbol{0}$ holds under Assumptions A2.1(c) and A2.4, $\left\|\sum_{s=1}^T\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_s^0\epsilon_{is}\right\| = O_p(1)$ by replacing ϵ_{is} with $\boldsymbol{u}_t$ in Lemma B.3.2(c) and using arguments as used in the proof of Lemma B.5 in Horie and Yamamoto (2016) given that $\boldsymbol{f}_t = \kappa_T^{-1}\boldsymbol{F}_{t-1} + \boldsymbol{u}_t$ for $t = 2, \ldots, T$. Then

$$\|\mathbb{D}_{3i}\| \leq \left\|\boldsymbol{H}\left(\boldsymbol{\rho}_0^0\right)^T\right\|\|\bar{\mathbb{D}}_{3i}\| = O_p\left(\mathcal{J}_B\right)O_p(\mathcal{J}_B^{-2}\kappa_T) = O_p(\mathcal{J}_B^{-1})$$

where we use the fact that $\left\|\boldsymbol{H}\left(\boldsymbol{\rho}_0^0\right)^T\right\| = O_p\left(\mathcal{J}_B\right)$.

For $\mathbb{D}_{4i}$, we make the following decomposition

$$\begin{aligned}
\mathbb{D}_{4i} &= \mathcal{J}_B^{-2}\Big\{\left(N^{-p}\mathcal{J}_T^{-1}\right)\widehat{\boldsymbol{B}}'\left(\boldsymbol{\Lambda}^0\boldsymbol{u}'+\boldsymbol{\epsilon}\right)'\left(\boldsymbol{\Lambda}^0\boldsymbol{u}'+\boldsymbol{\epsilon}\right)\boldsymbol{\epsilon}_i+\left(N^{-p}\mathcal{J}_T^{-1}\right)\widehat{\boldsymbol{B}}'\boldsymbol{\epsilon}'\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}\boldsymbol{\epsilon}_i \\
&\quad+\left(N^{-p}\mathcal{J}_T^{-1}\right)\widehat{\boldsymbol{B}}'\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}\boldsymbol{\epsilon}_i\Big\} \\
&\equiv \mathcal{J}_B^{-2}\left\{\mathbb{D}_{4ia}+\mathbb{D}_{4ib}+\mathbb{D}_{4ic}\right\}.
\end{aligned}$$

Recall that in the Proofs of Lemma B.1.2(b), $\mathbb{A}_{2t1}$ and $\mathbb{A}_{3t1}$ are the leading terms. Here, the leading terms in $\mathbb{D}_{4i}$ are closely related to $\mathbb{A}_{2t1}$ and $\mathbb{A}_{3t1}$. In addition, it is easy to see that $\mathbb{D}_{4ia}$ will be dominated by $\mathbb{D}_{4ib}$ and $\mathbb{D}_{4ic}$.

For $\mathbb{D}_{4ib}$, we have

$$\mathbb{D}_{4ib}=\left(N^{-p}\mathcal{J}_T^{-1}\right)\boldsymbol{H}\boldsymbol{f}^{0\prime}\boldsymbol{\epsilon}'\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}\boldsymbol{\epsilon}_i+\left(N^{-p}\mathcal{J}_T^{-1}\right)(\widehat{\boldsymbol{B}}-\boldsymbol{f}^0\boldsymbol{H}')'\boldsymbol{\epsilon}'\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}\boldsymbol{\epsilon}_i\equiv\mathbb{D}_{4ib1}+\mathbb{D}_{4ib2}.$$

It is easy to show that $\mathbb{D}_{4ib2}$ is dominated by $\mathbb{D}_{4ib1}$ and so we focus on $\mathbb{D}_{4ib1}$. As the analysis of $\mathbb{A}_{2t1}$ in the proof of Lemma B.2(b), we have

$$\begin{aligned}
&\|\mathbb{D}_{4ib1}\| \\
&= \left\|N^{-p}\mathcal{J}_T^{-1}\boldsymbol{H}\left(\boldsymbol{\rho}_0^0\right)^T\sum_{j=1}^N\sum_{s=1}^T\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_s^0\epsilon_{js}\boldsymbol{\lambda}_j^{0\prime}\sum_{t=2}^T\boldsymbol{B}_t^0\epsilon_{it}\right\| \\
&\le N^{-p}\left\|\mathcal{J}_T^{-1}\right\|\left\|\boldsymbol{H}\left(\boldsymbol{\rho}_0^0\right)^T\right\|\left\|\sum_{j=1}^N\sum_{s=1}^T\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_s^0\epsilon_{js}\boldsymbol{\lambda}_j^{0\prime}\right\|\left\|\sum_{t=2}^T\boldsymbol{B}_t^0\epsilon_{it}\right\| \\
&= N^{-p}O_p\left(\mathcal{J}_B^{-1}\left[\left(\rho_{01}^0\right)^{-T}+\ldots+\left(\rho_{0R_0}^0\right)^{-T}\right]\right)O_p\left(\mathcal{J}_B\right)O_p\left(N^{p/2}\right)O_p\left[\left(\rho_{01}^0\right)^{T}+\ldots+\left(\rho_{0R_0}^0\right)^{T}\right] \\
&= O_p\left(N^{-p/2}\left[\left(\rho_{01}^0\right)^{-T}+\ldots+\left(\rho_{0R_0}^0\right)^{-T}\right]\right)O_p\left[\left(\rho_{01}^0\right)^{T}+\ldots+\left(\rho_{0R_0}^0\right)^{T}\right] \\
&= O_p\left(N^{-p/2}\right),
\end{aligned}$$

where we use the fact that $\left\|\sum_{j=1}^N\sum_{s=1}^T\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_s^0\epsilon_{js}\boldsymbol{\lambda}_j^{0\prime}\right\|=O_p\left(N^{p/2}\right)$ and that $\left\|\sum_{t=2}^T\boldsymbol{B}_t^0\epsilon_{it}\right\|=$

$O_p[(\rho_{01}^0)^T + \ldots + \left(\rho_{0R_0}^0\right)^T]$ by following similar arguments for Lemma B.3.2(c). Similarly,

$$\begin{aligned}
\mathbb{D}_{4ic} &= \left(N^{-p}\mathcal{J}_T^{-1}\right)\boldsymbol{H}\boldsymbol{f}^{0\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}\boldsymbol{\epsilon}_i + \left(N^{-p}\mathcal{J}_T^{-1}\right)(\widehat{\boldsymbol{B}} - \boldsymbol{f}^0\boldsymbol{H}')'\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}\boldsymbol{\epsilon}_i \\
&= \left(N^{-p}\mathcal{J}_T^{-1}\right)\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}\boldsymbol{\epsilon}_i + \left(N^{-p}\mathcal{J}_T^{-1}\right)\boldsymbol{H}\boldsymbol{u}'\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}\boldsymbol{\epsilon}_i \\
&\quad + \left(N^{-p}\mathcal{J}_T^{-1}\right)(\widehat{\boldsymbol{B}} - \boldsymbol{f}^0\boldsymbol{H}')'\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}\boldsymbol{\epsilon}_i \\
&\equiv \mathbb{D}_{4ic1} + \mathbb{D}_{4ic2} + \mathbb{D}_{4ic3}.
\end{aligned}$$

It is easy to show that $\mathbb{D}_{4ic2}$ and $\mathbb{D}_{4ic3}$ are dominated by $\mathbb{D}_{4ic1}$ and so we focus on $\mathbb{D}_{4ic1}$. As the analysis of $\mathbb{A}_{3t1}$ in the proof of Lemma B.2(b), we have $\mathbb{D}_{4ic1} = \left(N^{-p}\mathcal{J}_T^{-1}\right)\boldsymbol{H}\left(\boldsymbol{\rho}_0^0\right)^T\bar{\mathbb{D}}_{4ic1}$ where $\bar{\mathbb{D}}_{4ic1} = \left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}^{0\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}\boldsymbol{\epsilon}_i$.

$$\begin{aligned}
\left\|\mathbb{D}_{4ic1}\right\| &= \left\|\mathcal{J}_T^{-1}N^{-p}\sum_{s=1}^T \boldsymbol{H}\boldsymbol{B}_s^0\boldsymbol{B}_s^{0\prime}\sum_{j=1}^N\boldsymbol{\lambda}_j^0\sum_t \epsilon_{jt}\epsilon_{it}\right\| \\
&\leq \left\|\mathcal{J}_T^{-1}N^{-p}\sum_{s=1}^T \boldsymbol{H}\boldsymbol{B}_s^0\boldsymbol{B}_s^{0\prime}\right\|\left\|\sum_{j=1}^N\boldsymbol{\lambda}_j^0\sum_t \epsilon_{jt}\epsilon_{it}\right\| \\
&= O_p\left(N^{-p}\right)O_p\left(N^{p/2}T^{1/2}\right),
\end{aligned}$$

where the bound of $\left\|\mathcal{J}_T^{-1}N^{-p}\sum_{s=1}^T \boldsymbol{H}\boldsymbol{B}_s^0\boldsymbol{B}_s^{0\prime}\right\|$ follows directly from the same arguments as used in the analysis of $\mathbb{A}_{3t1}$ in the proof of Lemma B.2(b), and the bound of $\left\|\sum_{j=1}^N\boldsymbol{\lambda}_j^0\sum_t \epsilon_{jt}\epsilon_{it}\right\|$ is obtained as follows:

$$\begin{aligned}
E\left\|\sum_{j=1}^N\boldsymbol{\lambda}_j^0\sum_t \epsilon_{jt}\epsilon_{it}\right\|^2 &= \sum_{j=1}^N\sum_{l=1}^N\sum_{t=2}^T\sum_{q=1}^T E\left(\boldsymbol{\lambda}_l^{0\prime}\boldsymbol{\lambda}_j^0\epsilon_{jt}\epsilon_{it}\epsilon_{iq}\epsilon_{lq}\right) = \sum_{j=1}^N\sum_{t=2}^T E\left(\boldsymbol{\lambda}_j^{0\prime}\boldsymbol{\lambda}_j^0\epsilon_{jt}^2\epsilon_{it}^2\right) \\
&= \sum_{j=1}^N E\left(\boldsymbol{\lambda}_j^{0\prime}\boldsymbol{\lambda}_j^0\right)\sum_{t=2}^T E\left(\epsilon_{jt}^2\epsilon_{it}^2\right) = O_p\left(N^pT\right),
\end{aligned}$$

where the second equality holds by Assumption A2.1 and the law of iterated expectations, and the third equality holds by Assumptions A2.1(a), A2.2 and A2.4. It follows that $\left\|\mathbb{D}_{4i}\right\| = O_p\left(\mathcal{J}_B^{-2}N^{-p/2}T^{1/2}\right)$.

For $\mathbb{D}_{2i}$, we make the following decomposition.

$$\begin{aligned}
\mathbb{D}_{2i} &= \mathcal{J}_B^{-2}\boldsymbol{H}\boldsymbol{f}^{0\prime}(\widehat{\boldsymbol{B}}\boldsymbol{H}^{\prime -1} - \boldsymbol{f}^0)\boldsymbol{\lambda}_i \\
&= \mathcal{J}_B^{-2}\Big\{N^{-p}\boldsymbol{H}\boldsymbol{f}^{0\prime}\left(\boldsymbol{\Lambda}^0\boldsymbol{u}' + \boldsymbol{\epsilon}\right)'\left(\boldsymbol{\Lambda}^0\boldsymbol{u} + \boldsymbol{\epsilon}\right)\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}\boldsymbol{\lambda}_i + \left(N^{-p}\right)\boldsymbol{H}\boldsymbol{f}^{0\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}\boldsymbol{\lambda}_i \\
&\qquad + N^{-p}\boldsymbol{H}\boldsymbol{f}^{0\prime}\boldsymbol{\epsilon}'\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}\boldsymbol{\lambda}_i\Big\} \\
&= \mathbb{D}_{2ia} + \mathbb{D}_{2ib} + \mathbb{D}_{2ic}.
\end{aligned}$$

As in the analysis of $\mathbb{D}_{4i}$, we can show that $\mathbb{D}_{2ia}$ is dominated by $\mathbb{D}_{2ib}$ and $\mathbb{D}_{2ic}$. For $\mathbb{D}_{2ib}$ and , we have

$$\begin{aligned}
\mathbb{D}_{2ib} &= N^{-p}\mathcal{J}_B^{-2}\boldsymbol{H}\boldsymbol{f}^{0\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}\boldsymbol{\lambda}_i \\
&= N^{-p}\mathcal{J}_B^{-2}\boldsymbol{H}\boldsymbol{f}^{0\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}\boldsymbol{f}^0\boldsymbol{H}'\mathcal{J}_T^{-1}\boldsymbol{\lambda}_i + N^{-p}\mathcal{J}_B^{-2}\boldsymbol{H}\boldsymbol{f}^{0\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}(\widehat{\boldsymbol{B}} - \boldsymbol{f}^0\boldsymbol{H}')\mathcal{J}_T^{-1}\boldsymbol{\lambda}_i \\
&= N^{-p}\mathcal{J}_B^{-2}\left\{\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}\boldsymbol{f}^0\boldsymbol{H}'\mathcal{J}_T^{-1}\boldsymbol{\lambda}_i + \boldsymbol{H}\boldsymbol{u}^{0\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}\boldsymbol{f}^0\boldsymbol{H}'\mathcal{J}_T^{-1}\boldsymbol{\lambda}_i\right. \\
&\qquad \left. + \boldsymbol{H}\boldsymbol{f}^{0\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}(\widehat{\boldsymbol{B}} - \boldsymbol{f}^0\boldsymbol{H}')\mathcal{J}_T^{-1}\boldsymbol{\lambda}_i\right\} \\
&\equiv \mathbb{D}_{2ib,1} + \mathbb{D}_{2ib,2} + \mathbb{D}_{2ib,3},
\end{aligned}$$

where $\mathbb{D}_{2ib,2}$ and $\mathbb{D}_{2ib,3}$ are dominated by $\mathbb{D}_{2ib,1}$. Following the analysis of $\|\mathbb{A}_{2t1}\|$ in the proof of Lemma B.2(b), we have

$$\begin{aligned}
&\|\mathbb{D}_{2ib,1}\| \\
=& \left\|N^{-p}\mathcal{J}_B^{-2}\boldsymbol{\lambda}_i'\mathcal{J}_T^{-1}\sum_{j=1}^{N}\sum_{s=1}^{T}\boldsymbol{H}\boldsymbol{f}_s^0\epsilon_{js}\boldsymbol{\lambda}_j^{0\prime}\sum_{t=2}^{T}\boldsymbol{B}_t^0\boldsymbol{B}_t^{0\prime}\boldsymbol{H}'\right\| \\
\leq& \mathcal{J}_B^{-2}\|\boldsymbol{\lambda}_i\|\left\|\mathcal{J}_T^{-1}N^{-p}\sum_{j=1}^{N}\sum_{s=1}^{T}\boldsymbol{H}\boldsymbol{f}_s^0\epsilon_{js}\boldsymbol{\lambda}_j^{0\prime}\right\|\left\|\sum_{t=2}^{T}\boldsymbol{B}_t^0\boldsymbol{B}_t^{0\prime}\boldsymbol{H}'\right\| \\
=& \mathcal{J}_B^{-2}O_p(1)O_p\left(N^{-p/2}\left[\left(\rho_{01}^0\right)^{-T} + \ldots + \left(\rho_{0R_0}^0\right)^{-T}\right]\right)O_p\left(\mathcal{J}_B\left[\left(\rho_{01}^0\right)^{T} + \ldots + \left(\rho_{0R_0}^0\right)^{T}\right]\right) \\
=& O_p\left(\mathcal{J}_B^{-1}N^{-p/2}\right),
\end{aligned}$$

where $\left\|\mathcal{J}_T^{-1}N^{-p}\sum_{j=1}^{N}\sum_{s=1}^{T}\boldsymbol{H}\boldsymbol{f}_s^0\epsilon_{js}\boldsymbol{\lambda}_j^{0\prime}\right\| = O_p\left(N^{-p/2}[(\rho_{01}^0)^{-T} + \ldots + \left(\rho_{0R_0}^0\right)^{-T}]\right)$ holds by the same argument as used in the study of $\|\mathbb{A}_{2t1}\|$ in the proof of Lemma B.2(b), and that $\left\|\sum_{t=2}^{T}\boldsymbol{B}_t^0\boldsymbol{B}_t^{0\prime}\boldsymbol{H}^{\prime}\right\| = O_p\left(\mathcal{J}_B\left[\left(\rho_{01}^0\right)^T + \ldots + \left(\rho_{0R_0}^0\right)^T\right]\right)$ by arguments as used to bound $\|\mathbb{A}_{3t1}\|$ in the proof of Lemma B.2(b). Similarly,

$$
\begin{aligned}
\mathbb{D}_{2ic} &= N^{-p}\mathcal{J}_B^{-2}\boldsymbol{H}\boldsymbol{f}^{0\prime}\boldsymbol{\epsilon}^{\prime}\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}\boldsymbol{\lambda}_i \\
&= N^{-p}\mathcal{J}_B^{-2}\left\{\boldsymbol{H}\boldsymbol{f}^{0\prime}\boldsymbol{\epsilon}^{\prime}\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}\boldsymbol{f}^0\boldsymbol{H}^{\prime}\mathcal{J}_T^{-1}\boldsymbol{\lambda}_i + \boldsymbol{H}\boldsymbol{f}^{0\prime}\boldsymbol{\epsilon}^{\prime}\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}(\widehat{\boldsymbol{B}} - \boldsymbol{f}^0\boldsymbol{H}^{\prime})\mathcal{J}_T^{-1}\boldsymbol{\lambda}_i\right\} \\
&= N^{-p}\mathcal{J}_B^{-2}\left\{\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{\epsilon}^{\prime}\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}\boldsymbol{f}^0\boldsymbol{H}^{\prime}\mathcal{J}_T^{-1}\boldsymbol{\lambda}_i + \boldsymbol{H}\boldsymbol{u}^{0\prime}\boldsymbol{\epsilon}^{\prime}\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}\boldsymbol{f}^0\boldsymbol{H}^{\prime}\mathcal{J}_T^{-1}\boldsymbol{\lambda}_i\right. \\
&\quad \left. + \boldsymbol{H}\boldsymbol{f}^{0\prime}\boldsymbol{\epsilon}^{\prime}\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}(\widehat{\boldsymbol{B}} - \boldsymbol{f}^0\boldsymbol{H}^{\prime})\mathcal{J}_T^{-1}\boldsymbol{\lambda}_i\right\} \\
&\equiv \mathbb{D}_{2ic,1} + \mathbb{D}_{2ic,2} + \mathbb{D}_{2ic,3},
\end{aligned}
$$

where $\mathbb{D}_{2ic,2}$ and $\mathbb{D}_{2ic,3}$ are dominated by $\mathbb{D}_{2ic,1}$. Note that

$$
\begin{aligned}
&\|\mathbb{D}_{2ic,1}\| \\
= & \left\|N^{-p}\mathcal{J}_B^{-2}\boldsymbol{\lambda}_i^{\prime}\mathcal{J}_T^{-1}\sum_{s=1}^{T}\boldsymbol{H}\boldsymbol{B}_s^0\boldsymbol{B}_s^{0\prime}\sum_{j=1}^{N}\boldsymbol{\lambda}_j^0\sum_t \epsilon_{jt}\boldsymbol{B}_t^{0\prime}\boldsymbol{H}^{\prime}\right\| \\
\leq & N^{-p}\mathcal{J}_B^{-2}\left\|\mathcal{J}_T^{-1}\right\|\|\boldsymbol{\lambda}_i\|\left\|\sum_{s=1}^{T}\boldsymbol{H}\boldsymbol{B}_s^0\boldsymbol{B}_s^{0\prime}\right\|\left\|\sum_{j=1}^{N}\boldsymbol{\lambda}_j^0\sum_t \epsilon_{jt}\boldsymbol{B}_t^{0\prime}\boldsymbol{H}^{\prime}\right\| \\
= & N^{-p}\mathcal{J}_B^{-1}O_p((\rho_{01}^0)^{-T} + \ldots + (\rho_{0R_0}^0)^{-T})O_p(1)\,O_p\left(\mathcal{J}_B\left[\left(\rho_{01}^0\right)^T + \ldots + \left(\rho_{0R_0}^0\right)^T\right]\right)O_p\left(\mathcal{J}_B N^{p/2}\right) \\
= & O_p\left(\mathcal{J}_B^{-1}N^{-p/2}\right),
\end{aligned}
$$

where we use the fact that $\left\|\sum_{j=1}^{N}\boldsymbol{\lambda}_j^0\sum_t \epsilon_{jt}\boldsymbol{B}_t^{0\prime}\boldsymbol{H}^{\prime}\right\| = O_p\left(\mathcal{J}_B N^{p/2}\right)$ by similar arguments as used to bound $\|\mathbb{A}_{1t3a}\|$ in the proof of Lemma B.2(b). In sum, we have $\|\mathbb{D}_{2i}\| = O_p\left(\mathcal{J}_B^{-1}N^{-p/2}\right)$.

Collecting above all leading terms, namely, $\mathbb{D}_{2i}$, $\mathbb{D}_{3i}$, and $\mathbb{D}_{4i}$, it follows that:

$$
\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{H}^{\prime -1}\boldsymbol{\lambda}_i = O_p\left(\mathcal{J}_B^{-1}\left(1 + \mathcal{J}_B^{-1}N^{-p/2}T^{1/2} + N^{-p/2}\right)\right).
$$

B.2.3 Proofs of Lemma B.1.3

By (B.2.13) and (B.2.5), we have

$$
\begin{aligned}
& \left\{\left[N^{1-p}\left(\boldsymbol{\rho}_0^0\right)^{-T} \boldsymbol{V}_{Z,N}\left(\boldsymbol{\rho}_0^0\right)^{-T}\right]\left[\left(\boldsymbol{\rho}_0^0\right)^{2T} \mathcal{J}_T^{-1}\right]\right\}' \sum_{s=2}^{q} \widehat{\boldsymbol{B}}_s \\
& -\left[\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0}{N^p}\left(\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}\right)+\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0}{N^p}\boldsymbol{u}'\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}\right]' \sum_{s=2}^{q} \boldsymbol{B}_s-\left[\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0}{N^p}\left(\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}\right)\right]' \sum_{s=2}^{q} \boldsymbol{u}_s \\
= \; & \left\{\left[N^{1-p}\left(\boldsymbol{\rho}_0^0\right)^{-T} \boldsymbol{V}_{Z,N}\left(\boldsymbol{\rho}_0^0\right)^{-T}\right]\left[\left(\boldsymbol{\rho}_0^0\right)^{2T} \mathcal{J}_T^{-1}\right]\right\}' \sum_{s=2}^{q} \widehat{\boldsymbol{B}}_s-\boldsymbol{H}_B' \sum_{s=2}^{q} \boldsymbol{B}_s^0-\boldsymbol{H}_u' \sum_{s=2}^{q} \boldsymbol{u}_s \\
= \; & \left(N^{-p}\mathcal{J}_T^{-1}\right) \widehat{\boldsymbol{B}}'\left(\boldsymbol{\Lambda}^0\boldsymbol{u}'+\boldsymbol{\epsilon}\right)' \sum_{s=2}^{q}\left(\boldsymbol{\Lambda}^0\boldsymbol{u}_s+\boldsymbol{\epsilon}_s\right)+\left(N^{-p}\mathcal{J}_T^{-1}\right) \widehat{\boldsymbol{B}}'\boldsymbol{\epsilon}'\boldsymbol{\Lambda}^0 \sum_{s=2}^{q} \boldsymbol{B}_s^0 \\
& +\left(N^{-p}\mathcal{J}_T^{-1}\right) \widehat{\boldsymbol{B}}'\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime} \sum_{s=2}^{q} \boldsymbol{\epsilon}_s \\
= \; & \left(N^{-p}\mathcal{J}_T^{-1}\right)\left(\mathbb{C}_{1q}+\mathbb{C}_{2q}+\mathbb{C}_{3q}+\mathbb{C}_{4q}+\mathbb{C}_{5q}+\mathbb{C}_{6q}\right), \qquad \text{(B.2.21)}
\end{aligned}
$$

where $\mathbb{C}_{1q}$ to $\mathbb{C}_{6q}$ are defined as follows:

$$
\begin{aligned}
\mathbb{C}_{1q} &= \boldsymbol{H}\boldsymbol{f}^{0\prime}\boldsymbol{\epsilon}' \sum_{s=1}^{q} \boldsymbol{\epsilon}_s+\left(\widehat{\boldsymbol{B}}-\boldsymbol{f}^0\boldsymbol{H}'\right)' \boldsymbol{\epsilon}' \sum_{s=1}^{q} \boldsymbol{\epsilon}_s \equiv \mathbb{C}_{1q1}+\mathbb{C}_{1q2}, \\
\mathbb{C}_{2q} &= \boldsymbol{H}\boldsymbol{f}^{0\prime}\boldsymbol{u}\boldsymbol{\Lambda}^{0\prime} \sum_{s=1}^{q} \boldsymbol{\epsilon}_s+\left(\widehat{\boldsymbol{B}}-\boldsymbol{f}^0\boldsymbol{H}'\right)' \boldsymbol{u}\boldsymbol{\Lambda}^{0\prime} \sum_{s=1}^{q} \boldsymbol{\epsilon}_s \equiv \mathbb{C}_{2q1}+\mathbb{C}_{2q2}, \\
\mathbb{C}_{3q} &= \boldsymbol{H}\boldsymbol{f}^{0\prime}\boldsymbol{\epsilon}'\boldsymbol{\Lambda}^0 \sum_{s=1}^{q} \boldsymbol{u}_s+\left(\widehat{\boldsymbol{B}}-\boldsymbol{f}^0\boldsymbol{H}'\right)' \boldsymbol{\epsilon}'\boldsymbol{\Lambda}^0 \sum_{s=1}^{q} \boldsymbol{u}_s \equiv \mathbb{C}_{3q1}+\mathbb{C}_{3q2}, \\
\mathbb{C}_{4q} &= \boldsymbol{H}\boldsymbol{f}^{0\prime}\boldsymbol{u}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0 \sum_{s=1}^{q} \boldsymbol{u}_s+\left(\widehat{\boldsymbol{B}}-\boldsymbol{f}^0\boldsymbol{H}'\right)' \boldsymbol{u}\boldsymbol{\Lambda}'\boldsymbol{\Lambda}^0 \sum_{s=1}^{q} \boldsymbol{u}_s \equiv \mathbb{C}_{4q1}+\mathbb{C}_{4q2}, \\
\mathbb{C}_{5q} &= \boldsymbol{H}\boldsymbol{f}^{0\prime}\boldsymbol{\epsilon}'\boldsymbol{\Lambda}^0 \sum_{s=1}^{q} \boldsymbol{B}_s^0+\left(\widehat{\boldsymbol{B}}-\boldsymbol{f}^0\boldsymbol{H}'\right)' \boldsymbol{\epsilon}'\boldsymbol{\Lambda}^0 \sum_{s=1}^{q} \boldsymbol{B}_s^0 \equiv \mathbb{C}_{5q1}+\mathbb{C}_{5q2}, \\
\mathbb{C}_{6q} &= \boldsymbol{H}\boldsymbol{f}^{0\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime} \sum_{s=1}^{q} \boldsymbol{\epsilon}_s+\left(\widehat{\boldsymbol{B}}-\boldsymbol{f}^0\boldsymbol{H}'\right)' \boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime} \sum_{s=1}^{q} \boldsymbol{\epsilon}_s \equiv \mathbb{C}_{6q1}+\mathbb{C}_{6q2}. \qquad \text{(B.2.22)}
\end{aligned}
$$

We bound $\max_{1\leq q\leq T}\|\mathbb{C}_{lq}\|$ for $l=1,...,6$ below. The arguments are similar to those used in the proof of Lemma 2 in Bai and Ng (2004). But the treatment of $\boldsymbol{H}$ and $\boldsymbol{f}^0$ are quite different because of the presence of mildly explosive factors here.

We first bound $\max_{1\leq q\leq T}\|C_{1q}\|$. Note that

$$\begin{aligned}\mathbb{C}_{1q1} &= \boldsymbol{H}\boldsymbol{f}^{0\prime}E\left(\boldsymbol{\epsilon}'\sum_{s=1}^{q}\boldsymbol{\epsilon}_s\right)+\boldsymbol{H}\boldsymbol{f}^{0\prime}\left[\boldsymbol{\epsilon}'\sum_{s=1}^{q}\boldsymbol{\epsilon}_s-E\left(\boldsymbol{\epsilon}'\sum_{s=1}^{q}\boldsymbol{\epsilon}_s\right)\right]\\ &= \mathbb{C}_{1q1a}+\mathbb{C}_{1q1b}.\end{aligned}$$

Note that $\mathbb{C}_{1q1a}=\boldsymbol{H}\left(\boldsymbol{\rho}_0^0\right)^T\mathcal{J}_B^{-1}\bar{\mathbb{C}}_{1q1a}$ where $\bar{\mathbb{C}}_{1q1a}=\mathcal{J}_B\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}^{0\prime}E\left(\boldsymbol{\epsilon}'\sum_{s=1}^{q}\boldsymbol{\epsilon}_s\right)$. Note that

$$\begin{aligned}\left\|(N\mathcal{J}_B\kappa_T^{1/2})^{-1}\bar{\mathbb{C}}_{1q1a}\right\| &= \left\|\sum_{t=2}^{T}\kappa_T^{-1/2}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_t^0\sum_{s=1}^{q}\left[N^{-1}\sum_{i=1}^{N}E(\epsilon_{it}\epsilon_{is})\right]\right\|\\ &\leq \left\|\sum_{t=2}^{T}\kappa_T^{-1/2}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_t^0\right\|\sum_{s=1}^{T}\left\|N^{-1}\sum_{i=1}^{N}E(\epsilon_{it}\epsilon_{is})\right\|\\ &\leq M\left\|\sum_{t=2}^{T}\kappa_T^{-1/2}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_t^0\right\|=O_p(1)\ \text{ uniformly in } q,\end{aligned}$$

where the second inequality follows from the fact that $\max_{1\leq s\leq T}\sum_{t=1}^{T}\left|N^{-1}\sum_{i=1}^{N}E(\epsilon_{it}\epsilon_{is})\right|\leq M$ by Assumption A2.3 and last inequality follows from similar arguments as used in the proof of Lemma B.3.2(b). This result, along with the fact that $\left\|\boldsymbol{H}\left(\boldsymbol{\rho}_0^0\right)^T\mathcal{J}_B^{-1}\right\|=O_p(1)$, implies that

$$\max_{1\leq q\leq T}\|\mathbb{C}_{1q1a}\|\leq\left\|\boldsymbol{H}\left(\boldsymbol{\rho}_0^0\right)^T\mathcal{J}_B^{-1}\right\|\max_{1\leq q\leq T}\|\bar{\mathbb{C}}_{1q1a}\|=O_p\left(N\mathcal{J}_B\kappa_T^{1/2}\right). \tag{B.2.23}$$

Let $\Phi_{q,s}=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{q}[\epsilon_{is}\epsilon_{it}-E\left(\epsilon_{is}\epsilon_{it}\right)]$. Then

$$\mathbb{C}_{1q1b}=\sqrt{NT}\boldsymbol{H}\left(\boldsymbol{\rho}_0^0\right)^T\sum_{s=1}^{T}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_s^0\Phi_{q,s}.$$

Following Proofs of Lemma B.1.2(b) in Bai and Ng (2004), it suffices to show that for every

$s = 1, \ldots, T \max_{1 \leq q \leq T} |\Phi_{q,s}| = O_p(1)$. Equivalently, we are going to prove that

$$P\left(\max_{1 \leq q \leq T} |\Phi_{q,s}| \geq K\right) = o(1)$$

where K is finite and chosen to be large enough. Let $\varphi_{it}^s = \epsilon_{is}\epsilon_{it} - E(\epsilon_{is}\epsilon_{it})$ and $\vartheta_{Nq} = N^{2/(4+\delta)} q^{(2+4m)/(4+\delta)}$, where $m > 0$ and can be small enough, δ is defined in Assumption A2.1. Let $\mathbf{1}_{it} = \mathbf{1}\{|\varphi_{it}^s| \leq \vartheta_{Nq}\}$, and $\bar{\mathbf{1}}_{it} = 1 - \mathbf{1}_{it}$. Define

$$\varphi_{1it}^s = \varphi_{it}^s \mathbf{1}_{it} - E(\varphi_{it}^s \mathbf{1}_{it}), \varphi_{2it}^s = \varphi_{it}^s \bar{\mathbf{1}}_{it}, \text{ and } \varphi_{3it}^s = E\left(\varphi_{it}^s \bar{\mathbf{1}}_{it}\right).$$

Apparently $\varphi_{1it}^s + \varphi_{2it}^s - \varphi_{3it}^s = \varphi_{it}^s$ as $\mathbb{E}(\varphi_{it}^s) = 0$. We prove the claim by showing that
(i1) $P\left(\max_{1 \leq q \leq T} \left|\frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^q \varphi_{1it}^s\right| \geq K\right) = o(1)$
(i2) $P\left(\max_{1 \leq q \leq T} \left|\frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^q \varphi_{2it}^s\right| \geq K\right) = o(1)$, and (i3) $\max_{1 \leq q \leq T} \left|\frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^q \varphi_{3it}^s\right| = o(1)$.

First, we prove (i3). Note that

$$\begin{aligned}
\max_{1 \leq q \leq T} \left|\frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^q \varphi_{3it}^s\right| &\leq \frac{1}{\sqrt{NT}} \max_{1 \leq q \leq T} \left\{\sum_{i=1}^N \sum_{t=1}^q E|\varphi_{it}^s|^2\right\}^{1/2} \left\{\sum_{i=1}^N \sum_{t=1}^q P(|\varphi_{it}^s| > \vartheta_{Nq})\right\}^{1/2} \\
&\leq \frac{1}{\sqrt{NT}} \max_{1 \leq q \leq T} (\sqrt{Nq}) \left\{\sum_{i=1}^N \sum_{t=1}^q P(|\varphi_{it}^s| > \vartheta_{Nq})\right\}^{1/2} \\
&\leq \frac{1}{\sqrt{NT}} \max_{1 \leq q \leq T} (\sqrt{Nq}) \left\{\sum_{i=1}^N \sum_{t=1}^q \vartheta_{Nq}^{-(4+\delta)/2} E|\varphi_{it}^s|^{(4+\delta)/2}\right\}^{1/2} \\
&\leq \frac{1}{\sqrt{NT}} \max_{1 \leq q \leq T} Nq\vartheta_{Nq}^{-(4+\delta)/4} = O\left(T^{-m}\right) = o(1).
\end{aligned}$$

where the first inequality holds due to Holder's inequality, the second and fourth inequalities hold because $E|\varphi_{it}^s|^{(4+\delta)/2} < M$ by the construction of φ_{it}^s under Assumption A2.1. The third inequality holds because of Markov inequality.

Next, we prove (i2). Noting that $\max_{1\leq q\leq T}\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{q}\varphi_{2it}^{s}\right|\geq K$ implies that

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left|\varphi_{it}^{s}\right|\bar{\mathbf{1}}_{it}\geq K,$$

using Holder and Markov inequalities, we have,

$$\begin{aligned}
& P\left(\max_{1\leq q\leq T}\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{q}\varphi_{2it}^{s}\right|\geq K\right) \\
\leq\;& P\left[\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left|\varphi_{it}^{s}\right|\bar{\mathbf{1}}_{it}\geq K\right] \\
\leq\;& \frac{(NT)^{-1/2}\left\{\sum_{i=1}^{N}\sum_{t=1}^{T}E|\varphi_{it}^{s}|^{2}\right\}^{1/2}\left\{\sum_{i=1}^{N}\sum_{t=1}^{T}P\left(|\varphi_{it}^{s}|>\vartheta_{Nq}\right)\right\}^{1/2}}{K} \\
\leq\;& (NT)^{-1/2}(NT)^{1/2}\left\{\sum_{i=1}^{N}\sum_{t=1}^{T}\vartheta_{Nq}^{-(4+\delta)/2}E|\varphi_{it}^{s}|^{(4+\delta)/2}\right\}^{1/2} \\
\leq\;& (NT)^{1/2}\vartheta_{Nq}^{-(4+\delta)/4}=O\left(T^{-m}\right)=o\left(1\right).
\end{aligned}$$

where the third and fourth inequalities hold because $E|\varphi_{it}^{s}|^{(4+\delta)/2}<M$ by the construction of φ_{it}^{s} under Assumption A2.1.

To prove (i1), we consider two typical cases for q, i.e., (i1a) $q\asymp T$, (i1b) q is finite and. We first prove (i1) when $q\asymp T$. Without loss of generality, let $\{a_T\}$ be a sequence of integers such that $0<a_T<T$, $a_T\to\infty$ as $T\to\infty$, and $T-a_T=o(\sqrt{T})$. We have

$$\begin{aligned}
& P\left(\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{q}\varphi_{1it}^{s}\right|\geq K\right) \\
\leq\;& P\left(\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{a_T}\varphi_{1it}^{s}\right|+\max_{1\leq q\leq T}\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=a_T+1}^{q}\varphi_{1it}^{s}\right|\geq K\right) \\
\leq\;& P\left(\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{a_T}\varphi_{1it}^{s}\right|\geq K/2\right)+P\left(\max_{1\leq q\leq T}\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=a_T+1}^{q}\varphi_{1it}^{s}\right|\geq K/2\right).
\end{aligned}$$

Using Markov inequality, we bound the second term in the above display as follows,

$$
\begin{aligned}
& P\left(\max_{1\leq q\leq T}\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=a_T+1}^{q}\varphi_{1it}^{s}\right|\geq K/2\right) \\
\leq\ & P\left(\max_{1\leq q\leq T}\frac{1}{\sqrt{T}}\sum_{t=a_T+1}^{q}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\varphi_{1it}^{s}\right|\geq K/2\right) \\
\leq\ & P\left(\frac{1}{\sqrt{T}}\sum_{t=a_T+1}^{T}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\varphi_{1it}^{s}\right|\geq K/2\right) \\
\leq\ & T^{-1/2}\sum_{t=a_T+1}^{T}E\left|N^{-1/2}\sum_{i=1}^{N}\varphi_{1it}^{s}\right| \\
\leq\ & T^{-1/2}\sum_{t=a_T+1}^{T}E\left|N^{-1/2}\sum_{i=1}^{N}\varphi_{it}^{s}\right| = O\left((T-a_T)T^{-1/2}\right) = o\left(1\right),
\end{aligned}
$$

where the fourth equality holds by constructions of φ_{1it}^{s}, and $E\left|N^{-1/2}\sum_{i=1}^{N}\varphi_{it}^{s}\right| < M$ by Assumption A2.3(c). Now, we are in the position to show $\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{a_T}\varphi_{1it}^{s}\right| = O_p\left(1\right)$, to this end, by Chebyshev's inequality, it suffices to show $E\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{a_T}\varphi_{1it}^{s}\right)^2 = O_p\left(1\right)$. Recall $\varphi_{1it}^{s} = \xi_{it}\mathbf{1}_{it} - E\left(\xi_{it}\mathbf{1}_{it}\right)$, and $\varphi_{it}^{s} = \epsilon_{is}\epsilon_{it} - E\left(\epsilon_{is}\epsilon_{it}\right)$. Therefore, under Assumption A2.1, $\{N^{-1/2}\sum_{i=1}^{N}\varphi_{1it}^{s}\}$ are still mixing sequence with zero mean. We have,

$$
\begin{aligned}
E\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{a_T}\varphi_{1it}^{s}\right)^2 &= \frac{1}{T}\sum_{t=1}^{a_T}\sum_{q=1}^{a_T}E\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\varphi_{1it}^{s}\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\varphi_{1jq}^{s}\right] \\
&\leq \frac{1}{T}\sum_{t=1}^{a_T}\sum_{q=1}^{a_T}\left(\alpha(|t-q|)\right)^{\delta/(4+\delta)}\left(E\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\varphi_{1it}^{s}\right|^{(4+\delta)/2}\right)^{4/(4+\delta)} \\
&\leq \frac{1}{T}\sum_{t=1}^{a_T}\sum_{q=1}^{a_T}\left(\alpha(|t-q|)\right)^{\delta/(4+\delta)}\left(E\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\varphi_{it}^{s}\right|^{(4+\delta)/2}\right)^{4/(4+\delta)} \\
&= O\left(a_T T^{-1}\right) = O\left(1\right),
\end{aligned}
$$

where the second equality holds by constructions of φ_{1it}^{s}, and $E\left|N^{-1/2}\sum_{i=1}^{N}\varphi_{it}^{s}\right|^{(4+\delta)/2} < M$ by Assumption A2.3(c) for some $\delta > 0$ defined Assumption A2.1. Collecting all above proofs for the claims (i1), (i2) and (i3), $\max_{1\leq q\leq T}|\Phi_{q,s}| = O_p\left(1\right)$ holds for every $s = 1,\ldots,T$. Besides, it is the trivial case to obtain $\max_{1\leq q\leq T}|\Phi_{q,s}| = O_p(1)$ when q is fixed and finite by similar arguments for the case $q \asymp T$. For other values of q, corresponding proofs can follow above two typical cases with slight modifications. Then, we have

$$\begin{aligned}
\max_{1\leq q\leq T}\|\mathbb{C}_{1q1b}\| &= \sqrt{NT}\max_{1\leq q\leq T}\left\|\boldsymbol{H}\left(\boldsymbol{\rho}_0^0\right)^T\sum_{s=1}^{T}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_s^0\Phi_{q,s}\right\| \\
&\leq \sqrt{NT}\left\|\boldsymbol{H}\left(\boldsymbol{\rho}_0^0\right)^T\right\|\left\|\sum_{s=1}^{T}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_s^0\right\|\max_{1\leq q\leq T}\|\Phi_{q,s}\| \\
&\leq \sqrt{NT}\mathcal{J}_B\left\|\sum_{s=1}^{T}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{f}_s^0\right\|\left(\max_{1\leq q\leq T}\|\Phi_{q,s}\|\right) \\
&= O_p\left(N^{1/2}\kappa_T^{1/2}T^{1/2}\mathcal{J}_B\right), \qquad \text{(B.2.24)}
\end{aligned}$$

where the first inequality holds by sub-multiplicity of the norm operator, and second inequality holds because $\|\boldsymbol{H}(\boldsymbol{\rho}_0^0)\| = O_p(\mathcal{J}_B)$ as discussed in proofs of Lemma B.1.2(a) and (b), and final quality holds by Lemma B.3.2(b). In sum, we have

$$\max_{1\leq q\leq T}\|\mathbb{C}_{1q1}\| = O_p\left(\mathcal{J}_B\kappa_T^{1/2}NT^{1/4}\right) + O_p\left(N^{1/2}\kappa_T^{1/2}T^{1/2}\mathcal{J}_B\right). \qquad \text{(B.2.25)}$$

It is straightforward to see that $\max_{1\leq q\leq T}\|\mathbb{C}_{1q2}\|$ will be dominated by $\max_{1\leq q\leq T}\|\mathbb{C}_{1q1}\|$ by similar arguments above for $\mathbb{C}_{1q1}$ due to the term $\left\|\widehat{\boldsymbol{B}} - \boldsymbol{f}^0\boldsymbol{H}'\right\|$ in corresponding derivations.

For $\max_{1\leq q\leq T}\|\mathbb{C}_{2q}\|$ to $\max_{1\leq q\leq T}\|\mathbb{C}_{4q}\|$, by similar arguments as done above, we can see that they are not leading terms and negligible compared with $\max_{1\leq q\leq T}\mathbb{C}_{5q}$ and $\max_{1\leq q\leq T}\mathbb{C}_{6q}$. Now, we are in the position to show the order of leading terms involving $\max_{1\leq q\leq T}\mathbb{C}_{5q}$ and $\max_{1\leq q\leq T}\mathbb{C}_{6q}$.

Again, $\max_{1\leq q\leq T}\|\mathbb{C}_{5q2}\|$ in (B.2.22) will be dominated by $\max_{1\leq q\leq T}\|\mathbb{C}_{5q1}\|$ as same arguments above for $\max_{1\leq q\leq T}\|\mathbb{C}_{1q}\|$, it is enough to bound $\mathbb{C}_{5q1}$. Note that

$$\max_{1\leq q\leq T}\|\mathbb{C}_{5q1}\| \leq \left\|\boldsymbol{H}\left(\boldsymbol{\rho}_0^0\right)^T\right\|_{sp}\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\sum_i\sum_r\boldsymbol{f}_r^0\epsilon_{ir}\boldsymbol{\lambda}_i^{0\prime}\right\|\max_{1\leq q\leq T}\left\|\sum_{s=1}^{q}\boldsymbol{B}_s^0\right\|.$$

Since the term $\left\|\boldsymbol{H}\left(\boldsymbol{\rho}_0^0\right)^T\right\|_{sp}$ does not involve q and is of order $O_p\left(\mathcal{J}_B\right)$ in Frobenius norm in proofs of Lemma B.1.2, and the term $\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\sum_i\sum_r \boldsymbol{f}_r^0\epsilon_{ir}'\boldsymbol{\lambda}_i^{0\prime}\right\|$ does not involve q and is of order $O_p\left(N^{p/2}\right)$ in Frobenius norm according to proofs of Lemma B.1.2. By the construction of mildly explosive factors, under Assumption A2.1(d) , $\|\sum_{s=1}^q \boldsymbol{B}_s^0\| \leq \sum_{s=1}^q \|\boldsymbol{B}_s^0\| \leq \sum_{s=1}^T \|\boldsymbol{B}_s^0\| = O_p\left(\left(\rho_{01}^0\right)^T+\ldots+\left(\rho_{0R_0}^0\right)^T\right)$ by direct calculations. Then, based on these facts, it follows that

$$\max_{1\leq q\leq T}\|\mathbb{C}_{5q}\| = O_p\left(\mathcal{J}_B N^{p/2}\left[\left(\rho_{01}^0\right)^T+\ldots+\left(\rho_{0R_0}^0\right)^T\right]\right). \tag{B.2.26}$$

Similarly, for $\mathbb{C}_{6q}$, it suffices to bound the dominant term $\|\mathbb{C}_{6q1}\| = \|\boldsymbol{H}\boldsymbol{f}^{0\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}'\sum_{s=1}^q \boldsymbol{\epsilon}_s\|$. Note that

$$\max_{1\leq q\leq T}\|\mathbb{C}_{6q1}\| \leq \left\|\boldsymbol{H}\left(\boldsymbol{\rho}_0^0\right)^T\right\|_{sp}\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\sum_r \boldsymbol{f}_r^0\boldsymbol{B}_r^{0\prime}\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|\left\|\left(\boldsymbol{\rho}_0^0\right)^T\right\|\max_{1\leq q\leq T}\left\|\sum_i\sum_{s=1}^q \boldsymbol{\lambda}_i^0\epsilon_{is}\right\|.$$

where $\left\|\boldsymbol{H}'\left(\boldsymbol{\rho}_0^0\right)^T\right\|_{sp} = O_p\left(\mathcal{J}_B\right)$ by the proofs of Lemma B.1.2, $\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\sum_r \boldsymbol{f}_r^0\boldsymbol{B}_r^{0\prime}\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\| = O_p(1)$ by Lemma B.3.2(e), and $\left\|\left(\boldsymbol{\rho}_0^0\right)^T\right\| = O_p\left(\left(\rho_{01}^0\right)^T+\ldots+\left(\rho_{0R_0}^0\right)^T\right)$ by direct calculations. These terms does not involve q. Now, we are in the position to bound $\max_{1\leq q\leq T}\|\sum_i\sum_{s=1}^q \boldsymbol{\lambda}_i^0\epsilon_{is}\|$, to this end, we can readily obtain that, uniformly in q, $E\left\|N^{-p/2}T^{-1/2}\sum_i\sum_{s=1}^q \boldsymbol{\lambda}_i^0\epsilon_{is}\right\| \leq M$ holds as lemma B.1(4) stated in Bai and Ng (2004). Then, we have $\max_{1\leq q\leq T}\|\sum_i\sum_{s=1}^q \boldsymbol{\lambda}_i\epsilon_{is}\| = O_p\left(N^{p/2}T^{1/2}\right)$. We have

$$\max_{1\leq q\leq T}\|\mathbb{C}_{6q}\| = O_p\left(\mathcal{J}_B N^{p/2}T^{1/2}\left[\left(\rho_{01}^0\right)^T+\ldots+\left(\rho_{0R_0}^0\right)^T\right]\right). \tag{B.2.27}$$

Note that $\left\|N^{-p}\mathcal{J}_T^{-1}\right\| = O_p\left(N^{-p}\mathcal{J}_B^{-1}\left[\left(\rho_{01}^0\right)^{-T}+\ldots+\left(\rho_{0R_0}^0\right)^{-T}\right]\right)$, neglecting those dominated terms, we can conclude

$$\begin{aligned}
&\max_{1\le q\le T}\left\|\sum_{s=1}^{q}\left(\boldsymbol{H}^{-1}\widehat{\boldsymbol{B}}_s-\boldsymbol{B}_s^0-\boldsymbol{u}_s\right)\right\| \\
\le\quad & \left\|\boldsymbol{H}_B'^{-1}\right\|\max_{1\le q\le T}\left\|\left\{\left[N^{1-p}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{V}_{Z,N}\left(\boldsymbol{\rho}_0^0\right)^{-T}\right]\left[\left(\boldsymbol{\rho}_0^0\right)^{2T}\mathcal{J}_T^{-1}\right]\right\}'\widehat{\boldsymbol{B}}_t-\boldsymbol{H}_B'\boldsymbol{B}_t^0-\boldsymbol{H}_u'\boldsymbol{u}_t\right\| \\
+\quad & \left\|\boldsymbol{H}_u\boldsymbol{H}_B^{-1}-\boldsymbol{I}\right\|\max_{1\le q\le T}\left\|\sum_{s=1}^{q}\boldsymbol{u}_s\right\| \\
\le\quad & \left\|\boldsymbol{H}_B^{-1}\right\|\left\|N^{-p}\mathcal{J}_T^{-1}\right\|\left(\max_{1\le q\le T}\left\|\mathbb{C}_{5q1}\right\|+\max_{1\le q\le T}\left\|\mathbb{C}_{6q1}\right\|\right) \\
+\quad & \left\|\boldsymbol{H}_u\boldsymbol{H}_B^{-1}-\boldsymbol{I}\right\|\max_{1\le q\le T}\left\|\sum_{s=1}^{q}\boldsymbol{u}_s\right\|=O_p\left(N^{-p/2}+N^{-p/2}T^{1/2}\right)=O_p\left(N^{-p/2}T^{1/2}\right)
\end{aligned}\tag{B.2.28}$$

where the first inequality follows by the construction of $(\boldsymbol{H}^{-1})'$ in (B.2.10), the sub-multiplicity of the norm and the triangular inequality. And the final equality follows by the bounds for $\max_{1\le q\le T}\|\mathbb{C}_{5q1}\|$ and $\max_{1\le q\le T}\|\mathbb{C}_{6q1}\|$ given above, and the bounds of $\left\|\boldsymbol{H}_B^{-1}\right\|$ and $\left\|\boldsymbol{H}_u\boldsymbol{H}_B^{-1}-\boldsymbol{I}\right\|$ in the proofs of Lemma B.1.2(a), and the fact that $\max_{1\le q\le T}\|\sum_{s=1}^{q}\boldsymbol{u}_s\|=O_p\left(T^{3/4}\right)$ by similar arguments used for $\mathbb{C}_{1q1a}$. Then, it follows $\max_{1\le t\le T}\frac{1}{\sqrt{T}}\left\|\sum_{s=2}^{t}\boldsymbol{H}^{-1}\hat{\boldsymbol{B}}_s-\boldsymbol{f}_s\right\|=O_p\left(N^{-p/2}\right)$ directly.

B.2.4 Proofs of Lemma B.1.4

Recall that $\widehat{\boldsymbol{\Lambda}}_i=(\widehat{\boldsymbol{B}}\widehat{\boldsymbol{B}}')^{-1}\widehat{\boldsymbol{B}}'\boldsymbol{Z}_i$ with $\boldsymbol{Z}_i=\boldsymbol{f}^0\boldsymbol{\Lambda}_i^0+\boldsymbol{\epsilon}_i$, where $\boldsymbol{Z}_i=(Z_{i1},\ldots,Z_{iT})'$ and $\boldsymbol{\epsilon}_i=(\epsilon_{i1},\ldots,\epsilon_{iT})'$. Then

$$\begin{aligned}
\left\|\boldsymbol{\Lambda}^0\boldsymbol{f}^{0\prime}-\widehat{\boldsymbol{\Lambda}}^0\widehat{\boldsymbol{B}}'\right\|^2 &= \sum_{i=1}^{N}\left\|\boldsymbol{f}^0\boldsymbol{\Lambda}_i^0-\widehat{\boldsymbol{B}}\widehat{\boldsymbol{\Lambda}}_i\right\|^2=\sum_{i=1}^{N}\left\|P_{\widehat{\boldsymbol{B}}}\boldsymbol{Z}_i-\boldsymbol{f}^0\boldsymbol{\Lambda}_i^0\right\|^2 \\
&= \sum_{i=1}^{N}\left\|P_{\widehat{\boldsymbol{B}}}\left(\boldsymbol{f}^0\boldsymbol{\Lambda}_i^0+\boldsymbol{\epsilon}_i\right)-\boldsymbol{f}^0\boldsymbol{\Lambda}_i^0\right\|^2=\sum_{i=1}^{N}\left\|M_{\widehat{\boldsymbol{B}}}\boldsymbol{f}^0\boldsymbol{\Lambda}_i^0-P_{\widehat{\boldsymbol{B}}}\boldsymbol{\epsilon}_i\right\|^2 \\
&\le 2\sum_{i=1}^{N}\left\|M_{\widehat{\boldsymbol{B}}}\boldsymbol{f}^0\boldsymbol{\Lambda}_i^0\right\|^2+2\sum_{i=1}^{N}\left\|P_{\widehat{\boldsymbol{B}}}\boldsymbol{\epsilon}_i\right\|^2\equiv 2\mathcal{B}_1+2\mathcal{B}_2.
\end{aligned}$$

It suffices to bound $\mathcal{B}_1$ and $\mathcal{B}_2$. For $\mathcal{B}_2$, we have $\mathcal{B}_2 \leq \sum_{i=1}^{N} \boldsymbol{\epsilon}_i' P_{\widehat{\boldsymbol{B}}} \boldsymbol{\epsilon}_i = O(N+T)$ by Lemma B.3.1(a). For $\mathcal{B}_1$, we apply Lemma B.1.2(a) and Assumption A2.2 to obtain

$$\begin{aligned}
\mathcal{B}_1 &= \sum_{i=1}^{N} \left\| M_{\widehat{\boldsymbol{B}}} \left(\boldsymbol{f}^0 - \hat{\boldsymbol{B}} \boldsymbol{H}'^{-1} \right) \boldsymbol{\Lambda}_i^0 \right\|^2 \leq \left\| M_{\widehat{\boldsymbol{B}}} \right\|_{sp}^2 \left\| \boldsymbol{f}^0 - \hat{\boldsymbol{B}} \boldsymbol{H}'^{-1} \right\|^2 \sum_{i=1}^{N} \left\| \boldsymbol{\Lambda}_i^0 \right\|^2 \\
&= O(1) \, O_p \left(T N^{-p} \right) O_p(N^p) = O_p(T).
\end{aligned}$$

Consequenly, $||\boldsymbol{\Lambda}^0 \boldsymbol{f}^{0\prime} - \widehat{\boldsymbol{\Lambda}} \widehat{\boldsymbol{B}}'||^2 = O_p(N+T)$.

B.2.5 Proofs of Lemmas B.5 and B.6

These results are obtained directly as a combination of the proofs for Lemma B.1.2 and Lemma B.1.3 presented above and the corresponding arguments as used in Bai and Ng (2004) for the unit root case.

B.3 Some Useful Lemmas

Lemma B.3.1. *Suppose Assumption A2.1 to A2.3 hold. Then*

(a) $\sup_{\boldsymbol{F} \in \mathbb{D}_P} (NT)^{-1} \sum_{i=1}^{N} \boldsymbol{\epsilon}_i' \boldsymbol{P}_F \boldsymbol{\epsilon}_i = O_P\left(N^{-1} + T^{-1}\right)$

(b) $(NT)^{-1} \|\boldsymbol{\epsilon}\boldsymbol{\epsilon}'\| = O_P\left(N^{-1/2} + T^{-1/2}\right)$ *and* $(NT)^{-1} \|\boldsymbol{\epsilon}'\boldsymbol{\epsilon}\| = O_P\left(N^{-1/2} + T^{-1/2}\right)$

where $\mathbb{D}_P = \left\{ \boldsymbol{F} \in \mathbb{R}^{T \times R} \right\}$*, and* (a)*,* (b) *hold under Assumptions A2.3.*

Proof. The derivations are exactly same as those in Peng et al. (2020). ■

Lemma B.3.2. *Suppose Assumption A2.1 and A2.5 hold. Then*

(a) $\left\| \kappa_T^{-1/2} \sum_{t=2}^{T} \left(\boldsymbol{\rho}_0^0\right)^{-t} \boldsymbol{u}_t \right\| = O_p(1)$

(b) $\left\| \sum_{t=2}^{T} \left(\boldsymbol{\rho}_0^0\right)^{-T} \boldsymbol{F}_t^0 \right\| = O_p\left(\kappa_T T^{1/2} \rho_{0*}^{-T}\right) + O_p\left(\kappa_T^{3/2}\right)$

(c) $\left\| \left(\boldsymbol{\rho}_0^0\right)^{-T} \kappa_T^{-1} \sum_{t=2}^{T} \boldsymbol{F}_{t-1}^0 \boldsymbol{u}_t' \right\| = O_p(1)$

(d) $\left\| \kappa_T^{-2} \sum_{t=2}^{T} \left(\boldsymbol{\rho}_0^0\right)^{-T} \boldsymbol{F}_t^0 \boldsymbol{F}_t^{0\prime} \left(\boldsymbol{\rho}_0^0\right)^{-T} \right\| = O_p\left(1\right).$

(e) $\left\| \sum_{t=2}^{T} \left(\boldsymbol{\rho}_0^0\right)^{-T} \boldsymbol{f}_t^0 \boldsymbol{f}_t^{0\prime} \left(\boldsymbol{\rho}_0^0\right)^{-T} \right\| = O_p\left(1\right) + O_p(T\rho_{0*}^{-2T})$

where $\rho_0^* = \max\left\{\rho_{0,1}^0, \ldots, \rho_{0,R_0}^0\right\}$, $\rho_{0*} = \min\left\{\rho_{0,1}^0, \ldots, \rho_{0,R_0}^0\right\}$

Proof. Proofs are similar to those derivations of Lemma B.5 in Horie and Yamamoto (2016) despite that Horie and Yamamoto (2016) only considered univariate case such that $R_0 = 1$. However, when $R_0 > 1$, all routines are similar if we define $c^* = \max\left\{c_1, \ldots, c_{R_0}\right\}$ and then, treat corresponding arguments as if $R_0 = 1$ as done in Horie and Yamamoto (2016). ■

The next lemma studies the asymptotic property of $\boldsymbol{V_{Z,N}}$.

Lemma B.3.3. *Under Assumption A2.1-A2.6, as* $N, T \to +\infty$*,*

$$N^{1-p}\left(\boldsymbol{\rho}_0^0\right)^{-T} \boldsymbol{V}_{Z,N} \left(\boldsymbol{\rho}_0^0\right)^{-T} \to \Upsilon_1 \tag{B.3.1}$$

where $\Upsilon_1 \equiv \lim_{N,T\to\infty} \left(\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}\right)^{-1} \left[\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\right] \frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0}{N^p} \left(\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}\right) \left(\boldsymbol{\rho}_0^0\right)^{-2T}$*, which is a positively definite matrix.*

Proof. Premultiplying $B^{0\prime}$ and postmultiplying $\left(\boldsymbol{\rho}_0^0\right)^{-2T}$ on both sides of (B.2.2), we have

$$\begin{aligned}
& \boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\boldsymbol{V}_{Z,N}\mathcal{J}_T^{-1}\left(\boldsymbol{\rho}_0^0\right)^{-2T} - N^{-1+p}\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\frac{1}{N^p}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0(\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1})\left(\boldsymbol{\rho}_0^0\right)^{-2T} \\
= \; & N^{-1}\boldsymbol{B}^{0\prime}\left(\boldsymbol{\Lambda}^0\boldsymbol{u}' + \boldsymbol{\epsilon}\right)'\left(\boldsymbol{\Lambda}^0\boldsymbol{u}' + \boldsymbol{\epsilon}\right)\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}\left(\boldsymbol{\rho}_0^0\right)^{-2T} \\
& + N^{-1}\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\left(\boldsymbol{\Lambda}^0\boldsymbol{u}' + \boldsymbol{\epsilon}\right)\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}\left(\boldsymbol{\rho}_0^0\right)^{-2T} + N^{-1}\boldsymbol{B}^{0\prime}\left(\boldsymbol{\Lambda}^0\boldsymbol{u}' + \boldsymbol{\epsilon}\right)'\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}\left(\boldsymbol{\rho}_0^0\right)^{-2T} \\
\equiv \; & \mathcal{A}_1 + \mathcal{A}_2 + \mathcal{A}_3.
\end{aligned}$$

We bound each term below. First, note that

$$\begin{aligned}
\mathcal{A}_1 &= N^{-1}\left(\boldsymbol{B}^{0\prime}\boldsymbol{u}\right)\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^{0}\boldsymbol{u}^{\prime}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}\left(\boldsymbol{\rho}_0^0\right)^{-2T}+N^{-1}\left(\boldsymbol{B}^{0\prime}\boldsymbol{u}\right)\left(\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}\right)\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}\left(\boldsymbol{\rho}_0^0\right)^{-2T}\\
&\quad+N^{-1}\left(\boldsymbol{\epsilon}\boldsymbol{B}^{0}\right)^{\prime}\left(\boldsymbol{\Lambda}^{0}\boldsymbol{u}^{\prime}\right)\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}\left(\boldsymbol{\rho}_0^0\right)^{-2T}+N^{-1}\left(\boldsymbol{\epsilon}\boldsymbol{B}^{0}\right)^{\prime}\boldsymbol{\epsilon}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}\left(\boldsymbol{\rho}_0^0\right)^{-2T}\\
&\equiv \mathcal{A}_{11}+\mathcal{A}_{12}+\mathcal{A}_{13}+\mathcal{A}_{14}.
\end{aligned}$$

For $\mathcal{A}_{11}$, we have

$$\begin{aligned}
\|\mathcal{A}_{11}\| &= N^{-1+p}T^{1/2}\left\|\left(\boldsymbol{B}^{0\prime}\boldsymbol{u}\left(\boldsymbol{\rho}_0^0\right)^{-T}\right)\left(\boldsymbol{\rho}_0^0\right)^{T}\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^{0}}{N^p}\frac{\boldsymbol{u}^{\prime}}{T^{1/2}}\mathcal{J}_B^{-1}\widehat{\boldsymbol{B}}\mathcal{J}_B\mathcal{J}_T^{-1}\left(\boldsymbol{\rho}_0^0\right)^{-2T}\right\|\\
&\leq N^{-1+p}T^{1/2}\left\|\boldsymbol{B}^{0\prime}\boldsymbol{u}\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|\left\|\left(\boldsymbol{\rho}_0^0\right)^{T}\right\|_{sp}\left\|\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^{0}}{N^p}\right\|\left\|\frac{\boldsymbol{u}^{\prime}}{T^{1/2}}\right\|\left\|\mathcal{J}_B^{-1}\widehat{\boldsymbol{B}}\right\|\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|_{sp}^{3}\\
&= N^{-1+p}T^{1/2}O_p\left(\kappa_T\right)\left\|\left(\boldsymbol{\rho}_0^0\right)^{T}\right\|_{sp}\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|_{sp}^{3}\\
&= N^{-1+p}T^{1/2}\kappa_T O_p\left(\exp\left(-2c_{r,min}T\kappa_T^{-1}\right)\right),
\end{aligned}$$

where we use the fact that $\mathcal{J}_T=\left(\boldsymbol{\rho}_0^0\right)^{T}\mathcal{J}_B$ to obtain the inequality, the second equality holds by Assumptions A2.1(b) and A2.2, Lemma B.3.2(c), and the normalization condition, and the last equality holds by the fact that as $T\to\infty$, $\left\|\left(\boldsymbol{\rho}_0^0\right)^{T}\right\|_{sp}=\left(1+\frac{c_{r,max}}{\kappa_T}\right)^{T}\asymp\exp\left(c_{r,max}T\kappa_T^{-1}\right)$, $\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|_{sp}=\left(1+\frac{c_{r,min}}{\kappa_T}\right)^{-T}\asymp\exp\left(-c_{r,min}T\kappa_T^{-1}\right)$, and $\left|c_{r,max}-c_{r,min}\right|=O_p\left(\frac{\kappa_T}{T}\right)$ as assumed in A2.5(a). Similarly,

$$\begin{aligned}
&\|\mathcal{A}_{12}\|\\
&\leq N^{-1+p/2}T^{1/2}\left\|\boldsymbol{B}^{0\prime}\boldsymbol{u}\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|\left\|\left(\boldsymbol{\rho}_0^0\right)^{T}\right\|_{sp}\left\|\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}}{N^{p/2}T^{1/2}}\right\|\left\|\mathcal{J}_B^{-1}\widehat{\boldsymbol{B}}\right\|\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|_{sp}^{3}\\
&= N^{-1+p/2}T^{1/2}O_p\left(\kappa_T\right)O_p\left(\exp\left(-2c_{r,min}T\kappa_T^{-1}\right)\right),
\end{aligned}$$

where we use Lemma B.3.2(c), (B.2.4), and the normalization condition. Next,

$$\begin{aligned}\|\mathcal{A}_{13}\| &\leq N^{-1+p/2}T^{1/2}\left\|\boldsymbol{B}^{0\prime}\boldsymbol{\epsilon}'\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|\left\|\left(\boldsymbol{\rho}_0^0\right)^{T}\right\|_{sp}\left\|\frac{\boldsymbol{\Lambda}^0}{N^{p/2}}\right\|\left\|\frac{\boldsymbol{u}'}{\sqrt{T}}\right\|\left\|\mathcal{J}_B^{-1}\widehat{\boldsymbol{B}}\right\|\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|_{sp}^3 \\ &= N^{-1+p/2}T^{1/2}O_p\left(N^{1/2}\right)\left\|\left(\boldsymbol{\rho}_0^0\right)^{T}\right\|_{sp}\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|_{sp}^3 \\ &= N^{-1/2+p/2}T^{1/2}O_p\left(\exp\left(-2c_{r,min}T\kappa_T^{-1}\right)\right),\end{aligned}$$

where we use Assumptions A2.1(b) and A2.2, the normalization condition, and the fact that $\left\|\boldsymbol{B}^{0\prime}\boldsymbol{\epsilon}'\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\| = O_p\left(N^{1/2}\right)$ by Assumptions A2.1 and A2.4 and similar arguments as used to obtain Lemma B.3.2(c) (c.f., Lemma B.5(c) in Horie and Yamamoto (2016)). Similarly, we have

$$\begin{aligned}\|\mathcal{A}_{14}\| &\leq N^{-1}\left\|\boldsymbol{B}^0\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|\left\|\left(\boldsymbol{\rho}_0^0\right)^{T}\right\|_{sp}\|\boldsymbol{\epsilon}'\boldsymbol{\epsilon}\|_{sp}\left\|\mathcal{J}_B^{-1}\widehat{\boldsymbol{B}}\right\|\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|_{sp}^3 \\ &= N^{-1}O_p\left(N+T\right)O_p\left(\exp\left(-2c_{r,min}T\kappa_T^{-1}\right)\right),\end{aligned}$$

where we use Lemma B.3.2, Assumption A2.3, and the normalization condition. It follows that

$$\|\mathcal{A}_1\| = O_p\left(N^{-1+p}\left[T^{1/2}\kappa_T + N^{1/2-p/2}T^{1/2} + N^{-p}\left(N+T\right)\right]\right)O_p\left(\exp\left(-2c_{r,min}T\kappa_T^{-1}\right)\right). \tag{B.3.2}$$

Next, we study $\mathcal{A}_2$. Note that

$$\mathcal{A}_2 = N^{-1}\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\boldsymbol{u}'\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}\left(\boldsymbol{\rho}_0^0\right)^{-2T} + N^{-1}\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}\left(\boldsymbol{\rho}_0^0\right)^{-2T} \equiv \mathcal{A}_{21} + \mathcal{A}_{22}.$$

For $\mathcal{A}_{21}$, we have

$$
\begin{aligned}
\|\mathcal{A}_{21}\| &\leq N^{-1+p}T^{1/2}\left\|\left(\boldsymbol{\rho}_0^0\right)^T\right\|_{sp}\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|\left\|\left(\boldsymbol{\rho}_0^0\right)^T\right\|_{sp} \\
&\quad \left\|\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0}{N^p}\right\|\left\|\frac{\boldsymbol{u}'}{T^{1/2}}\right\|\left\|\mathcal{J}_B^{-1}\widehat{\boldsymbol{B}}\right\|\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|_{sp}^3 \\
&= N^{-1+p}T^{1/2}O_p(1)\left\|\left(\boldsymbol{\rho}_0^0\right)^T\right\|_{sp}^2\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|_{sp}^3 = N^{-1+p}T^{1/2}O_p\left(\exp\left(-c_{r,min}T\kappa_T^{-1}\right)\right),
\end{aligned}
$$

where the first equality holds by Assumptions A2.1(b) and A2.2, Lemma B.3.2(e), the normalization condition, the fact that $\mathcal{J}_T = \left(\boldsymbol{\rho}_0^0\right)^T\mathcal{J}_B$. Similarly,

$$
\begin{aligned}
\|\mathcal{A}_{22}\| &\leq N^{-1+p/2}T^{1/2}\left\|\left(\boldsymbol{\rho}_0^0\right)^T\right\|_{sp}\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|\left\|\left(\boldsymbol{\rho}_0^0\right)^T\right\|_{sp} \\
&\quad \left\|\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}}{N^{p/2}T^{1/2}}\right\|\left\|\mathcal{J}_B^{-1}\widehat{\boldsymbol{B}}\right\|\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|_{sp}^3 \\
&= N^{-1+p/2}T^{1/2}O_p(1)\left\|\left(\boldsymbol{\rho}_0^0\right)^T\right\|_{sp}^2\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|_{sp}^3 = N^{-1+\frac{p}{2}}T^{1/2}O_p\left(\exp\left(-c_{r,min}T\kappa_T^{-1}\right)\right),
\end{aligned}
$$

where the first equality holds by Lemma B.3.2(e), (B.2.4), the normalization condition, and the fact that $\mathcal{J}_T = \left(\boldsymbol{\rho}_0^0\right)^T\mathcal{J}_B$. In sum, we have

$$
\|\mathcal{A}_2\| = N^{-1+p}T^{1/2}O_p\left(\exp\left(-c_{r,min}T\kappa_T^{-1}\right)\right). \tag{B.3.3}
$$

Now, we study $\mathcal{A}_3$. Note that

$$
\mathcal{A}_3 = N^{-1}\boldsymbol{B}^{0\prime}\boldsymbol{u}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}\left(\boldsymbol{\rho}_0^0\right)^{-2T} + N^{-1}\boldsymbol{B}^{0\prime}\boldsymbol{\epsilon}'\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}\left(\boldsymbol{\rho}_0^0\right)^{-2T} \equiv \mathcal{A}_{31} + \mathcal{A}_{32}.
$$

For $\mathcal{A}_{31}$, we have

$$
\begin{aligned}
&\|\mathcal{A}_{31}\| \\
\leq\; & N^{-1+p}\left\|\left(\boldsymbol{\rho}_0^0\right)^T\right\|_{sp}\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{B}^{0\prime}\boldsymbol{u}\right\|\left\|\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0}{N^p}\right\|\left\|\left(\boldsymbol{\rho}_0^0\right)^T\right\|_{sp}\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{B}^{0\prime}\right\|\left\|\mathcal{J}_B^{-1}\widehat{\boldsymbol{B}}\right\|\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|_{sp}^3 \\
\leq\; & N^{-1+p}O_p(\kappa_T)\left\|\left(\boldsymbol{\rho}_0^0\right)^T\right\|_{sp}^2\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|_{sp}^3 = N^{-1+p}\kappa_T O_p\left(\exp\left(-c_{r,min}T\kappa_T^{-1}\right)\right),
\end{aligned}
$$

where the first equality follows by Assumption A2.2, Lemma B.3.2(c) and (e), the normalization condition, and the fact that $\mathcal{J}_T = \left(\boldsymbol{\rho}_0^0\right)^T \mathcal{J}_B$. Similarly,

$$\begin{aligned}
& \left\|\mathcal{A}_{32}\right\| \\
\leq\; & N^{-1+p/2}\left\|\left(\boldsymbol{\rho}_0^0\right)^T\right\|_{sp}\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{\epsilon}\boldsymbol{B}^0\right\|\left\|\frac{\boldsymbol{\Lambda}^0}{N^{p/2}}\right\|\left\|\left(\boldsymbol{\rho}_0^0\right)^T\right\|_{sp}\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{B}^{0\prime}\right\|\left\|\mathcal{J}_B^{-1}\widehat{\boldsymbol{B}}\right\|\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|_{sp}^3 \\
\leq\; & N^{-1+p}O_p\left(N^{1/2}\right)\left\|\left(\boldsymbol{\rho}_0^0\right)^T\right\|_{sp}^2\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\right\|_{sp}^3 = N^{-1/2+p}O_p\left(\exp\left(-c_{r,min}T\kappa_T^{-1}\right)\right),
\end{aligned}$$

where the equality holds by Lemma B.3.2(e), Assumption A2.2, the normalization condition, the fact that $\mathcal{J}_T = \left(\boldsymbol{\rho}_0^0\right)^T \mathcal{J}_B$, and the fact that $\left\|\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{\epsilon}\boldsymbol{B}^0\right\| = O_p\left(N^{1/2}\right)$ used above. In sum, we have

$$\left\|\mathcal{A}_3\right\| = N^{-1+p}\left(\kappa_T + N^{1/2-p/2}\right)O_p\left(\exp\left(-c_{r,min}T\kappa_T^{-1}\right)\right). \tag{B.3.4}$$

Combining (B.3.2), (B.3.3), and (B.3.4), we have

$$\begin{aligned}
& \left\|\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\boldsymbol{V}_{Z,N}\mathcal{J}_T^{-1}\left(\boldsymbol{\rho}_0^0\right)^{-2T} - N^{-1+p}\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0}{N^p}\left(\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}\right)\left(\boldsymbol{\rho}_0^0\right)^{-2T}\right\| \\
=\; & \left(N^{-1+p}\left[T^{1/2}\kappa_T + N^{1/2-p/2}T^{1/2} + N^{-p}\left(N+T\right)\right]\right)O_p\left(\exp\left(-2c_{r,min}T\kappa_T^{-1}\right)\right) \\
& + N^{-1+p}(T^{1/2} + \kappa_T + N^{1/2-p/2})O_p\left(\exp\left(-c_{r,min}T\kappa_T^{-1}\right)\right) \\
=\; & o_p\left(1\right).
\end{aligned} \tag{B.3.5}$$

Noting that $\mathcal{J}_T \equiv \left(\boldsymbol{\rho}_0^0\right)^T \mathcal{J}_B$, $\boldsymbol{V}_{Z,N}$, and $\left(\boldsymbol{\rho}_0^0\right)^T$ are all $R_0 \times R_0$ diagonal matrices, so they are interchangeable, (B.3.5) implies that

$$\left(\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1}\right)\left(N^{1-p}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{V}_{Z,N}\left(\boldsymbol{\rho}_0^0\right)^{-T}\right) = \left[\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\right]\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0}{N^p}(\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\mathcal{J}_T^{-1})\left(\boldsymbol{\rho}_0^0\right)^{-2T} + o_p(1)$$

and $N^{1-p}\left(\boldsymbol{\rho}_0^0\right)^{-T}\boldsymbol{V}_{Z,N}\left(\boldsymbol{\rho}_0^0\right)^{-T} \xrightarrow{p} \Upsilon_1$. ■

B.4 Supplementary Results for Monte Carlo Simulations

In this section, we first report the simulation results based on the asymptotic normal critical value. Then we report the performance of the transformed double ridge ratio criterion to determine the number of factors.

B.4.1 Simulation results based on the asymptotic normal critical value

In this subsection, we present simulation results of our proposed test based on the asymptotic critical value.

Table B.1: Finite sample properties of the PPT under the null and alternatives when $R_0 = 3$ that is taken as known.

	Size of the Proposed Test					
DGP 1	$\rho_i = 1$ for $i = 1, \ldots, N$					
$\boldsymbol{\rho}_0^0 \backslash (N, T)$	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
$diag\,(1.00, 1.00, 1.00)$	0.056	0.052	0.042	0.062	0.038	0.034
$diag\,(1.02, 1.02, 1.02)$	0.020	0.002	0.000	0.016	0.006	0.000
$diag\,(1.04, 1.04, 1.04)$	0.002	0.000	0.000	0.006	0.000	0.000
$diag\,(1.08, 1.08, 1.08)$	0.000	0.000	0.002	0.000	0.000	0.000
$diag\,(1.02, 1.04, 1.06)$	0.000	0.002	0.000	0.000	0.000	0.000
$diag\,(1.02, 1.05, 1.08)$	0.000	0.000	0.000	0.000	0.000	0.000
	Power of the Proposed Test					
DGP 2	$\rho_i \overset{i.i.d}{\sim} Uniform\,(1.08, 1.10)$ for any $i \in \{1, \ldots, N\}$					
$\boldsymbol{\rho}_0^0 \backslash (N, T)$	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
$diag\,(1.00, 1.00, 1.00)$	0.180	0.910	1.000	0.382	0.998	1.000
$diag\,(1.02, 1.02, 1.02)$	0.118	0.604	1.000	0.162	0.606	1.000
$diag\,(1.04, 1.04, 1.04)$	0.156	0.754	1.000	0.210	0.758	1.000
$diag\,(1.08, 1.08, 1.08)$	0.378	0.918	1.000	0.504	0.928	1.000
$diag\,(1.02, 1.04, 1.06)$	0.174	0.728	1.000	0.214	0.794	1.000
$diag\,(1.02, 1.05, 1.08)$	0.170	0.758	1.000	0.210	0.810	1.000

Table B.1 reports the results of the proposed PPT test based on the asymptotic critical value when $R_0 = 3$ is assumed to be known. From the table, we can see the proposed test suffers from severe downward size distortions in the presence of mildly explosive factors. When the factors follow a unit root process, it is consistent with our anticipation that our proposed test can have nominal size based on the asymptotic critical value. However, the power of the proposed test is reasonably good in many cases, which is attributable to the exponential explosiveness in the idiosyncratic components.

Table B.2: Finite sample properties of the PPT under the null and alternatives when $R_0 = 3$ that is taken as unknown and estimated by Algorithm 2.2.

	Size of the Proposed Test					
DGP 1	$\rho_i = 1$ for $i = 1, \ldots, N$					
$\boldsymbol{\rho}_0^0 \backslash (N, T)$	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
$diag\,(1.00, 1.00, 1.00)$	0.042	0.040	0.038	0.048	0.064	0.052
$diag\,(1.02, 1.02, 1.02)$	0.022	0.002	0.000	0.020	0.004	0.002
$diag\,(1.04, 1.04, 1.04)$	0.006	0.002	0.000	0.008	0.000	0.000
$diag\,(1.08, 1.08, 1.08)$	0.000	0.000	0.000	0.000	0.000	0.000
$diag\,(1.02, 1.04, 1.06)$	0.000	0.000	0.006	0.006	0.002	0.006
$diag\,(1.02, 1.05, 1.08)$	0.002	0.000	0.012	0.000	0.000	0.002
	Power of the Proposed Test					
DGP 2	$\rho_i \overset{i.i.d}{\sim}$ *Uniform* $(1.08, 1.10)$ for any $i \in \{1, \ldots, N\}$					
$\boldsymbol{\rho}_0^0 \backslash (N, T)$	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
$diag\,(1.00, 1.00, 1.00)$	0.288	0.478	0.940	0.544	0.506	0.990
$diag\,(1.02, 1.02, 1.02)$	0.038	0.602	0.998	0.110	0.676	0.996
$diag\,(1.04, 1.04, 1.04)$	0.072	0.544	0.474	0.168	0.598	0.548
$diag\,(1.08, 1.08, 1.08)$	0.046	0.268	1.000	0.054	0.366	0.992
$diag\,(1.02, 1.04, 1.06)$	0.112	0.440	0.754	0.116	0.542	0.776
$diag\,(1.02, 1.05, 1.08)$	0.040	0.472	0.572	0.088	0.430	0.624

Table B.2 shows the results of the proposed PPT test based on the asymptotic critical value when R_0 is estimated by Algorithm 2.2. As above, we can also see the significant downward size distortions in the presence of mildly explosive factors except for factors. The power of the proposed test remains good in many cases.

When $R_0 = 1$, Tables B.3 and B.4 display similar results as those in Tables B.1 and B.2.

In summary, when the asymptotic normal critical value is used, the proposed PPT test is undersized in the presence of mildly explosive factors. These results highlight the need of a wild-bootstrap-based test as shown in the main text.

Table B.3: Finite sample properties of the PPT under the null and alternatives when $R_0 = 1$ that is taken as known.

	Size of the Proposed Test					
DGP 1	$\rho_i = 1$ for $i = 1, \dots, N$					
$\rho_0^0 \backslash (N,T)$	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
$\rho = 1.00$	0.050	0.038	0.042	0.050	0.052	0.040
$\rho = 1.02$	0.022	0.014	0.010	0.036	0.020	0.002
$\rho = 1.04$	0.012	0.006	0.002	0.020	0.006	0.056
$\rho = 1.08$	0.008	0.000	0.000	0.004	0.000	0.000
	Power of the Proposed Test					
DGP 2	$\rho_i \overset{i.i.d}{\sim}$ *Uniform* $(1.08, 1.10)$ for any $i \in \{1, \dots, N\}$					
$\rho_0^0 \backslash (N,T)$	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
$\rho = 1.00$	0.120	1.000	1.000	0.172	1.000	1.000
$\rho = 1.02$	0.904	1.000	1.000	0.982	0.998	1.000
$\rho = 1.04$	0.898	0.952	1.000	0.934	0.950	1.000
$\rho = 1.08$	0.780	0.998	1.000	0.952	1.000	1.000

Table B.4: Finite sample properties of the PPT under the null and alternatives when $R_0 = 1$ that is taken as unknown and estimated by Algorithm 2.2.

	Size of the Proposed Test					
DGP 1	$\rho_i = 1$ for $i = 1, \dots, N$					
$\rho_0^0 \backslash (N,T)$	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
$\rho = 1.00$	0.040	0.056	0.054	0.048	0.042	0.044
$\rho = 1.02$	0.042	0.056	0.048	0.040	0.046	0.044
$\rho = 1.04$	0.042	0.048	0.040	0.044	0.044	0.056
$\rho = 1.08$	0.040	0.044	0.048	0.054	0.044	0.044
	Power of the Proposed Test					
DGP 2	$\rho_i \overset{i.i.d}{\sim}$ *Uniform* $(1.08, 1.10)$ for any $i \in \{1, \dots, N\}$					
$\rho_0^0 \backslash (N,T)$	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
$\rho = 1.00$	1.000	1.000	1.000	1.000	1.000	1.000
$\rho = 1.02$	0.904	1.000	1.000	0.982	0.998	1.000
$\rho = 1.04$	0.898	0.952	1.000	0.934	0.950	1.000
$\rho = 1.08$	0.780	0.998	1.000	0.952	1.000	1.000

B.4.2 The selection frequency of transformed double ridge ratio

In this subsection, we report the selection frequency for the transformed double ridge ratio criterion when R_0 is estimated by Algorithm 2.2 described in Section 2.4. In Table B.5 we present the correct selection frequency ($\hat{R} = R_0$) under the null; In Table B.6, the selection frequency of $\hat{R} \geq R_0$ is of interest and reported.

Table B.5: selection frequency of TDDR criterion under the null when $R_0 = 3$ that is estimated by Algorithm 2.2.

	Correct Selection Frequency with $\hat{R} = R_0$ (%)					
DGP 1	$\rho_i = 1$ for $i = 1, \ldots, N$					
$\boldsymbol{\rho}_0^0 \backslash (N, T)$	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
$diag\,(1.00, 1.00, 1.00)$	60.4	38.4	27.0	57.2	28.8	10.2
$diag\,(1.02, 1.02, 1.02)$	58.2	47.6	73.4	54.6	43.6	67.4
$diag\,(1.04, 1.04, 1.04)$	67.8	78.0	77.2	65.8	76.2	77.4
$diag\,(1.08, 1.08, 1.08)$	84.2	80.6	80.6	85.2	81.0	75.4
$diag\,(1.02, 1.04, 1.06)$	70.8	80.8	96.2	67.6	81.0	96.6
$diag\,(1.02, 1.05, 1.08)$	79.2	88.6	96.8	79.4	92.0	99.0

Table B.5 exhibits simulation results of correct selection frequency of Algorithm 2.2. Except for the case that factors follow unit-root processes, Algorithm 2.2 chooses R_0 with high probabilities in most cases for different specifications of mildly explosive factors.

Table B.6: selection fequency of TDDR criterion under the null when $R_0 = 3$ that is estimated by Algorithm 2.2.

	Selection Frequency of $\hat{R} \geq R_0$(%)					
DGP 2	$\rho_i \overset{i.i.d}{\sim} \textit{Uniform}\,(1.08, 1.10)$ for any $i \in \{1, \ldots, N\}$					
$\boldsymbol{\rho}_0^0 \backslash (N, T)$	(100,51)	(100,101)	(100,201)	(200,51)	(200,101)	(200,201)
$diag\,(1.02, 1.04, 1.06)$	70.8	80.8	96.2	67.6	81.0	96.6
$diag\,(1.02, 1.05, 1.08)$	79.2	88.6	96.8	79.4	92.0	99.0
$diag\,(1.00, 1.00, 1.00)$	61.4	41.6	100.0	50.6	36.2	100.0
$diag\,(1.02, 1.02, 1.02)$	70.4	40.8	100.0	59.4	29.6	100.0
$diag\,(1.04, 1.04, 1.04)$	71.6	48.2	100.0	56.0	41.0	100.0
$diag\,(1.08, 1.08, 1.08)$	89.2	48.2	100.0	89.6	54.2	100.0

However, the results in Table B.6 reveal that Algorithm 2.2 overestimates the true number of factors with high probabilities if data are generated under the alternatives. This issue results from spurious factors under the alternatives. However, as analyzed in the main texts, the slight over-extraction of factors does not lead to the total loss of the power against the alternatives for our proposed test.

Appendix C

Appendix to Chapter 3

C.1 Proofs of Main Results

To prove the main results in the paper, especially Theorem 3.3.2 and Theorem 3.3.4, we need some technical lemmas. Below we first state the technical lemmas whose proofs can be found on the online supplement, and then prove the main results in the paper.

C.1.1 Technical Lemmas

We first state some technical lemmas related to the consistency of the estimated local-to-unity explosive factors under the alternative of local-to-unity. In the case where the factors exhibit a unit root process, the results are relatively simple and similar to those in Westerlund (2015).

Consistency of the estimated local-to-unity explosive factors when idiosyncratic error terms are local-to-unity explosive

The following four lemmas establish the consistency of the PC estimation for the first-differenced form of data, they hold under some mild assumptions when $\rho^0_{0,r} = 1 + \nu_r/T$ with $\nu_r > 0$ being finite and fixed for all r, and $\rho_i^0 = 1 + c_i/T$ with $c_i \geq 0$ being finite and fixed for all i as specified in (3.2 .7).

Lemma C.1.1. *Suppose that Assumptions A3.1–A3.5 hold. Then $T^{-1}\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}$ is asymptotically invertible.*

Lemma C.1.2. *Suppose that Assumptions A3.1–A3.5 hold and the null hypothesis in (3.2 .5) holds. Then there exists $\boldsymbol{H}$ with asymptotic rank R_0 such that as $(N,T)\to\infty$,*

(a) $T^{-1}\sum_{t=2}^{T}\left\|\boldsymbol{H}^{-1}\widehat{\boldsymbol{B}}_t - \boldsymbol{f}_t^0\right\|^2 = O_p(N^{-1})$;

(b) $\left(\boldsymbol{H}^{-1}\widehat{\boldsymbol{B}}_t - \boldsymbol{B}_t^0\right) = O_p(N^{-1/2} + T^{-1/2} + N^{-1}T^{1/2})$ *for each given* t;

(c) $\left(\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{H}^{\prime -1}\boldsymbol{\lambda}_i^0\right) = O_p\left(T^{-1/2} + N^{-1}\right)$ *for each given* i.

Lemma C.1.2 above shows that the PC estimation for the first-dfferenced form of data can yield the consistent estimates of unobserved common factors in first difference and factor loadings, and it is also possible to allow N and T to diverge at different rates as in Bai (2009) and Lu and Su (2016). Lemma C.1.3 is similar to the corresponding result in Bai and Ng (2004), and demonstrates that the cumulative sum of $\widehat{\boldsymbol{B}}_s$ is uniformly close to the cumulative sum of $\boldsymbol{B}_s^0$,

Lemma C.1.3. *Under the assumptions of Lemma C.1.2,*

$$\max_{1\leq t\leq T}\frac{1}{\sqrt{T}}\left\|\sum_{s=2}^{t}\left(\boldsymbol{H}^{-1}\widehat{\boldsymbol{B}}_s - \boldsymbol{B}_s^0\right)\right\| \quad = \quad O_p\left(N^{-1/2}\right).$$

Lemma C.1.4. *Under the assumptions of Lemma C.1.2,* $\left\|\boldsymbol{\Lambda}^0\boldsymbol{f}^{0\prime} - \widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{B}}^{\prime}\right\|^2 = O_p\left(N+T\right)$.

To study the asymptotic properties of the test statistic, we need the following lemma, which is similar to Lemma B.2 in Bai and Ng (2004). Lemma C.1.5 are presented for completeness and can readily obtained by following proofs in Bai and Ng (2004) directly based on Lemma C.1.2-C.1.3, and assumptions in the current paper.

Lemma C.1.5. *Consider estimation of* (3.2 .10) *by the method of principal components. Let* $\boldsymbol{B}_t^0$ *be defined by* (3.2 .9)*, and* $\widehat{\boldsymbol{F}}_t$ *be defined by* (3.2 .12)*. Besides, denote the sample means by* $\overline{\widehat{\boldsymbol{F}}} = (T-1)^{-1}\sum_{t=2}^{T}\widehat{\boldsymbol{F}}_t$*,* $\overline{\boldsymbol{F}} = (T-1)^{-1}\sum_{t=2}^{T}\boldsymbol{F}_t^0$*. Let* $\widehat{\boldsymbol{F}}^c = \widehat{\boldsymbol{F}}_t - \overline{\widehat{\boldsymbol{F}}}$ *be the demeaned series and we define* $\boldsymbol{F}_t^{0,c}$ *similarly. Suppose that Assumptions A3.1-A3.5 hold. Then there exists an* $\boldsymbol{H}$ *with rank* R *such that as* $N, T \to \infty$*,*

(i) $(1/\sqrt{T})\widehat{\boldsymbol{F}}_t = \boldsymbol{H}(1/\sqrt{T})\boldsymbol{F}_t^0 + O_p\left(N^{-1/2} + T^{-1/2} + N^{-1}T^{1/2}\right)$ *uniformly in* $t \in [2, T]$*;*

(ii) $(1/T^2)\sum_{t=2}^{T}\widehat{\boldsymbol{F}}_t\widehat{\boldsymbol{F}}_t' = \boldsymbol{H}\left((1/T^2)\sum_{t=2}^{T}\boldsymbol{F}_t^0\boldsymbol{F}_t^{0\prime}\right)\boldsymbol{H}' + O_p\left(N^{-1/2} + T^{-1/2} + N^{-1}T^{1/2}\right)$

(iii) $(1/T)\sum_{t=2}^{T}\widehat{\boldsymbol{B}}_t\widehat{\boldsymbol{B}}_t' = \boldsymbol{H}\left((1/T)\sum_{t=2}^{T}\boldsymbol{B}_t^0\boldsymbol{B}_t^{0\prime}\right)\boldsymbol{H}' + O_p\left(N^{-1/2} + T^{-1/2} + N^{-1}T^{1/2}\right)$

(iv) $(1/T)\sum_{t=3}^{T}\left(\widehat{\boldsymbol{F}}_{t-1}\widehat{\boldsymbol{B}}_t' + \widehat{\boldsymbol{B}}_t\widehat{\boldsymbol{F}}_{t-1}'\right) = (1/T)\boldsymbol{H}\sum_{t=3}^{T}\left(\boldsymbol{F}_{t-1}\boldsymbol{B}_t^{0\prime} + \boldsymbol{B}_t^0\boldsymbol{F}_{t-1}^{0\prime}\right)\boldsymbol{H}' + O_p\left(N^{-1/2} + T^{-1/2} + N^{-1}T^{1/2}\right)$

(v) $(1/\sqrt{T})\overline{\widehat{\boldsymbol{F}}} = (1/\sqrt{T})\boldsymbol{H}\bar{\boldsymbol{F}} + O_p\left(N^{-1/2} + T^{-1/2} + N^{-1}T^{1/2}\right)$

(vi) $(1/\sqrt{T})\widehat{\boldsymbol{F}}^c = (1/\sqrt{T})\boldsymbol{H}\boldsymbol{F}^c + O_p\left(N^{-1/2} + T^{-1/2} + N^{-1}T^{1/2}\right)$

(vii) $(1/T^2)\sum_{t=2}^{T}\widehat{\boldsymbol{F}}_t^c\widehat{\boldsymbol{F}}_t^{c\prime} = \boldsymbol{H}\left((1/T^2)\sum_{t=2}^{T}\boldsymbol{F}_t^{0,c}\boldsymbol{F}_t^{0,c\prime}\right)\boldsymbol{H}' + O_p\left(N^{-1/2} + T^{-1/2} + N^{-1}T^{1/2}\right)$

(viii) $(1/T)\sum_{t=3}^{T}\left(\widehat{\boldsymbol{F}}_{t-1}^c\widehat{\boldsymbol{B}}_t' + \widehat{\boldsymbol{B}}_t\widehat{\boldsymbol{F}}_{t-1}^{c\prime}\right) = \boldsymbol{H}\left[(1/T)\sum_{t=3}^{T}\left(\boldsymbol{F}_{t-1}^c\boldsymbol{B}_t^{0\prime} + \boldsymbol{B}_t^0\boldsymbol{F}_{t-1}^{0,c}\right)\right]\boldsymbol{H}' + O_p\left(N^{-1/2} + T^{-1/2} + N^{-1}T^{1/2}\right)$

C.1.2 Proof of Theorem 3.2

In this part, we focus on the case that $\rho_i^0 = 1 + c_i/T$ with $c_i > 0$ for $i = 1, \ldots, N$. The case with $c_i \leq 0$ can be obtained similarly.

The Asymptotic distribution of test statistic under the null

Note that $\rho_{0,r}^0 = 1 + \nu_r/T$ with $\nu_r = 0$ for $r = 1, \ldots, R_0$ under the null, which can be regarded as the special case of generic series $\{\nu_r\}_{r=1}^R$ such that $\nu_r \neq 0$ in general. Thus, we omit the details of proofs here, and refer readers to the next subsection.

The Asymptotic distribution of test statistic under the alternative of local-to-unity

Since proofs for other estimated factors are exactly same as that for the first estimated factor, we focus on the first estimated factor for illustrations below.

Recall that $\widehat{\boldsymbol{F}}_t = \sum_{s=2}^{t} \widehat{\boldsymbol{B}}_s$. Let $\widehat{\boldsymbol{B}}^{(1)}$ be the first column of $\widehat{\boldsymbol{B}}$, and $\widehat{\boldsymbol{F}}^{(1)}$ be the first column of $\widehat{\boldsymbol{F}}$. By regressing $\widehat{\boldsymbol{B}}_t^{(1)}$ on $\widehat{\boldsymbol{F}}_{t-1}^{(1)}$, we can obtain the OLS estimator $\widehat{\nu}_1$ from (3.2 .14) as follows,

$$\widehat{\nu}_1 = \left[T_2 \sum_{t=3}^{T} \widehat{\boldsymbol{B}}_t^{(1)} \widehat{\boldsymbol{F}}_{t-1}^{(1)\prime} - \sum_{t=3}^{T} \widehat{\boldsymbol{B}}_t^{(1)} \sum_{t=3}^{T} \widehat{\boldsymbol{F}}_{t-1}^{(1)\prime}\right] \left[T_2 \sum_{t=3}^{T} \widehat{\boldsymbol{F}}_{t-1}^{(1)} \widehat{\boldsymbol{F}}_{t-1}^{(1)\prime} - \left(\sum_{t=3}^{T} \widehat{\boldsymbol{F}}_{t-1}^{(1)}\right) \left(\sum_{t=3}^{T} \widehat{\boldsymbol{F}}_{t-1}^{(1)}\right)'\right]^{-1}.$$

Because we derive the limiting distribution when T goes to infinity, $T_2 = T - 2$ will go to infinity as $T \to +\infty$. It follows that

$$\begin{aligned} & T_2 \widehat{\nu}_1 \\ = & \left\{\left(T_2^{-1} \sum_{t=3}^{T} \widehat{\boldsymbol{B}}_t^{(1)} \widehat{\boldsymbol{F}}_{t-1}^{(1)\prime}\right) - \left(T_2^{-1/2} \sum_{t=3}^{T} \widehat{\boldsymbol{B}}_t^{(1)}\right) \left(T_2^{-3/2} \sum_{t=3}^{T} \widehat{\boldsymbol{F}}_{t-1}^{(1)\prime}\right)\right\} \\ & \left\{\left(T_2^{-2} \sum_{t=3}^{T} \widehat{\boldsymbol{F}}_{t-1}^{(1)} \widehat{\boldsymbol{F}}_{t-1}^{(1)\prime}\right) - \left(T_2^{-3/2} \sum_{t=3}^{T} \widehat{\boldsymbol{F}}_{t-1}^{(1)}\right) \left(T_2^{-3/2} \sum_{t=3}^{T} \widehat{\boldsymbol{F}}_{t-1}^{(1)}\right)'\right\}^{-1} \end{aligned}$$

Let $\boldsymbol{H}_{(1)}'$ be the first row of $\boldsymbol{H}$, thus $\boldsymbol{H}_{(1)}$ is a $R_0 \times 1$ vector. Using Lemma C.1.5, for the denominator of $T_2 \widehat{\nu}_1$, we have,

$$\begin{aligned} & \left(T_2^{-2} \sum_{t=3}^{T} \widehat{\boldsymbol{F}}_{t-1}^{(1)} \widehat{\boldsymbol{F}}_{t-1}^{(1)\prime}\right) - \left(T_2^{-3/2} \sum_{t=3}^{T} \widehat{\boldsymbol{F}}_{t-1}\right) \left(T_2^{-3/2} \sum_{t=3}^{T} \widehat{\boldsymbol{F}}_{t-1}\right)' \\ = & \left\{\boldsymbol{H}_{(1)}' \left(T_2^{-2} \sum_{t=3}^{T} \boldsymbol{F}_{t-1}^{0} \boldsymbol{F}_{t-1}^{0\prime}\right) \boldsymbol{H}_{(1)} - \boldsymbol{H}_{(1)}' \left(T_2^{-3/2} \sum_{t=3}^{T} \boldsymbol{F}_{t-1}^{0}\right) \left(T_2^{-3/2} \sum_{t=3}^{T} \boldsymbol{F}_{t-1}^{0}\right)' \boldsymbol{H}_{(1)}\right\} \\ & \times \left[1 + o_p(1)\right]. \end{aligned}$$

Similarly, for the numerator of $T_2\widehat{\nu}_1$, we have

$$\begin{aligned}
&\left(T_2^{-1}\sum_{t=3}^{T}\widehat{\boldsymbol{B}}_t^{(1)}\widehat{\boldsymbol{F}}_{t-1}^{(1)\prime}\right)-\left(T_2^{-1/2}\sum_{t=3}^{T}\widehat{\boldsymbol{B}}_t^{(1)}\right)\left(T_2^{-3/2}\sum_{t=3}^{T}\widehat{\boldsymbol{F}}_{t-1}^{(1)\prime}\right)\\
=\ &\left\{\boldsymbol{H}_{(1)}^{\prime}\left(T_2^{-1}\sum_{t=3}^{T}\boldsymbol{B}_t^{0}\boldsymbol{F}_{t-1}^{0\prime}\right)\boldsymbol{H}_{(1)}-\boldsymbol{H}_{(1)}^{\prime}\left(T_2^{-1/2}\sum_{t=3}^{T}\boldsymbol{B}_t^{0}\right)\left(T_2^{-3/2}\sum_{t=3}^{T}\boldsymbol{F}_{t-1}^{0}\right)^{\prime}\boldsymbol{H}_{(1)}\right\}\\
&\times\left[1+o_p\left(1\right)\right].
\end{aligned}$$

Further, recall that $\boldsymbol{B}_t^0=T^{-1}\boldsymbol{\nu}\boldsymbol{F}_{t-1}^0+\boldsymbol{u}_t$ with $\boldsymbol{\nu}\equiv diag(\nu_1,\ldots,\nu_{R_0})$ under the alternative of local-to-unity, then, we define that

$$\begin{aligned}
&\overline{\nu}^{(1)} \qquad\qquad\qquad\qquad \text{(C.1.1)}\\
\equiv\ &\left\{\boldsymbol{H}_{(1)}^{\prime}\boldsymbol{\nu}\left(T_2^{-2}\sum_{t=3}^{T}\boldsymbol{F}_{t-1}^{0}\boldsymbol{F}_{t-1}^{0\prime}\right)\boldsymbol{H}_{(1)}-\boldsymbol{H}_{(1)}^{\prime}\boldsymbol{\nu}\left(T_2^{-3/2}\sum_{t=3}^{T}\boldsymbol{F}_{t-1}^{0}\right)\left(T_2^{-3/2}\sum_{t=3}^{T}\boldsymbol{F}_{t-1}^{0}\right)^{\prime}\boldsymbol{H}_{(1)}\right\}\\
&\left\{\boldsymbol{H}_{(1)}^{\prime}\left(T_2^{-2}\sum_{t=3}^{T}\boldsymbol{F}_{t-1}^{0}\boldsymbol{F}_{t-1}^{0\prime}\right)\boldsymbol{H}_{(1)}-\boldsymbol{H}_{(1)}^{\prime}\left(T_2^{-3/2}\sum_{t=3}^{T}\boldsymbol{F}_{t-1}^{0}\right)\left(T_2^{-3/2}\sum_{t=3}^{T}\boldsymbol{F}_{t-1}^{0}\right)^{\prime}\boldsymbol{H}_{(1)}\right\}^{-1}.
\end{aligned}$$

Using Lemma C.3.2 1(a)-(d), it is straightforward to see that $\nu_R^{(1)}=O_p\left(1\right)$ from (C.1.2), we then have,

$$\begin{aligned}
&T_2\widehat{\nu}_1\\
=\ &\left(\overline{\nu}_R^{(1)}+\left\{\boldsymbol{H}_{(1)}^{\prime}\left(T_2^{-1}\sum_{t=3}^{T}\boldsymbol{u}_t\boldsymbol{F}_{t-1}^{0\prime}\right)\boldsymbol{H}_{(1)}-\boldsymbol{H}_{(1)}^{\prime}\left(T_2^{-1/2}\sum_{t=3}^{T}\boldsymbol{u}_t\right)\left(T_2^{-3/2}\sum_{t=3}^{T}\boldsymbol{F}_{t-1}^{0}\right)^{\prime}\boldsymbol{H}_{(1)}\right\}\right.\\
&\left.\left\{\boldsymbol{H}_{(1)}^{\prime}\left(T_2^{-2}\sum_{t=3}^{T}\boldsymbol{F}_{t-1}^{0}\boldsymbol{F}_{t-1}^{0\prime}\right)\boldsymbol{H}_{(1)}-\boldsymbol{H}_{(1)}^{\prime}\left(T_2^{-3/2}\sum_{t=3}^{T}\boldsymbol{F}_{t-1}^{0}\right)\left(T_2^{-3/2}\sum_{t=3}^{T}\boldsymbol{F}_{t-1}^{0}\right)^{\prime}\boldsymbol{H}_{(1)}\right\}^{-1}\right)\\
&\times\left[1+o_p\left(1\right)\right].
\end{aligned}$$

Then, let $\boldsymbol{H}_{(1),\infty}^{\prime}$ be the fist row of $\boldsymbol{H}_{\infty}=\lim_{N,T\to\infty}\boldsymbol{H}$, we further have that

$$T\widehat{\nu}_1\Rightarrow\overline{\nu}_{\infty}^{(1)}+\Psi_{\nu}^{(1)},$$

where

$$\overline{\nu}_{\infty}^{(1)} \equiv \left\{\boldsymbol{H}_{(1),\infty}'\boldsymbol{\nu}\Sigma_u^{1/2}\left[\int_0^1 \mathbf{J}_\nu(r)\mathbf{J}_\nu(r)'\mathrm{dr} - \left(\int_0^1 \mathbf{J}_\nu(r)'\mathrm{dr}\right)\left(\int_0^1 \mathbf{J}_\nu(r)\mathrm{dr}\right)'\right]\Sigma_u^{1/2}\boldsymbol{H}_{(1),\infty}\right\}$$
$$\left\{\boldsymbol{H}_{(1),\infty}'\Sigma_u^{1/2}\left[\int_0^1 \mathbf{J}_\nu(r)\mathbf{J}_\nu(r)'\mathrm{dr} - \left(\int_0^1 \mathbf{J}_\nu(r)'\mathrm{dr}\right)\left(\int_0^1 \mathbf{J}_\nu(r)\mathrm{dr}\right)'\right]\Sigma_u^{1/2}\boldsymbol{H}_{(1),\infty}\right\}^{-1},$$

and

$$\Psi_\nu^{(1)} \equiv \left\{\boldsymbol{H}_{(1),\infty}'\Sigma_u^{1/2}\left[\int_0^1 \mathrm{d}\mathbf{W}(r)'\mathbf{J}_\nu(r) - \boldsymbol{W}(1)\int_0^1 \mathbf{J}_\nu(r)'\mathrm{dr}\right]\Sigma_u^{1/2}\boldsymbol{H}_{(1),\infty} + \boldsymbol{H}_{(1),\infty}'\Omega_u\boldsymbol{H}_{(1),\infty}\right\}$$
$$\left\{\boldsymbol{H}_{(1),\infty}'\Sigma_u^{1/2}\left[\int_0^1 \mathbf{J}_\nu(r)\mathbf{J}_\nu(r)'\mathrm{dr} - \left(\int_0^1 \mathbf{J}_\nu(r)'\mathrm{dr}\right)\left(\int_0^1 \mathbf{J}_\nu(r)\mathrm{dr}\right)'\right]\Sigma_u^{1/2}\boldsymbol{H}_{(1),\infty}\right\}^{-1}.$$

where $\mathbf{J}_\nu(r) \equiv (\mathrm{J}_{\nu_1}(r), \ldots, \mathrm{J}_{\nu_R}(r))$ is a Ornstein-Uhlenbeck process such that $\mathbf{J}_\nu(0) = \mathbf{0}$ and $\mathbf{J}_\nu(r) = \boldsymbol{W}(r) + \boldsymbol{\nu}\int_0^r e^{(r-s)c}\boldsymbol{W}(s)\mathrm{d}s$; besides, $\boldsymbol{W}(r)$ is the R_0-vector standard Brownian motion on $\mathcal{C}[0,1]$ that is given by the weak limit of the partial sum $\Sigma_u^{-1/2}T^{-1/2}\sum_{t=1}^{[Tr]}\boldsymbol{u}_t$.

The estimated variance of $\widehat{\nu}_1$ is calculated as

$$\widehat{\varpi}_{\widehat{\nu}_1} = \widehat{\sigma}_{(1)}^2\left[\sum_{t=3}^{T}\widehat{\boldsymbol{F}}_{t-1}^{(1)}\widehat{\boldsymbol{F}}_{t-1}^{(1)\prime} - T^{-1}\left(\sum_{t=3}^{T}\widehat{\boldsymbol{F}}_{t-1}^{(1)}\right)\left(\sum_{t=3}^{T}\widehat{\boldsymbol{F}}_{t-1}^{(1)}\right)'\right]^{-1},$$

where $\widehat{\sigma}_{(1)}^2 = T_2^{-1}\sum_{t=3}^{T}\widehat{\boldsymbol{B}}_t^{(1)}\widehat{\boldsymbol{B}}_t^{(1)\prime}$. Recall that $\boldsymbol{B}_t^0 = T^{-1}\boldsymbol{\nu}\boldsymbol{F}_{t-1}^0 + \boldsymbol{u}_t$, it is straightforward to see that

$$\begin{aligned} T_2^{-1}\sum_{t=3}^{T}\widehat{\boldsymbol{B}}_t^{(1)}\widehat{\boldsymbol{B}}_t^{(1)\prime} &= T_2^{-1}\sum_{t=3}^{T}\boldsymbol{H}_{(1)}'\boldsymbol{B}_t^0\boldsymbol{B}_t^{0\prime}\boldsymbol{H}_{(1)} \\ &= T_2^{-1}\sum_{t=3}^{T}\boldsymbol{H}_{(1)}'\boldsymbol{u}_t\boldsymbol{u}_t'\boldsymbol{H}_{(1)} + O_p\left(T^{-1}\right) \end{aligned}$$

Using Lemma C.1.5 and Lemma C.3.2, we then have $\widehat{\sigma}_{(1)}^2 \xrightarrow{p} \boldsymbol{H}_{(1),\infty}'\Sigma_u\boldsymbol{H}_{(1),\infty}$.

It follows that

$$
\begin{aligned}
T^2Var(\widehat{\nu}_1) &= \widehat{\sigma}^2_{(1)}\left[T^{-2}\sum_{t=3}^{T}\widehat{\boldsymbol{F}}^{(1)}_{t-1}\widehat{\boldsymbol{F}}^{(1)\prime}_{t-1} - \left(T^{-3/2}\sum_{t=3}^{T}\widehat{\boldsymbol{F}}_{t-1}\right)\left(T^{-3/2}\sum_{t=3}^{T}\widehat{\boldsymbol{F}}_{t-1}\right)'\right]^{-1} \\
\Rightarrow & \left(\boldsymbol{H}'_{(1),\infty}\Sigma_u\boldsymbol{H}_{(1),\infty}\right) \\
& \times\left\{\boldsymbol{H}'_{(1),\infty}\Sigma_u^{1/2}\left[\int_0^1\mathbf{J}_\nu(r)\mathbf{J}_\nu(r)'\mathrm{dr} - \left(\int_0^1\mathbf{J}_\nu(r)'\mathrm{dr}\right)\left(\int_0^1\mathbf{J}_\nu(r)\mathrm{dr}\right)'\right]\Sigma_u^{1/2}\boldsymbol{H}_{(1),\infty}\right\}^{-1}.
\end{aligned}
$$

Collecting above immediate results, we can obtain the weak convergence of the proposed test statistic constructed from $\{\widehat{\boldsymbol{B}}^{(1)}_t\}_{t=2}^T$ as shown below,

$$
\begin{aligned}
DF^{(1),f} &= \{T\widehat{\nu}_1\}\left\{T^2Var(\widehat{\nu}_1)\right\}^{-1/2} \\
&\Rightarrow \overline{\chi}^{(1),f}_\nu + \psi^{(1),f}_\nu,
\end{aligned}
$$

where

$$
\begin{aligned}
\overline{\chi}^{(1),f}_\nu \equiv \quad & \left(\boldsymbol{H}'_{(1),\infty}\Sigma_u\boldsymbol{H}_{(1),\infty}\right)^{-1/2} \\
& \left\{\boldsymbol{H}'_{(1),\infty}\boldsymbol{\nu}\Sigma_u^{1/2}\left[\int_0^1\mathbf{J}_\nu(r)\mathbf{J}_\nu(r)'\mathrm{dr} - \left(\int_0^1\mathbf{J}_\nu(r)'\mathrm{dr}\right)\left(\int_0^1\mathbf{J}_\nu(r)\mathrm{dr}\right)'\right]\Sigma_u^{1/2}\boldsymbol{H}_{(1),\infty}\right\} \\
& \left\{\boldsymbol{H}'_{(1),\infty}\Sigma_u^{1/2}\left[\int_0^1\mathbf{J}_\nu(r)\mathbf{J}_\nu(r)'\mathrm{dr} - \left(\int_0^1\mathbf{J}_\nu(r)'\mathrm{dr}\right)\left(\int_0^1\mathbf{J}_\nu(r)\mathrm{dr}\right)'\right]\Sigma_u^{1/2}\boldsymbol{H}_{(1),\infty}\right\}^{-1/2},
\end{aligned}
$$

and

$$
\begin{aligned}
\psi^{(1),f}_\nu \equiv \quad & \left(\boldsymbol{H}'_{(1),\infty}\Sigma_u\boldsymbol{H}_{(1),\infty}\right)^{-1/2} \\
& \left\{\boldsymbol{H}'_{(1),\infty}\Sigma_u^{1/2}\left[\int_0^1\mathrm{d}\mathbf{W}(r)'\mathbf{J}_\nu(r) - \boldsymbol{W}(1)\int_0^1\mathbf{J}_\nu(r)'\mathrm{dr}\right]\Sigma_u^{1/2}\boldsymbol{H}_{(1),\infty} + \boldsymbol{H}'_{(1),\infty}\Omega_u\boldsymbol{H}_{(1),\infty}\right\} \\
& \left\{\boldsymbol{H}'_{(1),\infty}\Sigma_u^{1/2}\left[\int_0^1\mathbf{J}_\nu(r)\mathbf{J}_\nu(r)'\mathrm{dr} - \left(\int_0^1\mathbf{J}_\nu(r)'\mathrm{dr}\right)\left(\int_0^1\mathbf{J}_\nu(r)\mathrm{dr}\right)'\right]\Sigma_u^{1/2}\boldsymbol{H}_{(1),\infty}\right\}^{-1/2}.
\end{aligned}
$$

As shown above, we complete proofs for the limiting distribution of the proposed test statistic under the alternative of local-to-unity.

Note that under the null that $\nu_r = 0$ for all r, $\overline{\chi}_\nu^{(1),f} = \overline{\chi}_0^{(1),f} = 0$, and $\psi_\nu^{(1),f}$ turns out to be $\psi_0^{(1),f}$ with R_0-dimensional vector standard Brownian motion $\boldsymbol{W}(r)$ in place of the Ornstein-Uhlenbeck diffusion process $\mathbf{J}_\nu(r)$.

C.1.3 Proof of Theorem 3.3

Arguments of the proofs for Theorem 3.3.3 are almost the same as those for Theorem 3.3.2 in the last section. To see this point, recall that $\widehat{\nu}_{R_0}$ is obtained by regressing $\widehat{\boldsymbol{B}}_t^{R,c}$ on $\widehat{\boldsymbol{F}}_{t-1}^{R,c}$. Then, similar to $\widehat{\nu}_{(r)}$, we have,

$$\widehat{\nu}_{R_0} = \left[T_2 \sum_{t=3}^{T} \widehat{\boldsymbol{B}}_t^{R_0} \widehat{\boldsymbol{F}}_{t-1}^{R_0\prime} - \sum_{t=3}^{T} \widehat{\boldsymbol{B}}_t^{R_0} \sum_{t=3}^{T} \widehat{\boldsymbol{F}}_{t-1}^{R_0\prime}\right] \left[T_2 \sum_{t=3}^{T} \widehat{\boldsymbol{F}}_{t-1}^{R_0} \widehat{\boldsymbol{F}}_{t-1}^{R_0\prime} - \left(\sum_{t=3}^{T} \widehat{\boldsymbol{F}}_{t-1}^{R_0}\right) \left(\sum_{t=3}^{T} \widehat{\boldsymbol{F}}_{t-1}^{R_0}\right)'\right]^{-1}.$$

Similar to (C.1.2), we have

$$\begin{aligned} & T_2 \widehat{\nu}_{R_0} \qquad\qquad \text{(C.1.2)} \\ = & \left(\overline{\nu}_{R_0} + \left\{\boldsymbol{H}_{R_0}' \left(T_2^{-1} \sum_{t=3}^{T} \boldsymbol{u}_t \boldsymbol{F}_{t-1}^{0\prime}\right) \boldsymbol{H}_{R_0} - \boldsymbol{H}_{R_0}' \left(T_2^{-1/2} \sum_{t=3}^{T} \boldsymbol{u}_t\right) \left(T_2^{-3/2} \sum_{t=3}^{T} \boldsymbol{F}_{t-1}^{0}\right)' \boldsymbol{H}_{R_0}\right\}\right. \\ & \left.\left\{\boldsymbol{H}_{R_0}' \left(T_2^{-2} \sum_{t=3}^{T} \boldsymbol{F}_{t-1}^{0} \boldsymbol{F}_{t-1}^{0\prime}\right) \boldsymbol{H}_{R_0} - \boldsymbol{H}_{R_0}' \left(T_2^{-3/2} \sum_{t=3}^{T} \boldsymbol{F}_{t-1}^{0}\right) \left(T_2^{-3/2} \sum_{t=3}^{T} \boldsymbol{F}_{t-1}^{0}\right)' \boldsymbol{H}_{R_0}\right\}^{-1}\right) \\ & \times \left[1 + o_p(1)\right]. \end{aligned}$$

where $\boldsymbol{H}_{R_0} \equiv \sum_{r=1}^{R_0} \boldsymbol{H}_{(r)}$, and $\overline{\nu}^{R_0}$ is defined as

$$\begin{aligned} & \overline{\nu}^{R_0} \qquad\qquad \text{(C.1.3)} \\ \equiv & \left\{\boldsymbol{H}_{R_0}' \boldsymbol{\nu} \left(T_2^{-2} \sum_{t=3}^{T} \boldsymbol{F}_{t-1}^{0} \boldsymbol{F}_{t-1}^{0\prime}\right) \boldsymbol{H}_{R_0} - \boldsymbol{H}_{R_0}' \boldsymbol{\nu} \left(T_2^{-3/2} \sum_{t=3}^{T} \boldsymbol{F}_{t-1}^{0}\right) \left(T_2^{-3/2} \sum_{t=3}^{T} \boldsymbol{F}_{t-1}^{0}\right)' \boldsymbol{H}_{R_0}\right\} \\ & \left\{\boldsymbol{H}_{R_0}' \left(T_2^{-2} \sum_{t=3}^{T} \boldsymbol{F}_{t-1}^{0} \boldsymbol{F}_{t-1}^{0\prime}\right) \boldsymbol{H}_{R_0} - \boldsymbol{H}_{R_0}' \left(T_2^{-3/2} \sum_{t=3}^{T} \boldsymbol{F}_{t-1}^{0}\right) \left(T_2^{-3/2} \sum_{t=3}^{T} \boldsymbol{F}_{t-1}^{0}\right)' \boldsymbol{H}_{R_0}\right\}^{-1}. \end{aligned}$$

From (C.1.2) and (C.1.3) above, we can see that $T_2 \widehat{\nu}_{R_0}$ has the similar structure as $T_2 \widehat{\nu}_{(r)}$ for

$r = 1, \ldots, R_0$.

Similarly, $\widehat{\varpi}_{\widehat{\nu}_{R_0}}$ also has similar structure as $\widehat{\varpi}_{\widehat{\nu}_1}$ shown in the last section.

By imitating the arguments of proofs for Theorem 3.3.2 in the last section, the desired result follows directly.

C.1.4 Proof of Theorem 3.4

In this section, we demonstrate the consistency of dependent wild bootstrap (DWB) for the proposed test when the alternative of local-to-unity holds, namely, $\rho_{0,r}^0 = 1 + \nu_r/T$ for $r = 1, \ldots, R_0$. Under the null, $\nu_r = 0$ holds for all $r = 1, \ldots, T$, proofs below show the consistency of DWB in approximating the limiting distributions under the null. We follow proofs for Theorem 3.1 and 3.2 presented in Rho and Shao (2019). Without loss of generality, we also focus on the first estimated as done in the proofs of Theorem 3.1, and proofs for other estimated factors are exactly same as those for the first estimated factor presented below.

Let P^* denote the probability measure induced by the dependent wild bootstrap conditional on $\mathcal{X} = \{Z_{it}\}$. Let E^* and Var^* denote the expectation and variance under P^* and O_p^* and o_p^* the probability order under P^*.

(a) We will follow lines in Rho and Shao (2019) to apply large-block-small-block method. Let $L_T = \left\lfloor (T/l_T)^{1/2} \right\rfloor$ be the length of a large-block and $l_T \asymp M \cdot T^\eta$ with $\eta \in (0, 1/3)$ be that of a small-block. Our goal is to assign points $t \in \{1, 2, \ldots, \lfloor Tr \rfloor\}$ to alternating large and small blocks. Let $K_T = K_{T,r} = \left\lfloor \lfloor Tr \rfloor (L_T + l_T)^{-1} \right\rfloor$ be the number of the large (small) blocks. For $0 \le r_1 < r_2 \le 1$, $K_1 = K_{T,r_1}$ and $K_1 = K_{T,r_2}$.

Define the k-th large-block $\mathcal{L}_k = \{j \in \mathbb{N} : (k-1)(L_T + l_T) + 1 \le j \le k(l_T + L_T) - l_T\}$ for $1 \le k \le K_T$, the k th small-block $\mathcal{S}_k = \{j \in \mathbb{N} : k(L_T + l_T) - l_T + 1 \le j \le k(l_T + L_T)\}$ for $1 \le k \le K_T - 1$, and $\mathcal{S}_{K_T} = \{j \in \mathbb{N} : K_T(L_T + l_T) - l_T + 1 \le j \le \lfloor Tr \rfloor\}$. And note that $L_T \to \infty$ and $l_T = o(L_T)$.

For notation convenience, define $u_{t,(1)} \equiv \boldsymbol{H}'_{(1)}\boldsymbol{u}_t$ for all t. Denote the residual $\{\widehat{u}_{t,(1)}\}_{t=3}^{T}$ that are obtained by regressing $\widehat{\boldsymbol{B}}_t^{(1)}$ on $\widehat{\boldsymbol{F}}_{t-1}^{(1)}$. Denote the bootstrap version of $\widehat{u}_{t,,(1)}$ by $u^*_{t,(1)}$. In addition, let $U_k = \sum_{j\in\mathcal{L}_k} W_j u_{j,(1)}$ and $V_k = \sum_{j\in\mathcal{S}_k} W_j u_{j,(1)}$ for $k = 1,\ldots,K_T$.

Before the proofs of Theorem 3.2, we state an important result as the lemma below used in following proofs. Let $u^*_{t,(r)} \equiv \widehat{u}_{t,(r)}W_t$, then, for $r = 1,\ldots,R_0$

Lemma C.1.6. *For any $\rho^0_{0,r} = 1 + \nu_r/T, \nu_r \geq 0$, then, for r-th row of the rotation matrix $\boldsymbol{H}$, as (N,T) goes to infinity simultaneously,*

$$T^{-1/2}\sum_{t=3}^{\lfloor Tr\rfloor} u^*_{t,(r)} \Rightarrow \boldsymbol{H}'_{(r),\infty}\Sigma_u^{1/2}\boldsymbol{W}(r) \quad \textit{in probability.}$$

Lemma C.1.6 above states that, in the bootstrap world, the functional CLT still holds for both the null hypothesis with $\rho^0_{0,r} = 1$ for all r and under the alternative of local-to-unity with $\rho^0_{0,r} = 1 + \nu_r/T$ with $\nu_r > 0$ for all r.

Next, we prove the consistency of DWB for the proposed test as an application of lemma C.1.6 above.

We claim that under the local alternatives,

$$T^{-1}\sum_{t=3}^{T}\left\{\left(u^*_{t,(1)}\right)^2 - E^*\left(u^*_{t,(1)}\right)^2\right\} = o^*_p(1), \tag{C.1.4}$$

and

$$T^{-1}\sum_{t=3}^{T}\left\{E^*\left(u^*_{t,(1)}\right)^2 - u^2_{t,(1)}\right\} = o_p(1). \tag{C.1.5}$$

Once (C.1.4) and (C.1.5) are established, it follows that $T^{-1}\sum_{t=3}^{T}\left\{\left(u^*_{t,(1)}\right)^2 - u^2_{t,(1)}\right\} = o^*_p(1)$. Then, by repeating arguments of proofs for Theorem 3.1 shown in the previous section, Theorem 3.2 holds directly as an application of the continuous mapping theorem, Lemma C.1.6, and the fact that $T^{-1}\sum_{t=3}^{T}\widehat{u}_{t,(1)}^{2}\xrightarrow{p}\boldsymbol{H}_{(1),\infty}'\Sigma_{u}\boldsymbol{H}_{(1),\infty}$.

To prove (C.1.5), using Lemma C.1.5, Lemma C.3.2, $\boldsymbol{B}_t^0=T^{-1}\boldsymbol{\nu}\boldsymbol{F}_{t-1}^0+\boldsymbol{u}_t$ by construction, and the fact that $\widehat{\nu}_1=O_p\left(T^{-1}\right)$ justified in the proofs of Theorem 3.1, we can obtain (C.1.5) by direct calculations as follows,

$$
\begin{aligned}
& T^{-1}\sum_{t=3}^{T}E^{*}\left(u_{t,(1)}^{*}\right)^{2}=T^{-1}\sum_{t=3}^{T}\widehat{u}_{t,(1)}^{2} \qquad \text{(C.1.6)}\\
= \;& T^{-1}\sum_{t=3}^{T}\left[\widehat{\boldsymbol{B}}_{t}^{(1)}-\widehat{\nu}_{1}\widehat{\boldsymbol{F}}_{t-1}^{(1)}\right]^{2}\\
= \;& T^{-1}\sum_{t=3}^{T}\left(\widehat{\boldsymbol{B}}_{t}^{(1)}\right)^{2}+\widehat{\nu}_{1}^{2}T^{-1}\sum_{t=3}^{T}\left(\widehat{\boldsymbol{F}}_{t-1}^{(1)}\right)^{2}-2\widehat{\nu}_{1}T^{-1}\sum_{t=3}^{T}\widehat{\boldsymbol{B}}_{t}^{(1)}\widehat{\boldsymbol{F}}_{t-1}^{(1)}\\
= \;& \left[\boldsymbol{H}_{(1)}'\left(T^{-1}\sum_{t=3}^{T}\boldsymbol{B}_{t}^{0}\boldsymbol{B}_{t}^{0\prime}\right)\boldsymbol{H}_{(1)}\right]+\widehat{\nu}_{1}^{2}T\left[\boldsymbol{H}_{(1)}'\left(T^{-2}\sum_{t=3}^{T}\boldsymbol{F}_{t-1}^{0}\boldsymbol{F}_{t-1}^{0\prime}\right)\boldsymbol{H}_{(1)}\right]\\
& -2\widehat{\nu}_{1}\left[\boldsymbol{H}_{(1)}'\left(T^{-1}\sum_{t=3}^{T}\boldsymbol{B}_{t}^{0}\boldsymbol{F}_{t-1}^{0\prime}\right)\boldsymbol{H}_{(1)}\right]+O_p\left(N^{-1/2}+T^{-1/2}+N^{-1}T^{1/2}\right)\\
= \;& \left(T^{-1}\sum_{t=3}^{T}\boldsymbol{H}_{(1)}'\boldsymbol{u}_{t}^{0}\boldsymbol{u}_{t}^{0\prime}\boldsymbol{H}_{(1)}\right)+O_p\left(T^{-1}\right)+O_p\left(N^{-1/2}+T^{-1/2}+N^{-1}T^{1/2}\right).
\end{aligned}
$$

Now we shall prove (C.1.4). Observe that

$$
\sum_{t=3}^{T}\left\{\left(u_{t,(1)}^{*}\right)^{2}-E^{*}\left(u_{t,(1)}^{*}\right)^{2}\right\}=\sum_{t=3}^{T}\widehat{u}_{t,(1)}^{2}\left(W_t^2-1\right).
$$

For any $\delta>0$,

$$
\begin{aligned}
P^{*}\left\{\left|\sum_{t=3}^{T}\widehat{u}_{t,(1)}^{2}\left(W_t^2-1\right)\right|>T\delta\right\} &\leq (T\delta)^{-2}E^{*}\left\{\sum_{t=3}^{T}\widehat{u}_{t,(1)}^{2}\left(W_t^2-1\right)\right\}^{2}\\
&\leq (T\delta)^{-2}C\left\{\sum_{t=3}^{T}\sum_{h=0}^{l_T}\widehat{u}_{t,(1)}^{2}\widehat{u}_{t+h,(1)}^{2}\right\},
\end{aligned}
$$

where the second line above follows by the properties of $\{W_t\}$ under the DWB scheme, and it remains to show $T^{-2}\sum_{t=3}^{T}\sum_{h=0}^{l_T}\widehat{u}_{t,(1)}^2\widehat{u}_{t+h,(1)}^2 = o_p(1)$ as below.

Substitute $\widehat{u}_{t,(1)} = \widehat{\boldsymbol{B}}_t^{(1)} - \widehat{\nu}_1\widehat{\boldsymbol{F}}_{t-1}^{(1)}$ into $T^{-2}\sum_{t=3}^{T}\sum_{h=0}^{l_T}\widehat{u}_{t,(1)}^2\widehat{u}_{t+h,(1)}^2$, denote $\Theta_{NT} = (N^{-1/2} + T^{-1/2} + N^{-1}T^{1/2}$, by direct calculations, we have

$$
\begin{aligned}
& T^{-2}\sum_{t=3}^{T}\sum_{h=0}^{l_T}\widehat{u}_{t,(1)}^2\widehat{u}_{t+h,(1)}^2 \\
= & \sum_{t=3}^{T}\sum_{h=0}^{l_T}\left\{\left(T^{-1/2}\widehat{\boldsymbol{B}}_t^{(1)} - T^{-1/2}\widehat{\nu}_1\widehat{\boldsymbol{F}}_{t-1}^{(1)}\right)^2\left(T^{-1/2}\widehat{\boldsymbol{B}}_{t+h}^{(1)} - T^{-1/2}\widehat{\nu}_1\widehat{\boldsymbol{F}}_{t+h-1}^{(1)}\right)^2\right\} \\
= & \sum_{t=3}^{T}\sum_{h=0}^{l_T}\Big\{\boldsymbol{H}_{(1)}'\left[T^{-1/2}\boldsymbol{B}_t^0 - T^{-1/2}\widehat{\nu}_1\boldsymbol{F}_{t-1}^0 + o_p(\Theta_{NT})\right]\left[T^{-1/2}\boldsymbol{B}_t^0 - T^{-1/2}\widehat{\nu}_1\boldsymbol{F}_{t-1}^0 + o_p(\Theta_{NT})\right]' \\
& \boldsymbol{H}_{(1)}\boldsymbol{H}_{(1)}'\left[T^{-1/2}\boldsymbol{B}_t^0 - T^{-1/2}\widehat{\nu}_1\boldsymbol{F}_{t-1}^0 + o_p(\Theta_{NT})\right]\left[T^{-1/2}\boldsymbol{B}_t^0 - T^{-1/2}\widehat{\nu}_1\boldsymbol{F}_{t-1}^0 + o_p(\Theta_{NT})\right]'\boldsymbol{H}_{(1)}\Big\} \\
= & \sum_{t=3}^{T}\sum_{h=0}^{l_T}\left\{\boldsymbol{H}_{(1)}'\left[T^{-1/2}\left(T^{-1}\boldsymbol{\nu} - \widehat{\nu}_1\cdot\boldsymbol{I}_R\right)\boldsymbol{F}_{t-1}^0 + T^{-1/2}\boldsymbol{u}_t + o_p(\Theta_{NT})\right]\right. \\
& \quad\times\left[T^{-1/2}\left(T^{-1}\boldsymbol{\nu} - \widehat{\nu}_1\cdot\boldsymbol{I}_R\right)\boldsymbol{F}_{t-1}^0 + T^{-1/2}\boldsymbol{u}_t + o_p(\Theta_{NT})\right]'\boldsymbol{H}_{(1)} \\
& \quad\times\boldsymbol{H}_{(1)}'\left[T^{-1/2}\left(T^{-1}\boldsymbol{\nu} - \widehat{\nu}_1\cdot\boldsymbol{I}_R\right)\boldsymbol{F}_{t+h-1}^0 + T^{-1/2}\boldsymbol{u}_{t+h} + o_p(\Theta_{NT})\right] \\
& \quad\left.\times\left[T^{-1/2}\left(T^{-1}\boldsymbol{\nu} - \widehat{\nu}_1\cdot\boldsymbol{I}_R\right)\boldsymbol{F}_{t+h-1}^0 + T^{-1/2}\boldsymbol{u}_{t+h} + o_p(\Theta_{NT})\right]'\boldsymbol{H}_{(1)}\right\} \\
= & \left\{T^{-2}\sum_{t=3}^{T}\sum_{h=0}^{l_T}\boldsymbol{H}_{(1)}'\boldsymbol{u}_t\boldsymbol{u}_t'\boldsymbol{H}_{(1)}\boldsymbol{H}_{(1)}'\boldsymbol{u}_{t+h}\boldsymbol{u}_{t+h}'\boldsymbol{H}_{(1)}\right\}[1 + o_p(1)].
\end{aligned}
$$

where the second to the last line above follows because we use following facts implicitly:

(1) $\boldsymbol{B}_t^0 = T^{-1}\boldsymbol{\nu}\boldsymbol{F}_{t-1}^0 + \boldsymbol{u}_t$,

(2) $T^{-1/2}\widehat{\boldsymbol{B}}_t = T^{-1/2}\boldsymbol{H}\boldsymbol{B}_t^0 + O_p\left(T^{-1/2}\Theta_{NT}\right)$,

(3) $T^{-1/2}\widehat{\boldsymbol{F}}_t = T^{-1/2}\boldsymbol{H}\boldsymbol{F}_t^0 + O_p\left(T^{-1/2}\Theta_{NT}\right)$ uniformly in t by Lemma C.1.5(i),

(4) $\|T^{-1}\boldsymbol{\nu} - \widehat{\nu}_1\cdot\boldsymbol{I}_R\| \le \|T^{-1}\boldsymbol{\nu}\| + \|\widehat{\nu}_1\| = O\left(T^{-1}\right)$ because $\|T^{-1}\boldsymbol{\nu}\| = O\left(T^{-1}\right)$ under the alternative of local-to-unity and $\widehat{\nu}_1 = O_p\left(T^{-1}\right)$ by Theorem 3.3.2, and

(5) $T^{-1/2}\boldsymbol{F}_{[Tr]}^0 = O_p(1)$ by Lemma C.3.2 1(a).

By Cauchy-Schwarz inequality, we have that

$$
\begin{aligned}
& T^{-2}\sum_{t=3}^{T}\sum_{h=0}^{l_T}\boldsymbol{H}_{(1)}'\boldsymbol{u}_t\boldsymbol{u}_t'\boldsymbol{H}_{(1)}\boldsymbol{H}_{(1)}'\boldsymbol{u}_{t+h}\boldsymbol{u}_{t+h}'\boldsymbol{H}_{(1)} \\
\leq\; & T^{-1}l_T\left[T^{-1}\sum_{t=3}^{T}\left(\boldsymbol{H}_{(1)}'\boldsymbol{u}_t\boldsymbol{u}_t'\boldsymbol{H}_{(1)}\right)^2\right]^{1/2}\left[T^{-1}\sum_{t=3}^{T}\left(l_T^{-1}\sum_{h=0}^{l_T}\boldsymbol{H}_{(1)}'\boldsymbol{u}_{t+h}\boldsymbol{u}_{t+h}'\boldsymbol{H}_{(1)}\right)^2\right]^{1/2} \\
\leq\; & T^{-1}l_T\left(\|\boldsymbol{H}\|^4T^{-1}\sum_{t=3}^{T}\|\boldsymbol{u}_t\|^4\right)^{1/2}\left[\|\boldsymbol{H}\|^4T^{-1}\sum_{t=3}^{T}\left(l_T^{-1}\sum_{h=0}^{l_T}\|\boldsymbol{u}_{t+h}\|^2\right)\right]^{1/2} \\
=\; & O_p\left(T^{-1}l_T\right)=o_p(1)
\end{aligned}
$$

where the last line above holds under Assumption A3.1. Collecting above arguments, (C.1.4) follows directly.

Proofs of Lemma C.1.6: Note that $E^*\left(T^{-1/2}\sum_{t=2}^{\lfloor Tr\rfloor}u_{t,(1)}^*\right)=0$. Meanwhile, because $\widehat{u}_{t,(1)}=\widehat{\boldsymbol{B}}_t^{(1)}-\widehat{\nu}_1\widehat{\boldsymbol{F}}_{t-1}^{(1)}$, we have

$$
T^{-1/2}\sum_{t=2}^{\lfloor Tr\rfloor}u_{t,(1)}^* \;=\; T^{-1/2}\sum_{t=2}^{\lfloor Tr\rfloor}\widehat{u}_{t,(1)}W_t=\sum_{t=2}^{\lfloor Tr\rfloor}\left(T^{-1/2}\widehat{\boldsymbol{B}}_t^{(1)}-\widehat{\nu}_1T^{-1/2}\widehat{\boldsymbol{F}}_{t-1}^{(1)}\right)W_t.
$$

We use following facts again: $\widehat{\boldsymbol{B}}_t=\left(\widehat{\boldsymbol{B}}_t-\boldsymbol{H}\boldsymbol{B}_t^0\right)+\boldsymbol{H}\boldsymbol{B}_t^0$, $\boldsymbol{B}_t^0=T^{-1}\boldsymbol{\nu}\boldsymbol{F}_{t-1}^0+\boldsymbol{u}_t$, $\|T^{-1}\boldsymbol{\nu}\|_{sp}=O\left(T^{-1}\right)$ under the alternative of local-to-unity, $\widehat{\nu}_1=O_p\left(T^{-1}\right)$, $T^{-1/2}\boldsymbol{F}_{[Tr]}^0\Rightarrow\Sigma_u^{1/2}\mathbf{J}_\nu(r)$ by Lemma C.3.2 1(a), and the results in Lemma C.1.5. Then, by direct calculations, we have following decompression,

$$
\begin{aligned}
T^{-1/2}\sum_{t=2}^{\lfloor Tr\rfloor}u_{t,(1)}^* \;=\;& T^{-1/2}\sum_{t=2}^{\lfloor Tr\rfloor}\left(\boldsymbol{H}_{(1)}'\boldsymbol{u}_tW_t+\boldsymbol{H}_{(1)}'\left(T^{-1}\boldsymbol{\nu}-\widehat{\nu}_1\cdot\boldsymbol{I}_R\right)\boldsymbol{F}_{t-1}^0W_t\right) \\
& +T^{-1/2}\sum_{t=2}^{\lfloor Tr\rfloor}\left(\widehat{\boldsymbol{B}}_t^{(1)}-\boldsymbol{H}_{(1)}'\boldsymbol{B}_t^0\right)W_t+\widehat{\nu}_1T^{-1/2}\sum_{t=2}^{\lfloor Tr\rfloor}\left(\widehat{\boldsymbol{F}}_{t-1}^{(1)}-\boldsymbol{H}_{(1)}'\boldsymbol{F}_{t-1}^0\right)W_t \\
\equiv\;& \mathcal{I}_{1,r}+\mathcal{I}_{2,r}+\mathcal{I}_{3,r}.
\end{aligned}
$$

Once the fact that $\mathcal{I}_{2,r}=o_p^*(1)$, $\mathcal{I}_{3,r}=o_p^*(1)$ and $\mathcal{I}_{1,r}\Rightarrow W(r)$ are established, the proof is complete. To this end, we study $\mathcal{I}_{1,r}$, $\mathcal{I}_{2,r}$ and $\mathcal{I}_{3,r}$ in turns below.

For $\mathcal{I}_{2,r}$, note that $E^*\left(\mathcal{I}_{2,r}\right)=0$, using Lemma C.1.2(i)-(ii) and Cauchy-Schwarz inequality, it follows that

$$
\begin{aligned}
E^*(\mathcal{I}_{2,r}^2) &\leq T^{-1}\sum_{t=2}^{\lfloor Tr\rfloor}\sum_{h=0}^{l_T}\left\|\widehat{\boldsymbol{B}}_t-\boldsymbol{H}\boldsymbol{B}_t^0\right\|\left\|\widehat{\boldsymbol{B}}_{t+h}-\boldsymbol{H}\boldsymbol{B}_{t+h}^0\right\|E^*\left(W_tW_{t+h}\right)\\
&= T^{-1}\sum_{t=2}^{\lfloor Tr\rfloor}\sum_{h=0}^{l_T}\left\|\widehat{\boldsymbol{B}}_t-\boldsymbol{H}\boldsymbol{B}_t^0\right\|\left\|\widehat{\boldsymbol{B}}_{t+h}-\boldsymbol{H}\boldsymbol{B}_{t+h}^0\right\|a\left(h/l_T\right)\\
&= O_p\left(l_T\left(N^{-1}+T^{-1}+N^{-2}T\right)\right)=o_p\left(1\right).
\end{aligned}
$$

where the first line holds by the properties of $\{W_t\}$ under the DWB scheme, and the last line holds as long as $l_T/N+l_T/T+l_TT/N^2\to 0$ as (N,T) go to infinity jointly. Then, by Chebyshev's inequality, we have $\mathcal{I}_{2,r}=o_p^*\left(1\right)$.

For $\mathcal{I}_{3,r}$, similar to above arguments, using Lemma C.1.2 and Lemma C.1.3, we also can obtain $\mathcal{I}_{3,r}=o_p^*\left(1\right)$.

For $\mathcal{I}_{1,r}$, we write $\mathcal{I}_{1,r}=T^{-1/2}\sum_{t=2}^{\lfloor Tr\rfloor}\boldsymbol{H}_{(1)}'\boldsymbol{u}_tW_t+T^{-1/2}\sum_{t=2}^{\lfloor Tr\rfloor}\boldsymbol{H}_{(1)}'\left(T^{-1}\boldsymbol{\nu}-\widehat{\nu}_1\cdot\boldsymbol{I}_R\right)\boldsymbol{F}_{t-1}^0W_t\equiv\mathcal{I}_{11,r}+\mathcal{I}_{12,r}$. Below, we are going to show that (i) $\mathcal{I}_{12,r}=o_p^*\left(1\right)$ and (ii) $\mathcal{I}_{11,r}\Rightarrow W(r)$.

First, we will show $\mathcal{I}_{12,r}=o_p^*\left(1\right)$. To this end, our proofs follow lines in Rho and Shao (2019) to show following two conditions holds

$$
\left|T^{-1/2}\sum_{t=2}^{\lfloor Tr\rfloor}\boldsymbol{H}_{(1)}'\left(T^{-1}\boldsymbol{\nu}-\widehat{\nu}_1\cdot\boldsymbol{I}_R\right)\boldsymbol{F}_{t-1}^0W_t\right|=o_p^*(1)\text{ for any }r\in[0,1] \tag{C.1.7}
$$

and

$$
E^*\left|T^{-1/2}\sum_{t=\lfloor Tr_1\rfloor+2}^{\lfloor Tr_2\rfloor}\boldsymbol{H}_{(1)}'\left(T^{-1}\boldsymbol{\nu}-\widehat{\nu}_1\cdot\boldsymbol{I}_R\right)\boldsymbol{F}_{t-1}^0W_t\right|^4=O_p\left(T^{-2}l_T^2+T^{-3}l_T^2L_T\right)=o_p\left(1\right) \tag{C.1.8}
$$

Note that $\|T^{-1}\boldsymbol{\nu}-\widehat{\nu}_1\cdot\boldsymbol{I}_R\|=O_p\left(T^{-1}\right)$ can be deduced by proofs of Theorem 3.3.2, which is used explicitly in following arguments.

When r in (C.1.7) equals 0, it is trivial case to hold for sure; when $r \in$ (0,1] holds in (C.1.7), by Chebyshev's inequality, for any $M > 0$,

$$\begin{aligned} & P^*\left(\left|T^{-1/2}\sum_{t=2}^{\lfloor Tr\rfloor}\boldsymbol{H}'_{(1)}\left(T^{-1}\boldsymbol{\nu}-\widehat{\nu}_1\cdot\boldsymbol{I}_R\right)\boldsymbol{F}^0_{t-1}W_t\right|>M\right) \\ \leq\ & M^{-2}E^*\left|T^{-1/2}\sum_{t=2}^{\lfloor Tr\rfloor}\boldsymbol{H}'_{(1)}\left(T^{-1}\boldsymbol{\nu}-\widehat{\nu}_1\cdot\boldsymbol{I}_R\right)\boldsymbol{F}^0_{t-1}W_t\right|^2 \\ \equiv\ & M^{-2}\mathcal{I}_{12a,r}. \end{aligned} \tag{C.1.9}$$

Below, we bound $\mathcal{I}_{12a,r}$ defined above.

$$\begin{aligned} \mathcal{I}_{12a,r} &\leq \left\|T\left(T^{-1}\boldsymbol{\nu}-\widehat{\nu}_1\cdot\boldsymbol{I}_R\right)\right\|_{sp}E^*\left|T^{-3/2}\sum_{t=2}^{\lfloor Tr\rfloor}\boldsymbol{H}'_{(1)}\boldsymbol{F}^0_{t-1}W_t\right|^2 \\ &\leq O_p(1)\,T^{-3}\sum_{t=2}^{\lfloor Tr\rfloor}\sum_{h=0}^{l_T}\boldsymbol{H}'_{(1)}\boldsymbol{F}^0_{t-1}\boldsymbol{F}^{0\prime}_{t+h-1}\boldsymbol{H}_{(1)}a(h/l_T) \\ &= O_p(1)\,T^{-2}\sum_{t=2}^{\lfloor Tr\rfloor}\sum_{h=0}^{l_T}\boldsymbol{H}'_{(1)}\left(\frac{\boldsymbol{F}^0_{t-1}}{\sqrt{T}}\right)\left(\frac{\boldsymbol{F}^0_{t+h-1}}{\sqrt{T}}\right)'\boldsymbol{H}_{(1)}a(h/l_T) \\ &\leq O_p(1)\left\|\boldsymbol{H}\right\|^2T^{-2}\sum_{t=2}^{\lfloor Tr\rfloor}\sum_{h=0}^{l_T}a(h/l_T) \\ &= O_p\left(T^{-1}l_T\right)=o_p(1). \end{aligned}$$

where the second line follows by the properties of $\{W_t\}$ under the DWB scheme, the fourth line holds because $T^{-1/2}\boldsymbol{F}^0_{[Tr]} \Rightarrow \Sigma_u^{1/2}\mathbf{J}_\nu(r)$ due to Lemma C.3.2 1(a), and the final line holds by the fact $\|\boldsymbol{H}\| = O_p(1)$ and the boundness of the kernel $a(.)$ associated with $\{W_t\}$ under the DWB scheme. Then, based on (C.1.9), we immediately obtain that

$$E\left\{P^*\left(\left|T^{-1/2}\sum_{t=2}^{\lfloor Tr\rfloor}\boldsymbol{H}'_{(1)}\left(\boldsymbol{\nu}-\widehat{\nu}_1\cdot\boldsymbol{I}_R\right)\boldsymbol{F}^0_{t-1}W_t\right|>M\right)\right\}=O\left(T^{-1}l_T\right)=o(1).$$

Then, as claimed in Rho and Shao (2019), (C.1.7) is established.

To prove (C.1.8), we recall the constructions of large blocks and small block at the beginning of this section. $L_T = \left\lfloor (T/l_T)^{1/2} \right\rfloor$ is the length of a large-block and $l_T \asymp M \cdot T^{\eta}$ with $\eta \in (0, 1/3)$ be that of a small-block. We assign points $t \in \{1, 2, \ldots, \lfloor Tr \rfloor\}$ to alternating large and small blocks. $K_T = K_{T,r} = \left\lfloor \lfloor Tr \rfloor (L_T + l_T)^{-1} \right\rfloor$ be the number of the large (small) blocks. For $0 \leq r_1 < r_2 \leq 1$, $K_1 = K_{T,r_1}$ and $K_1 = K_{T,r_2}$. The k-th large-block $\mathcal{L}_k = \{j \in \mathbb{N} : (k-1)(L_T + l_T) + 1 \leq j \leq k(l_T + L_T) - l_T\}$ for $1 \leq k \leq K_T$, the k th small-block $\mathcal{S}_k = \{j \in \mathbb{N} : k(L_T + l_T) - l_T + 1 \leq j \leq k(l_T + L_T)\}$ for $1 \leq k \leq K_T - 1$, and $\mathcal{S}_{K_T} = \{j \in \mathbb{N} : K_T(L_T + l_T) - l_T + 1 \leq j \leq \lfloor Tr \rfloor\}$. And note that $L_T \to \infty$ and $l_T = o(L_T)$.

Now, we make the following decomposition based on large and small blocks constructed above,

$$\begin{aligned}
&\sum_{t=\lfloor Tr_1 \rfloor + 2}^{\lfloor Tr_2 \rfloor} \boldsymbol{H}'_{(1)} \left(T^{-1}\boldsymbol{\nu} - \widehat{\nu}_1 \cdot \boldsymbol{I}_R\right) \boldsymbol{F}^0_{t-1} W_t \\
=& \sum_{k=K_1+1}^{K_2} \left[\sum_{j \in \mathcal{L}_k} \boldsymbol{H}'_{(1)} \left(T^{-1}\boldsymbol{\nu} - \widehat{\nu}_1 \cdot \boldsymbol{I}_R\right) \boldsymbol{F}^0_{j-1} W_j\right] + \sum_{k=K_1+1}^{K_2} \left[\sum_{j \in \mathcal{S}_k} \boldsymbol{H}'_{(1)} \left(T^{-1}\boldsymbol{\nu} - \widehat{\nu}_1 \cdot \boldsymbol{I}_R\right) \boldsymbol{F}^0_{j-1} W_j\right] \\
\equiv& \sum_{k=K_1+1}^{K_2} U_k + \sum_{k=K_1+1}^{K_2} V_k
\end{aligned}$$

Since $\{W_t\}$ are l_T-dependent, $\{U_k\}_{k=1}^{K_{T,r}}$ are independent random variables conditional on $\mathcal{X}$. The same property holds for $\{V_k\}_{k=1}^{K_{T,r}}$.

Below, we first handle with terms involving large blocks. By the C_r inequality,

$$\begin{aligned}
E^* \left| \sum_{t=\lfloor Tr_1 \rfloor + 2}^{\lfloor Tr_2 \rfloor} \boldsymbol{H}'_{(1)} \left(T^{-1}\boldsymbol{\nu} - \widehat{\nu}_1 \cdot \boldsymbol{I}_R\right) \boldsymbol{F}^0_{t-1} W_t \right|^4 &= E^* \left| \sum_{k=K_1+1}^{K_2} U_k + \sum_{k=K_1+1}^{K_2} V_k \right|^4 \\
&\leq 2^3 \left(E^* \left| \sum_{k=K_1+1}^{K_2} U_k \right|^4 + E^* \left| \sum_{k=K_1+1}^{K_2} V_k \right|^4 \right)
\end{aligned}$$

By construction, we have that $E^*(U_k) = 0$ and $E^*(V_k) = 0$ conditional on $\mathcal{X}$, for the general case that $0 \leq r_1 < r_2 \leq 1$, we can have

$$E^*\left|\sum_{k=K_1+1}^{K_2} U_k\right|^4 = \sum_{k=K_1+1}^{K_2} E^*\left(U_k^4\right) + \sum_{k\neq k'} E^*\left(U_k^2 U_{k'}^2\right) \leq \sum_{k=K_1+1}^{K_2} E^*\left(U_k^4\right) + \left\{\sum_{k=K_1+1}^{K_2} E^*\left(U_k^2\right)\right\}^2$$

and similarly for $E^*\left|\sum_{k=K_1}^{K_2} V_k\right|^4$. We follow arguments of (A.3) in Shao (2010), denote $\|A\|_p^* = [E^*\|A\|^p]^{1/p}$. Then, by Rosenthal inequality, we have $\|\sum_i A_i\|_4^* \leq M\left(\|\sum_i A_i\|_2^*\right)^{1/2}$. Conditional on $\mathcal{X}$, for $k = 1, \ldots, K_T$, we have

$$\begin{aligned}
\|U_k\|_4^* &\leq \sum_{g=1}^{2l_T}\left\|\sum_{j=1}^{\lfloor(L_T-g)/(2l_T)\rfloor} \boldsymbol{H}'_{(1)}\left(T^{-1}\boldsymbol{\nu} - \widehat{\nu}_1\cdot\boldsymbol{I}_R\right)\boldsymbol{F}_{j-1}^0 W_j\right\|_4^* \\
&\leq M\sum_{g=1}^{2l_T}\left\{\left\|\sum_{j=1}^{\lfloor(L_T-g)/(2l_T)\rfloor}\left(\boldsymbol{H}'_{(1)}\left(T^{-1}\boldsymbol{\nu} - \widehat{\nu}_1\cdot\boldsymbol{I}_R\right)\boldsymbol{F}_{g+2(j-1)l-1}^0\right)^2 W_{g+2(j-1)l}^2\right\|_2^*\right\}^{1/2} \\
&\leq M\sum_{g=1}^{2l_T}\left\{\sum_{j=1}^{\lfloor(L_T-g)/(2l_T)\rfloor}\left(\boldsymbol{H}'_{(1)}\left(T^{-1}\boldsymbol{\nu} - \widehat{\nu}_1\cdot\boldsymbol{I}_R\right)\boldsymbol{F}_{g+2(j-1)l-}^0\right)^2 \left\|W_{g+2(j-1)l_T}^2\right\|_2^*\right\}^{1/2} \\
&\leq M\sum_{g=1}^{2l_T}\left\{\sum_{j=1}^{\lfloor(L_T-g)/(2l_T)\rfloor}\left(\boldsymbol{H}'_{(1)}\left(T^{-1}\boldsymbol{\nu} - \widehat{\nu}_1\cdot\boldsymbol{I}_R\right)\boldsymbol{F}_{g+2(j-1)l-1}^0\right)^2\right\}^{1/2} \\
&\leq M\sqrt{l_T}\left\{\sum_{g=1}^{2l_T}\sum_{j=1}^{\lfloor(L_T-g)/(2l_T)\rfloor}\left(\boldsymbol{H}'_{(1)}\left(T^{-1}\boldsymbol{\nu} - \widehat{\nu}_1\cdot\boldsymbol{I}_R\right)\boldsymbol{F}_{g+2(j-1)l-1}^0\right)^2\right\}^{1/2} \\
&= M\sqrt{l_T}\left(\sum_{t\in\mathcal{L}_k}\left(\boldsymbol{H}'_{(1)}\left(T^{-1}\boldsymbol{\nu} - \widehat{\nu}_1\cdot\boldsymbol{I}_R\right)\boldsymbol{F}_{t-1}^0\right)^2\right)^{1/2},
\end{aligned} \tag{C.1.10}$$

where the fist line above is due to the triangle inequality, the second line above follows from Rosenthal inequality conditional on $\mathcal{X}$, the third and fourth lines hold by the basic properties of the norm $\|\cdot\|_p^*$ defined above, and the last inequality is due to the Cauchy–Schwarz inequality.

From above upper bound of $\|U_k\|_4^*$ for each k, it implies that

$$\begin{aligned} E^*\left(U_k^4\right) &\leq Ml_T^2\left(\sum_{t\in\mathcal{L}_k}\left(\boldsymbol{H}_{(1)}'\left(T^{-1}\boldsymbol{\nu}-\widehat{\nu}_1\cdot\boldsymbol{I}_R\right)\boldsymbol{F}_{t-1}^0\right)^2\right)^2 \\ &\leq Ml_T^2L_T\sum_{t\in\mathcal{L}_k}\left(\boldsymbol{H}_{(1)}'\left(T^{-1}\boldsymbol{\nu}-\widehat{\nu}_1\cdot\boldsymbol{I}_R\right)\boldsymbol{F}_{t-1}^0\right)^4. \end{aligned}$$

Then, because $\|\boldsymbol{H}\|=O_p(1)$, $\|T^{-1}\boldsymbol{\nu}\|=O_p\left(T^{-1}\right)$, $\widehat{\nu}_1=O_p\left(T^{-1}\right)$ by Theorem 3.3.2,

$$\begin{aligned} \sum_{k=K_1+1}^{K_2}E^*\left(U_k^4\right) &\leq Ml_T^2L_T\sum_{k=K_1+1}^{K_2}\sum_{t\in\mathcal{L}_k}\left(\boldsymbol{H}_{(1)}'\left(T^{-1}\boldsymbol{\nu}-\widehat{\nu}_1\cdot\boldsymbol{I}_R\right)\boldsymbol{F}_{t-1}^0\right)^4 \\ &\leq M\|\boldsymbol{H}\|^4\left\|T\left(T^{-1}\boldsymbol{\nu}-\widehat{\nu}_1\cdot\boldsymbol{I}_R\right)\right\|^4T^{-2}l_T^2L_T\sum_{k=K_1+1}^{K_2}\sum_{t\in\mathcal{L}_k}\left\|T^{-1/2}\boldsymbol{F}_{t-1}\right\|^4 \\ &= O_p(1)\,O_p\left(T^{-2}l_T^2L_T\right)O_p(T) \\ &= O_p\left(T^{-1}l_T^2L_T\right). \end{aligned}$$

where the third line above holds because $\sum_{k=K_1+1}^{K_2}\sum_{t\in\mathcal{L}_k}\left\|T^{-1/2}\boldsymbol{F}_{t-1}\right\|^4=O_p(T)$ holds that we can use Lemma C.3.2 1(a) to obtain. Besides, we have

$$\begin{aligned} &\sum_{k=K_1+1}^{K_2}E^*\left(U_k^2\right) \\ =\;&\sum_{k=K_1+1}^{K_2}\sum_{t\in\mathcal{L}_k}\sum_{h=-l_T}^{l_T}\left(\boldsymbol{H}_{(1)}'\left(T^{-1}\boldsymbol{\nu}-\widehat{\nu}_1\cdot\boldsymbol{I}_R\right)\boldsymbol{F}_{t-1}^0\right)\left(\boldsymbol{H}_{(1)}'\left(T^{-1}\boldsymbol{\nu}-\widehat{\nu}_1\cdot\boldsymbol{I}_R\right)\boldsymbol{F}_{t+h-1}^0\right)'E^*\left(W_tW_{t+h}\right) \\ \leq\;&\|\boldsymbol{H}\|^2\left\|T\left(T^{-1}\boldsymbol{\nu}-\widehat{\nu}_1\cdot\boldsymbol{I}_R\right)\right\|^2T^{-1}\sum_{k=K_1+1}^{K_2}\sum_{t\in\mathcal{L}_k}\sum_{h=-l_T}^{l_T}\left\|T^{-1/2}\boldsymbol{F}_{t-1}^0\right\|\left\|T^{-1/2}\boldsymbol{F}_{t+h-1}^0\right\|a\left(h/l_T\right) \\ =\;&O_p(1)\,T^{-1}O_p\left(l_TT\right)=O_p\left(l_T\right). \end{aligned}$$

where the first line holds by the properties of $\{W_t\}$ under the DWB scheme, and the last line follows by the boundness of $a(.)$ associated with $\{W_t\}$ under the DWB scheme and Lemma C.3.2 1(a).

Then, $\left\{\sum_{k=K_1+1}^{K_2}E^*\left(U_k^2\right)\right\}^2=O_p\left(l_T^2\right)$ holds by the above bound.

The same arguments work for terms involving small blocks $\{V_k\}$, and note that these terms are not leading terms by construction of large and small blocks. We omit the details here for brevity.

Then, collecting above immediate results together, we can obtain (C.1.8).

As argued in proofs of Lemma A.11 in Rho and Shao (2019), $\mathcal{I}_{12,r} = o_p^*(1)$ follows when (C.1.7) and (C.1.8) hold together.

Next, we are in the position to show $\mathcal{I}_{11,r} \Rightarrow W(r)$ below. We decompose $\mathcal{I}_{11,r}$ into large and small blocks again as done above for (C.1.8). Namely, we have

$$
\begin{aligned}
T^{-1/2}\sum_{t=2}^{\lfloor Tr \rfloor} \boldsymbol{H}'_{(1)}\boldsymbol{u}_t W_t &= T^{-1/2}\sum_{k=1}^{K_{T,r}}\left[\sum_{j\in\mathcal{L}_k}\boldsymbol{H}'_{(1)}\boldsymbol{u}_j W_j\right] + T^{-1/2}\sum_{k=1}^{K_{T,r}}\left[\sum_{j\in\mathcal{S}_k}\boldsymbol{H}'_{(1)}\boldsymbol{u}_j W_j\right] \\
&\equiv T^{-1/2}\sum_{k=1}^{K_{T,r}}\widetilde{U}_k + T^{-1/2}\sum_{k=1}^{K_{T,r}}\widetilde{V}_k \equiv \mathcal{I}_{11\mathcal{L},r} + \mathcal{I}_{11\mathcal{S},r}
\end{aligned}
$$

By construction of $\widetilde{U}_k$ and $\widetilde{V}_k$ above and $E^*(.)$ represents the expectation condition on $\mathcal{X}_n$, then $E^*\left(\widetilde{U}_k\right) = 0$ and $E^*\left(\widetilde{V}_k\right) = 0$ hold for all $k = 1, \ldots, K_{T,r}$; since $\{W_t\}$ are l_T-dependent, $\{\widetilde{U}_k\}_{k=1}^{K_T}$ and $\{\widetilde{V}_k\}_{k=1}^{K_T}$ are independent random variables conditional on $\mathcal{X}$.

We will finish the proof once we show that (i) The contribution of small blocks is negligible compared with that of large blocks, namely, $\mathcal{I}_{11\mathcal{S},r} = T^{-1/2}\sum_{k=1}^{K_{T,r}} V_k = o_p^*(1)$; and (ii) The functional central limit theorem will apply to the term consisting of large blocks, namely, $\mathcal{I}_{11\mathcal{L},r} = T^{-1/2}\sum_{k=1}^{K_{T,r}} U_k \Rightarrow \boldsymbol{H}'_{(1)}\Sigma_u^{1/2}\boldsymbol{W}(r)$.

Now, we are in the position to show $\mathcal{I}_{11\mathcal{S},r} = o_p^*(1)$. Note that

$$
\begin{aligned}
E\left\{E^*\left(\widetilde{V}_k^2\right)\right\} &= E\left[\sum_{j,j'\in\mathcal{S}_k} \boldsymbol{H}'_{(1)}\boldsymbol{u}_j\boldsymbol{u}'_{j'}\boldsymbol{H}_{(1)}\ a\left\{(j-j')/l_n\right\}\right] \\
\leq\ \|\boldsymbol{H}\|^2 \sum_{j,j'\in\mathcal{S}_k} [\alpha(|j-j'|)]^{\delta/(4+\delta)} &\left(E\|\boldsymbol{u}_j\|^{4+\delta} E\|\boldsymbol{u}_{j'}\|^{4+\delta}\right)^{2/(4+\delta)} a\left(\{j-j'\}/l_n\right) \\
\leq\ O(1)l_T \sum_{h=0}^{l_T-1} [\alpha(h)]^{\delta/(4+\delta)}\, a\left(h/l_n\right) &\leq Ml_T.
\end{aligned}
$$

where the second line above holds by Davydov's inequality under Assumption A3.1 for some $\delta > 0$, and the third line above holds under Assumption A3.1 and the fact that $\|\boldsymbol{H}\| = O_p(1)$. Then $\mathcal{I}_{11\mathcal{S},r} = T^{-1/2}\sum_{k=1}^{K_{T,r}} V_k = o_p^*(1)$ follows directly by using the above bound, the independence of $\{\widetilde{V}_k\}_{k=1}^{K_{T,r}}$ conditional on $\mathcal{X}$, and Chebyshev's inequality.

Finally, we will show $\mathcal{I}_{11\mathcal{L},r} \Rightarrow \boldsymbol{H}'_{(1)}\Sigma_u^{1/2}\boldsymbol{W}(r)$. To this end, we verify those conditions stated in Phillips (1988). First, we have $E^*\left(\widetilde{U}_k\right) = 0$ conditional on $\mathcal{X}$ by the construction $\widetilde{U}_k$. Second, note that $\mathcal{I}_{11\mathcal{L},r} = T^{-1/2}\sum_{k=1}^{K_{T,r}}\widetilde{U}_k \leq K_{T,r}^{-1/2}\sum_{k=1}^{K_{T,r}}\left(L_T^{-1/2}\widetilde{U}_k\right)$, by using (A.3) in Shao (2010) and Hölder's inequality, for some $\delta > 0$, we have

$$
\begin{aligned}
E^*\left|L_T^{-1/2}\widetilde{U}_k\right|^{4+\delta} &\leq ML_T^{-(4+\delta)/2} l_T^{(4+\delta)/2}\left(\sum_{t\in\mathcal{L}_k}\left|\boldsymbol{H}'_{(1)}\boldsymbol{u}_t\right|^2\right)^{(4+\delta)/2} \\
&\leq ML_T^{-(4+\delta)/2} l_T^{(4+\delta)/2}\|\boldsymbol{H}\|^{4+\delta} L_T^{\delta/2}\sum_{t\in\mathcal{L}_k}\|\boldsymbol{u}_t\|^{4+\delta} \\
&= O_p\left(L_T^{-1} l_T^{(4+\delta)/2}\right) = O_p\left(T^{-1} l_t^{(6+\delta)/2}\right) = o_p(1),
\end{aligned}
$$

where the last line is due to the fact that $L_T = \lfloor (T/l_T)^{1/2}\rfloor$, $l_T \asymp T^\eta$ with $\eta \in (0, 1/3)$, and $E\|\boldsymbol{u}_t\|^{4+\delta} = O_p(1)$ holds under Assumption A3.1. Third, $\{\widetilde{U}_k\}$ are independent of each other conditional on $\mathcal{X}_1$. Therefore, based on such three facts, in the view of results in Phillips (1988), the desired result follows directly.

(b) The desired result follows directly by taking advantages of the argument of proofs in (a) above and the simple facts for the joint test statistic like (C.1.2) and (C.1.3).

C.2 Proofs for Technical Lemmas

This section is composed of 3 parts. Section C.2.1 contains the proofs of Lemmas A.1-A.5 in the above paper. Section C.2.2 contains some technical lemmas that are used in the proofs of the main results and Lemmas C.1-C.5. Section C.2.3 provides some additional simulation results.

C.2.1 Proofs of Lemma C.1.1

The proof is similar to that of Proposition 1 in Bai (2003). We omit details here for brevity.

C.2.2 Some preliminary results

In this subsection, we present the bound of some terms frequently used latter in the proofs of technical lemmas. Under Assumption A3.1 to A3.5, we can obtain the desired bound by following lines in Feng and Su (2020) together with Lemma C.3.2 and thus omit the details of proofs for the brevity.

$$
\begin{aligned}
&(1)\quad \left\|\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}\right\| = O_p\left(N^{1/2}T^{1/2}\right), \\
&(2)\quad \left|N^{-1}\sum_{i=1}^{N} c_i e_{it-1}\right| = O_p\left(N^{-1/2}T\right) \text{ holds for } \quad t = 2,\ldots,T, \\
&(3)\quad T^{-1}\sum_{s=2}^{T}\left|N^{-1}\sum_{i=1}^{N}\boldsymbol{B}_s^{0\prime}\boldsymbol{\lambda}_i\epsilon_{it}\right|^2 = O_p\left(N^{-1}\right) \text{ holds for } \quad t = 1,\ldots,T, \\
&(4)\quad \left\|N^{-1}T^{-1}\sum_{j=1}^{N}\sum_{s=2}^{T}\boldsymbol{B}_s^{0}\epsilon_{js}\boldsymbol{\lambda}_j^{0\prime}\right\| = O_p\left(N^{-1/2}T^{-1/2}\right), \\
&(5)\quad \left\|T^{-1}\sum_{t=2}^{T}\boldsymbol{B}_t^{0}\epsilon_{it}\right\| = O_p\left(T^{-1/2}\right) \text{ holds for each } i, \\
&(6)\quad \left\|N^{-1}T^{-1}\sum_{j=1}^{N}\sum_{t=1}^{T}\boldsymbol{\lambda}_j^{0}\epsilon_{jt}\epsilon_{it}\right\| = O_p\left(N^{-1/2}T^{-1/2}\right), \\
&(7)\quad \left\|\boldsymbol{e}_{-1}'\left(\boldsymbol{\rho}_N - \boldsymbol{I}_N\right)\right\| = O_p\left(N^{1/2}\right), \text{ where } \boldsymbol{\rho}_N \equiv diag\left(1 + c_1/T,\ldots,1 + c_N/T\right), \\
&(8)\quad \left\|\boldsymbol{\Lambda}^{0\prime}\left(\boldsymbol{\rho}_N - \boldsymbol{I}_N\right)\boldsymbol{e}_{-1}\right\| = O_p\left(N^{1/2}\right), \\
&(9)\quad \sum_{s=2}^{T}\left\|N^{-1}\sum_{i=1}^{N}\left(\rho_i - 1\right)^2 e_{is-1}e_{it-1}\right\|^2 = O_p\left(T^{-1}\right) \text{ holds for } \quad t = 2,\ldots,T.
\end{aligned}
$$

C.2.3 Proofs of Lemma C.1.2

In this proof, we focus on the case that both common factors and idiosyncratic errors follow the local-to-unit processes from the explosive side. Namely, we will take the local-to-unit-root alternatives with $\rho_{0,r}^0 = 1 + \nu_r/T$ for $r = 1, ..., R_0$ with $\nu_r > 0$ for all r; and $\rho_{0,i}^0 = 1 + c_i/T$ for $i = 1, ..., N$ with $c_i > 0$ for all i. For other cases, we have similar arguments for proofs in the context of the alternative of local-to-unity.

Proofs of Lemma C.1.2(a)

Let $\boldsymbol{W} = (\boldsymbol{\rho}_N - \boldsymbol{I}_N)\,\boldsymbol{e}_{-1} + \boldsymbol{\epsilon}$ and note that (3.2 .10) can be written in matrix form as follows

$$\boldsymbol{Z} = \boldsymbol{\Lambda}^0\left(\boldsymbol{F}^0 - \boldsymbol{F}^0_{-1}\right)' + \boldsymbol{W} \equiv \boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime} + \boldsymbol{W} \tag{C.2.1}$$

Recall that $\Sigma_{Z,N} = N^{-1}\boldsymbol{Z}'\boldsymbol{Z}$, by the eigenvalue problem, we have the identity

$$\Sigma_{Z,N}\widehat{\boldsymbol{B}} = \widehat{\boldsymbol{B}}\boldsymbol{V}_{\boldsymbol{Z},\boldsymbol{N}}, \tag{C.2.2}$$

where $\boldsymbol{V}_{\boldsymbol{Z},\boldsymbol{N}}$ is a diagonal matrix that consists of the first R_0 eigenvalues of $\Sigma_{Z,N}$ arranged in descending order along its main diagonal line. Premultiplying both sides of (C.2.2) by the recalling factor T^{-1} and substituting (C.2.1) into (C.2.2), we have

$$\begin{aligned} & T^{-1}\widehat{\boldsymbol{B}}\boldsymbol{V}_{Z,N} - \boldsymbol{B}^0\left(N^{-1}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\right)\left(T^{-1}\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\right) \\ = \;& N^{-1}T^{-1}\boldsymbol{W}'\boldsymbol{W}\widehat{\boldsymbol{B}} + N^{-1}T^{-1}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{W}\widehat{\boldsymbol{B}} + N^{-1}T^{-1}\boldsymbol{W}'\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}} \\ \equiv \;& A_1 + A_2 + A_3. \end{aligned} \tag{C.2.3}$$

Below, we bound each term on the last line. First, for $A_1 = N^{-1}T^{-1}\boldsymbol{W}'\boldsymbol{W}\widehat{\boldsymbol{B}}$, recall that $\rho_i = 1 + c_i/T$ for $i = 1, \ldots, N$ under Assumption A3.5 such that $\boldsymbol{\rho}_N \equiv diag\,(1 + c_1/T, \ldots, 1 + c_N/T)$, we have

$$\begin{aligned} A_1 \;=\;& N^{-1}T^{-1}\boldsymbol{e}'_{-1}(\boldsymbol{\rho}_N - \boldsymbol{I}_N)^2\,\boldsymbol{e}_{-1}\widehat{\boldsymbol{B}} + N^{-1}T^{-1}\boldsymbol{\epsilon}'\boldsymbol{\epsilon}\widehat{\boldsymbol{B}} + N^{-1}T^{-1}\boldsymbol{e}'_{-1}(\boldsymbol{\rho}_N - \boldsymbol{I}_N)\,\boldsymbol{\epsilon} \\ & + N^{-1}T^{-1}\boldsymbol{\epsilon}'(\boldsymbol{\rho}_N - \boldsymbol{I}_N)\,\boldsymbol{e}_{-1}\widehat{\boldsymbol{B}} \\ \equiv\;& A_{11} + A_{12} + A_{13} + A_{14}. \end{aligned}$$

For A_{11}, we have by the sub-multiplicity of Frobenius norm,

$$\begin{aligned}\|A_{11}\| &= N^{-1}T^{-1}\left\|\boldsymbol{e}_{-1}'(\boldsymbol{\rho}_N-\boldsymbol{I}_N)^2\boldsymbol{e}_{-1}\widehat{\boldsymbol{B}}\right\| \\ &\leq N^{-1}T^{-1/2}\left\|\boldsymbol{e}_{-1}'(\boldsymbol{\rho}_N-\boldsymbol{I}_N)\right\|^2\left\|T^{-1/2}\widehat{\boldsymbol{B}}\right\| \\ &\leq N^{-1}T^{-1/2}O_p(N)\,O_p(1)=O_p\left(T^{-1/2}\right),\end{aligned}$$

where the third line above follows by the normalization condition and the bound given in Section C.2.2. Similarly, we have

$$\|A_{12}\|\leq N^{-1}T^{-1/2}\|\boldsymbol{\epsilon}\|_{sp}^2\left\|T^{-1/2}\widehat{\boldsymbol{B}}\right\|=N^{-1}T^{-1/2}O_p(N+T)\,O_p(1)=O_p\left(T^{-1/2}+N^{-1}T^{1/2}\right)$$

where we also use the that $\|\boldsymbol{\epsilon}\|_{sp}^2=O_p(N+T)$ under Assumption A3.3(d). By Cauchy-Schwarz inequality, it is easy to see $\|A_{13}\|$ and $\|A_{14}\|$ will not be the dominant term in comparison with $\|A_{11}\|$ and $\|A_{12}\|$. In sum, we have shown that $\|A_1\|=O_p\left(T^{-1/2}+N^{-1}T^{1/2}\right)$.

Next, we study A_2. Note that $A_2=N^{-1}T^{-1}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}(\boldsymbol{\rho}_N-\boldsymbol{I}_N)\,\boldsymbol{e}_{-1}\widehat{\boldsymbol{B}}+N^{-1}T^{-1}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}\widehat{\boldsymbol{B}}\equiv A_{21}+A_{22}$. For A_{21}, we have

$$\begin{aligned}A_{21} &\leq N^{-1}T^{-1}\left\|\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}(\boldsymbol{\rho}_N-\boldsymbol{I}_N)\,\boldsymbol{e}_{-1}\widehat{\boldsymbol{B}}\right\| \\ &\leq N^{-1}\left\|T^{-1/2}\boldsymbol{B}^0\right\|\left\|\boldsymbol{\Lambda}^{0\prime}(\boldsymbol{\rho}_N-\boldsymbol{I}_N)\,\boldsymbol{e}_{-1}\right\|\left\|T^{-1/2}\widehat{\boldsymbol{B}}\right\| \\ &\leq N^{-1}O_p(1)\,O_p\left(N^{1/2}\right)O_p(1)=O_p\left(N^{-1/2}\right),\end{aligned}$$

where the third line above follows by Lemma C.3.2 1(e), the normalization condition and the bound shown in Section C.2.2. Similarly, we have

$$\begin{aligned}\|A_{22}\| &\leq N^{-1}\left\|T^{-1/2}\boldsymbol{B}^0\right\|\left\|\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}\right\|\left\|T^{-1/2}\widehat{\boldsymbol{B}}\right\| \\ &= N^{-1}O_p(1)\,O_p\left(N^{1/2}T^{1/2}\right)O_p(1)=O_p\left(N^{-1/2}T^{1/2}\right),\end{aligned}$$

where the second line above follows by Lemma C.3.2 1(e), the normalization condition and the bound shown in Section C.2.2. In sum, we have $\|A_2\| = O_p\left(N^{-1/2} + N^{-1/2}T^{1/2}\right)$.

Next, note that $A_3 = N^{-1}T^{-1}\boldsymbol{e}'_{-1}\left(\boldsymbol{\rho}_N - \boldsymbol{I}_N\right)\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}} + N^{-1}T^{-1}\boldsymbol{\epsilon}'\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}} \equiv A_{31} + A_{32}$. By the bound shown in Section C.2.2 and Lemma C.1.1, we have

$$\|A_{31}\| \leq N^{-1}\left\|\boldsymbol{e}'_{-1}\left(\boldsymbol{\rho}_N - \boldsymbol{I}_N\right)\boldsymbol{\Lambda}^0\right\|\left\|T^{-1}\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\right\| = O_p\left(N^{-1/2}\right),$$

and

$$\|A_{32}\| \leq N^{-1/2}T^{1/2}\left\|\boldsymbol{\epsilon}'\boldsymbol{\Lambda}^0\right\|\left\|T^{-1}\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\right\| = N^{-1}O_p\left(N^{1/2}T^{1/2}\right)O_p\left(1\right) = O_p\left(N^{-1/2}T^{1/2}\right).$$

We then have $\|A_3\| = O_p\left(N^{-1/2} + N^{-1/2}T^{1/2}\right)$. Then, collecting all above immediate results, we have

$$\|A_1\| + \|A_2\| + \|A_3\| = O_p\left(N^{-1/2} + T^{-1/2} + N^{-1}T^{1/2} + N^{-1/2}T^{1/2}\right).$$

Then, it follows that

$$\begin{aligned} & \left\|\widehat{\boldsymbol{B}}\left(T^{-1}\boldsymbol{V}_{Z,N}\right) - \boldsymbol{B}^0\left(N^{-1}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\right)\left(T^{-1}\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\right)\right\| \\ \leq\ & \|A_1\| + \|A_2\| + \|A_3\| \\ =\ & O_p\left(N^{-1/2} + T^{-1/2} + N^{-1}T^{1/2} + N^{-1/2}T^{1/2}\right). \end{aligned}$$

Consequently, we have

$$\begin{aligned} & T^{-1}\left\|\widehat{\boldsymbol{B}}\left(T^{-1}\boldsymbol{V}_{Z,N}\right) - \boldsymbol{B}^0\left(N^{-1}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\right)\left(T^{-1}\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\right)\right\|^2 \\ =\ & T^{-1}\sum_{t=2}^{T}\left\|\left(T^{-1}\boldsymbol{V}_{Z,N}\right)\widehat{\boldsymbol{B}}_t - \left(T^{-1}\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\right)'\left(N^{-1}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\right)\boldsymbol{B}^0_t\right\|^2 = O_p\left(N^{-1}\right). \end{aligned} \tag{C.2.4}$$

Now we are in a position to define the rotation matrix $\boldsymbol{H}$. Following Bai and Ng (2004), let $\boldsymbol{H}_B \equiv \left(N^{-1}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\right)\left(T^{-1}\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\right)$. Define

$$\left(\boldsymbol{H}^{-1}\right)' \equiv \left(T^{-1}\boldsymbol{V}_{Z,N}\right)\boldsymbol{H}_B^{-1} \tag{C.2.5}$$

According to Lemma C.3.3 , and the fact that $\|\boldsymbol{H}_B\| = O_p(1)$ and $\left\|\boldsymbol{H}_B^{-1}\right\| = O_p(1)$ by constructions using Assumption A3.2 and Lemma C.1.1, we can readily show that

$$\left\|\boldsymbol{H}^{-1}\right\| = O_p(1), \quad \|\boldsymbol{H}\| = O_p(1). \tag{C.2.6}$$

Then, by (C.2.5) and (C.2.5), we have

$$T^{-1}\left\|\widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1} - \boldsymbol{B}^0\right\|^2 = O_p\left(N^{-1}\right). \tag{C.2.7}$$

The desired result follows directly.

Proofs of Lemma C.1.2(b)

Using (C.2.2), (C.2.5), and Lemma C.3.3, and the fact that $\boldsymbol{H}_B = \left(N^{-1}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\right)\left(T^{-1}\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\right)$, for each t, we have

$$\begin{aligned}
&\left\{T^{-1}\boldsymbol{V}_{Z,N}\right\}'\widehat{\boldsymbol{B}}_t - \left[\left(N^{-1}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\right)\left(T^{-1}\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\right)\right]'\boldsymbol{B}_t^0 \\
=\;& \left\{T^{-1}\boldsymbol{V}_{Z,N}\right\}'\widehat{\boldsymbol{B}}_t - \boldsymbol{H}_B'\boldsymbol{B}_t^0 \\
=\;& \left(N^{-1}T^{-1}\right)\widehat{\boldsymbol{B}}'\boldsymbol{W}'\boldsymbol{W}_t + \left(N^{-1}T^{-1}\right)\widehat{\boldsymbol{B}}'\boldsymbol{W}'\boldsymbol{\Lambda}^0\boldsymbol{B}_t^0 + \left(N^{-1}T^{-1}\right)\widehat{\boldsymbol{B}}'\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{W}_t \\
\equiv\;& \left(N^{-1}T^{-1}\right)\left(\mathbb{A}_{1t} + \mathbb{A}_{2t} + \mathbb{A}_{3t}\right),
\end{aligned} \tag{C.2.8}$$

We first study $\mathbb{A}_{1t}$, recall that $\boldsymbol{W} = (\boldsymbol{\rho}_N - \boldsymbol{I}_N)\boldsymbol{e}_{-1} + \boldsymbol{\epsilon}$ and thus $\boldsymbol{W}_t = (\boldsymbol{\rho}_N - \boldsymbol{I}_N)\boldsymbol{e}_{t-1} + \boldsymbol{\epsilon}_t$, we use these facts to make the following decomposition:

$$\begin{aligned}\mathbb{A}_{1t} &= \widehat{\boldsymbol{B}}'\boldsymbol{\epsilon}'\boldsymbol{\epsilon}_t + \widehat{\boldsymbol{B}}'\boldsymbol{e}'_{-1}\left(\boldsymbol{\rho}_N - \boldsymbol{I}_N\right)^2 \boldsymbol{e}_{t-1} + \widehat{\boldsymbol{B}}'\boldsymbol{e}'_{-1}\left(\boldsymbol{\rho}_N - \boldsymbol{I}_N\right)\boldsymbol{\epsilon}_t + \widehat{\boldsymbol{B}}'\boldsymbol{\epsilon}'\left(\boldsymbol{\rho}_N - \boldsymbol{I}_N\right)\boldsymbol{e}_{t-1} \\ &\equiv \mathbb{A}_{1t1} + \mathbb{A}_{1t2} + \mathbb{A}_{1t3} + \mathbb{A}_{1t4}.\end{aligned}$$

Noting that $\widehat{\boldsymbol{B}} = (\widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1} - \boldsymbol{B}^0)\boldsymbol{H}' + \boldsymbol{B}^0\boldsymbol{H}'$, we make further decomposition for $\mathbb{A}_{1t1}$:

$$\mathbb{A}_{1t1} = \boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{\epsilon}'\boldsymbol{\epsilon}_t + \boldsymbol{H}\left(\widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1} - \boldsymbol{B}^0\right)'\boldsymbol{\epsilon}'\boldsymbol{\epsilon}_t = \mathbb{A}_{1t1a} + \mathbb{A}_{1t1b}.$$

Recall that $\gamma_N(s,t) = \sum_{i=1}^N E\left(\epsilon_{is}\epsilon_{it}\right)$, we make further decomposition for $\mathbb{A}_{1t1a}$:

$$\mathbb{A}_{1t1a} = \boldsymbol{H}\sum_{s=2}^T \boldsymbol{B}_s^0 \gamma_N(s,t) + \boldsymbol{H}\sum_{s=2}^T \boldsymbol{B}_s^0 \left(\sum_{i=1}^N \epsilon_{is}\epsilon_{it} - \gamma_N(s,t)\right) \equiv \mathbb{A}_{1t1a1} + \mathbb{A}_{1t1a2}.$$

Note that $\sum_{s=2}^T \left|\gamma_N(s,t)\right|^2 \le \left(\max_{s,t}\left|\gamma_N(s,t)\right|\right)\sum_{s=2}^T \left|\gamma_N(s,t)\right| = O(N^2)$ by Assumption A3.3(b). In addition, $\sum_{s=2}^T \left\|\boldsymbol{B}_s^0\right\|^2 = O_p(T)$ by Lemma C.3.2(e). Then by the Cauchy-Schwarz inequality,

$$\left\|\mathbb{A}_{1t1a1}\right\| = \left\|\boldsymbol{H}\right\|\left\|\sum_{s=2}^T \boldsymbol{B}_s^0\gamma_N(s,t)\right\| \le \left\|\boldsymbol{H}\right\|\left[\sum_{s=2}^T \left\|\boldsymbol{B}_s^0\right\|^2\right]^{1/2}\left[\sum_{s=2}^T \left|\gamma_N(s,t)\right|^2\right]^{1/2} = O_p\left(NT^{1/2}\right).$$

For $\mathbb{A}_{1t1a2}$, we have by the Cauchy-Schwarz inequality

$$\left\|\mathbb{A}_{1t1a2}\right\| \le \left\|\boldsymbol{H}\right\|\left[\sum_{s=1}^T \left\|\boldsymbol{B}_s^0\right\|^2\right]^{1/2}\left[\sum_{s=1}^T\left|\sum_{i=1}^N\left[\epsilon_{is}\epsilon_{it} - E\left(\epsilon_{is}\epsilon_{it}\right)\right]\right|^2\right]^{1/2} = O_p\left(N^{1/2}T\right),$$

where we use the fact that $\left\|\boldsymbol{H}\right\| = O_p(1)$, $\sum_{s=2}^T \left\|\boldsymbol{B}_s^0\right\|^2 = O_p(T)$ stated in Lemma C.3.2 1(e) and Assumption A3.3(b).

Now, write $\mathbb{A}_{1t1b} = \sum_{s=1}^T(\widehat{\boldsymbol{B}}_s - \boldsymbol{H}\boldsymbol{B}_s^0)\gamma_N(s,t) + \sum_{s=1}^T(\widehat{\boldsymbol{B}}_s - \boldsymbol{H}\boldsymbol{B}_s^0)[\sum_{i=1}^N \epsilon_{is}\epsilon_{it} - \gamma_N(s,t)]$

$\equiv \mathbb{A}_{1t1b1} + \mathbb{A}_{1t1b2}$. Because of the result in (C.2.6) and (C.2.7), we also have

$$\left\|\widehat{\boldsymbol{B}} - \boldsymbol{B}^0\boldsymbol{H}'\right\| \leq \left\|\widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1} - \boldsymbol{B}^0\right\| \|\boldsymbol{H}\| = O_p\left(T^{1/2}N^{-1/2}\right). \tag{C.2.9}$$

Then, using Cauchy-Schwarz inequality, (C.2.9) above and Assumption A3.3(a), we have

$$\|\mathbb{A}_{1t1b1}\| \leq \left\|\widehat{\boldsymbol{B}} - \boldsymbol{B}^0\boldsymbol{H}'\right\| \left[\sum_{s=1}^{T} |\gamma_N(s,t)|^2\right]^{1/2} = O_p\left(T^{1/2}N^{-1/2}\right) O_p(N) = O_p\left(N^{1/2}T^{1/2}\right).$$

Similarly, by Cauchy-Schwarz inequality, (C.2.9) above, and Assumption A3.3(b) we have

$$\begin{aligned} \|\mathbb{A}_{1t1b2}\| &\leq \left\|\widehat{\boldsymbol{B}} - \boldsymbol{B}^0\boldsymbol{H}'\right\| \left[\sum_{s=1}^{T}\left|\sum_{i=1}^{N}\left[\epsilon_{is}\epsilon_{it} - E\left(\epsilon_{is}\epsilon_{it}\right)\right]\right|^2\right]^{1/2} \\ &= O_p\left(T^{1/2}N^{-1/2}\right) O_p\left(N^{1/2}T^{1/2}\right) = O_p(T). \end{aligned}$$

In sum, we have

$$\|\mathbb{A}_{1t1}\| = O_p\left(NT^{1/2} + N^{1/2}T + N^{1/2}T^{1/2} + T\right).$$

For $\mathbb{A}_{1t2}$, we have $\mathbb{A}_{1t2} = \boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{e}'_{-1}(\boldsymbol{\rho}_N - \boldsymbol{I}_N)^2 \boldsymbol{e}_{t-1} + \boldsymbol{H}(\widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1} - \boldsymbol{B}^0)'\boldsymbol{e}'_{-1}(\boldsymbol{\rho}_N - \boldsymbol{I}_N)^2 \boldsymbol{e}_{t-1} \equiv \mathbb{A}_{1t2a} + \mathbb{A}_{1t2b}$. For $\mathbb{A}_{1t2a}$, by expansions, we have

$$\mathbb{A}_{1t2a} = \boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{e}'_{-1}(\boldsymbol{\rho}_N - \boldsymbol{I}_N)\left(T^{-1}\sum_{i=1}^{N} c_i e_{it-1}\right).$$

It follows that

$$\begin{aligned} \|\mathbb{A}_{1t2a}\| &\leq \|\boldsymbol{H}\| \left\|\boldsymbol{B}^{0\prime}\right\| \left\|\boldsymbol{e}'_{-1}(\boldsymbol{\rho}_N - \boldsymbol{I}_N)\right\| \left|T^{-1}\sum_{i=1}^{N} c_i e_{it-1}\right| \\ &= O_p(1)\, O_p\left(T^{1/2}\right) O_p\left(N^{1/2}\right) O_p\left(N^{1/2}\right) \\ &= O_p\left(NT^{1/2}\right), \end{aligned}$$

where the second line above holds by following bounds: $\|\boldsymbol{H}\| = O_p(1)$ as stated in (C.2.6), $\|\boldsymbol{B}^0\| = O_p\left(T^{1/2}\right)$ by Lemma C.3.2 1(e), $T^{-1}\left|\sum_{i=1}^N c_i e_{it-1}\right| = O_p\left(N^{1/2}\right)$ holds as shown in C.2.2 by the weak cross-sectional dependence imposed in Assumption A3.1 and the fact that $e_{it}/T = O_p(1)$ for each i. For $\mathbb{A}_{1t2b}$, we have

$$
\begin{aligned}
\|\mathbb{A}_{1t2b}\| &= \left\|\sum_{s=2}^T \left(\widehat{\boldsymbol{B}}_s - \boldsymbol{H}\boldsymbol{B}_s^0\right)\left(\sum_{i=1}^N \left(\rho_i^0 - 1\right)^2 e_{is-1}e_{it-1}\right)\right\| \\
&\leq \left\|\widehat{\boldsymbol{B}} - \boldsymbol{B}^0\boldsymbol{H}'\right\| \left[\sum_{s=2}^T \left\|\sum_{i=1}^N \left(\rho_i^0 - 1\right)^2 e_{is-1}e_{it-1}\right\|^2\right]^{1/2} \\
&= O_p\left(T^{1/2}N^{-1/2}\right) O_p\left(NT^{-1/2}\right) = O_p\left(N^{1/2}\right),
\end{aligned}
$$

where the second line above follows because we use (C.2.9) and the result shown in Section C.2.2. In sum, we have

$$
\|\mathbb{A}_{1t2}\| = O_p\left(NT^{1/2}\right).
$$

Collecting the above results for $\mathbb{A}_{1tl}, l = 1, 2$, $\|\mathbb{A}_{1t}\| = O_p\left(N^{1/2}T + NT^{1/2} + N^{1/2}T^{1/2} + T\right)$ follows directly.

Now, we study $\mathbb{A}_{2t}$. Note that $\mathbb{A}_{2t} = \boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{W}'\boldsymbol{\Lambda}^0\boldsymbol{B}_t^0 + (\widehat{\boldsymbol{B}} - \boldsymbol{B}^0\boldsymbol{H}')'\boldsymbol{W}'\boldsymbol{\Lambda}^0\boldsymbol{B}_t^0 \equiv \mathbb{A}_{2t1} + \mathbb{A}_{2t2}$. In particular, let $\mathbb{A}_{2t1} = \boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{W}'\boldsymbol{\Lambda}^0\boldsymbol{B}_t^0 \equiv \boldsymbol{H}\bar{\mathbb{A}}_{2t1}\boldsymbol{B}_t^0$, where $\bar{\mathbb{A}}_{2t1} = \boldsymbol{B}^{0\prime}\boldsymbol{W}'\boldsymbol{\Lambda}^0 = \boldsymbol{B}^{0\prime}\boldsymbol{e}_{-1}'\left(\boldsymbol{\rho}_N - \boldsymbol{I}_N\right)\boldsymbol{\Lambda}^0 + \boldsymbol{B}^{0\prime}\boldsymbol{\epsilon}'\boldsymbol{\Lambda}^0 \equiv \bar{\mathbb{A}}_{2t1a} + \bar{\mathbb{A}}_{2t1b}$. So $\bar{\mathbb{A}}_{2t1b} = \sum_{i=1}^N\sum_{s=2}^T \boldsymbol{B}_s^0 e_{is-1}\boldsymbol{\lambda}_i^{0\prime}$ and $\bar{\mathbb{A}}_{2t1b} = \sum_{i=1}^N\sum_{s=2}^T \boldsymbol{B}_s^0 \epsilon_{is}\boldsymbol{\lambda}_i^{0\prime}$. Note that $E\bar{\mathbb{A}}_{2t1b} = 0$ under Assumptions A3.1 and A3.4; recall that $\boldsymbol{B}_t^0 = T^{-1}\boldsymbol{\nu}\boldsymbol{F}_{t-1}^0 + u_t$ for all t, and then,

$$
\begin{aligned}
& E\left\|\bar{\mathbb{A}}_{2t1b}\right\|^2 = E\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=2}^{T}\sum_{q=2}^{T}\boldsymbol{B}_s^0\epsilon_{is}\boldsymbol{\lambda}_i^{0\prime}\boldsymbol{\lambda}_j^0\epsilon_{jq}\boldsymbol{B}_q^{0\prime}\right) \\
= & \sum_{i=1}^{N}\sum_{j=1}^{N}E\left(\boldsymbol{\lambda}_i^{0\prime}\boldsymbol{\lambda}_j^0\right)\sum_{s=2}^{T}\sum_{q=2}^{T}E\left(\epsilon_{is}\epsilon_{jq}\right)E\left(\boldsymbol{B}_q^{0\prime}\boldsymbol{B}_s^0\right) \\
= & \sum_{i=1}^{N}\sum_{j=1}^{N}E\left(\boldsymbol{\lambda}_i^{0\prime}\boldsymbol{\lambda}_j^0\right) \\
& \times\left\{\sum_{s=2}^{T}\sum_{q=2}^{T}E\left(\epsilon_{is}\epsilon_{jq}\right)E\left[T^{-2}\boldsymbol{F}_{q-1}^0\boldsymbol{\nu}'\boldsymbol{\nu}\boldsymbol{F}_{s-1}^0 + \boldsymbol{u}_q'\boldsymbol{u}_s + T^{-1}\boldsymbol{F}_{q-1}^0\boldsymbol{\nu}'\boldsymbol{u}_s + T^{-1}\boldsymbol{u}_q'\boldsymbol{\nu}\boldsymbol{F}_{s-1}^0\right]\right\} \\
\equiv & \bar{\mathbb{A}}_{2t1b1} + \bar{\mathbb{A}}_{2t1b2} + \bar{\mathbb{A}}_{2t1b3} + \bar{\mathbb{A}}_{2t1b4} \qquad \text{(C.2.10)}
\end{aligned}
$$

where the second line holds under Assumption A3.4, and the $\{\bar{\mathbb{A}}_{2t1bh}\}_{h=1}^4$ are defined implicitly as above.

For $\bar{\mathbb{A}}_{2t1b1} = T^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left(\boldsymbol{\lambda}_i^{0\prime}\boldsymbol{\lambda}_j^0\right)\sum_{s=2}^{T}\sum_{q=2}^{T}E\left(\epsilon_{is}\epsilon_{jq}\right)E\left(T^{-2}\boldsymbol{F}_{q-1}^0\boldsymbol{\nu}'\boldsymbol{\nu}\boldsymbol{F}_{s-1}^0\right)$, it follows that

$$
\begin{aligned}
\bar{\mathbb{A}}_{2t1b1} & \leq M\left\|\boldsymbol{\nu}\right\|_{sp}T^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=2}^{T}\sum_{q=2}^{T}\alpha_{ij}\left(|s-q|\right)E\left(\boldsymbol{F}_{q-1}^0\boldsymbol{F}_{s-1}^0\right) \\
& \leq MT^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_{ij}\left(0\right)E\left(\sum_{s=2}^{T}\sum_{q=2}^{T}\boldsymbol{F}_{q-1}^0\boldsymbol{F}_{s-1}^0\right) \\
& = MT^{-2}O\left(N\right)O\left(T^3\right) = O\left(NT\right),
\end{aligned}
$$

where the first line holds by the Davydov's inequality under Assumptions A3.1 and A3.2, the second line follows from Assumption A3.1 too, and the third line follows from Assumption A3.1 and Lemma C.3.2 1(e). Similarly, for $\bar{\mathbb{A}}_{2t1b2}$, we have

$$
\begin{aligned}
\bar{\mathbb{A}}_{2t1b1} & \leq MT^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=2}^{T}\sum_{q=2}^{T}\alpha_{ij}\left(|s-q|\right)\alpha\left(|s-q|\right) \\
& \leq M\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=2}^{T}\sum_{q=2}^{T}\alpha_{ij}\left(|s-q|\right) = O\left(NT\right),
\end{aligned}
$$

where the first line holds by the Davydov's inequality under Assumptions A3.1 and A3.2, the second line follows from the property of mixing coefficients, and the third line follows from Assumption A3.1. Similarly, by Cauchy-Schwarz inequality, we can readily show $\bar{\mathbb{A}}_{2t1b3} = O(NT)$ and $\bar{\mathbb{A}}_{2t1b4} = O(NT)$. Then, by Chebyshev's inequality, it follows that $\bar{\mathbb{A}}_{2t1b} = O_p\left(N^{1/2}T^{1/2}\right)$. For $\bar{\mathbb{A}}_{2t1a}$,

$$\begin{aligned}\bar{\mathbb{A}}_{2t1a} &\leq \left\|\boldsymbol{B}^{0\prime}\boldsymbol{e}_{-1}^{\prime}\left(\boldsymbol{\rho}_N-\boldsymbol{I}_N\right)\boldsymbol{\Lambda}^0\right\| \\ &\leq \left\|\boldsymbol{B}^{0\prime}\right\|\left\|\boldsymbol{e}_{-1}^{\prime}\left(\boldsymbol{\rho}_N-\boldsymbol{I}_N\right)\boldsymbol{\Lambda}^0\right\| \\ &= O_p\left(N^{1/2}\left\|\boldsymbol{B}^0\right\|\right)=O_p\left(N^{1/2}T^{1/2}\right),\end{aligned}$$

where the last line follows by the results given in Section C.2.2 and Lemma C.3.2 1(e). In sum, we have

$$\begin{aligned}\left\|\mathbb{A}_{2t1}\right\| &\leq \|\boldsymbol{H}\|\left\|\bar{\mathbb{A}}_{2t1}\right\|\left\|\boldsymbol{B}^0\right\| \\ &= O_p(1)\,O_p\left(N^{1/2}\left\|\boldsymbol{B}^0\right\|\right)O_p\left(\left\|\boldsymbol{B}^0\right\|\right)=O_p\left(N^{1/2}T\right),\end{aligned}$$

where we use the fact that $\|\boldsymbol{B}_t^0\| \leq \|\boldsymbol{B}^0\| = O_p\left(T^{1/2}\right)$ by Lemma C.3.2 1(e) under Assumption A3.1 and A3.5. Similarly, we have

$$\begin{aligned}\left\|\mathbb{A}_{2t2}\right\| &\leq \left\|\widehat{\boldsymbol{B}}-\boldsymbol{B}^0\boldsymbol{H}^{\prime}\right\|\left[\sum_{s=1}^{T}\left|\sum_{i=1}^{N}\epsilon_{is}\boldsymbol{\lambda}_i\right|^2\right]^{1/2}\left\|\boldsymbol{B}^0\right\| \\ &= O_p\left(T^{1/2}N^{-1/2}\right)O_p\left(N^{1/2}T^{1/2}\right)O_p\left(\left\|\boldsymbol{B}^0\right\|\right) \\ &= O_p\left(T^{3/2}\right),\end{aligned}$$

where the second line follows by (C.2.9), the result given in Section C.2.2 and Lemma C.3.2 1(e). Then, collecting immediate results, we can conclude that $\|\mathbb{A}_{2t}\| = O_p\left(N^{1/2}T+T^{3/2}\right)$.

Next, we study $\mathbb{A}_{3t}$. Note that $\mathbb{A}_{3t} = \widehat{\boldsymbol{B}}^{\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}_t + \widehat{\boldsymbol{B}}^{\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\left(\boldsymbol{\rho}_N-\boldsymbol{I}_N\right)\boldsymbol{e}_{t-1} \equiv \mathbb{A}_{3t1} +$

$\mathbb{A}_{3t2}$. We have further decomposition for $\mathbb{A}_{3t1}$ as follows: $\mathbb{A}_{3t1} = \boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{B}^{0}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}_t + (\widehat{\boldsymbol{B}} - \boldsymbol{B}^0\boldsymbol{H}')'\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}_t \equiv \mathbb{A}_{3t1a} + \mathbb{A}_{3t1b}$.

$$\begin{aligned}
\|\mathbb{A}_{3t1a}\| &\leq \|\boldsymbol{H}\| \left\| \sum_{s=2}^{T} \boldsymbol{B}_s^0 \boldsymbol{B}_s^{0\prime} \right\| \left\| \sum_{i=1}^{N} \boldsymbol{\lambda}_i^0 \epsilon_{it} \right\| \\
&= O_p(1)\, O_p\left(\left\|\boldsymbol{B}^0\right\|^2\right) O_p\left(N^{1/2}\right) \\
&= O_p\left(N^{1/2}\left\|\boldsymbol{B}^0\right\|^2\right) = O_p\left(N^{1/2}T\right),
\end{aligned}$$

where the first equality holds by the fact that $\|\boldsymbol{H}\| = O_p(1)$, Lemma C.3.2 1(e), and the fact that $\left\|\sum_{i=1}^{N} \boldsymbol{\lambda}_i^0 \epsilon_{it}\right\| = O_p\left(N^{1/2}\right)$ under Assumptions A3.1 and A3.2 as given in Section C.2.2. Similarly,

$$\begin{aligned}
\|\mathbb{A}_{3t1b}\| &\leq \left\|\widehat{\boldsymbol{B}} - \boldsymbol{B}^0\boldsymbol{H}'\right\| \left[\sum_{s=2}^{T} \left|\sum_{i=1}^{N} \boldsymbol{B}_s^{0\prime} \boldsymbol{\lambda}_i \epsilon_{it}\right|^2\right]^{1/2} \\
&= O_p\left(T^{1/2}N^{-1/2}\right) O_p\left(N^{1/2}T^{1/2}\right) = O_p(T),
\end{aligned}$$

where we use (C.2.9) and the result give in Section C.2.2. Obviously, $\mathbb{A}_{3t1b}$ is dominated by $\mathbb{A}_{3t1a}$. Therefore, $\|\mathbb{A}_{3t1}\| = O_p\left(N^{1/2}T\right)$ holds.

Note that $\mathbb{A}_{3t2} = \boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{B}^{0}\boldsymbol{\Lambda}^{0\prime}(\boldsymbol{\rho}_N - \boldsymbol{I}_N)\boldsymbol{e}_{t-1} + (\widehat{\boldsymbol{B}} - \boldsymbol{B}^0\boldsymbol{H}')'\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}(\boldsymbol{\rho}_N - \boldsymbol{I}_N)\boldsymbol{e}_{t-1} \equiv \mathbb{A}_{3t2a} + \mathbb{A}_{3t2b}$. Similar to above arguments, it is easy to see that $\mathbb{A}_{3t2b}$ is dominated by $\mathbb{A}_{3t2a}$ due to the term $(\widehat{\boldsymbol{B}} - \boldsymbol{B}^0\boldsymbol{H}')$. Below, we bound $\mathbb{A}_{3t2a}$,

$$\begin{aligned}
\|\mathbb{A}_{3t2a}\| &\leq \|\boldsymbol{H}\| \left\|\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\right\| \left\|\boldsymbol{\Lambda}^{0\prime}(\boldsymbol{\rho}_N - \boldsymbol{I}_N)\boldsymbol{e}_{t-1}\right\| \\
&= O_p(1)\, O_p(T)\, O_p\left(N^{1/2}\right) \\
&= O_p\left(N^{1/2}T\right).
\end{aligned}$$

where the second line follows by the fact $\|\boldsymbol{H}\| = O_p(1)$, Lemma C.3.2 1(e) and the bound given in Section C.2.2. In sum, we have $\|\mathbb{A}_{3t}\| = O_p\left(N^{1/2}T\right)$. Collecting above immediate results for $\mathbb{A}_{1t}$, $\mathbb{A}_{2t}$, and $\mathbb{A}_{3t}$, we then have

$$\|\mathbb{A}_{1t}+\mathbb{A}_{2t}+\mathbb{A}_{3t}\| \leq \|\mathbb{A}_{1t}\|+\|\mathbb{A}_{2t}\|+\|\mathbb{A}_{3t}\| = O_p\left(N^{1/2}T^{1/2}+T+N^{1/2}T+NT^{1/2}+T^{3/2}\right). \tag{C.2.11}$$

Then, by (C.2.8) and the fact that $\|\boldsymbol{H}\| = O_p(1)$, it follows that

$$\begin{aligned}\left\|\boldsymbol{H}^{-1}\widehat{\boldsymbol{B}}_t - \boldsymbol{B}_t^0\right\| &\leq \|\boldsymbol{H}\|\left(N^{-1}T^{-1}\right)\left(\|\mathbb{A}_{1t}\|+\|\mathbb{A}_{2t}\|+\|\mathbb{A}_{3t}\|\right) \\ &= O_p\left(N^{-1/2}T^{-1/2}+N^{-1}+N^{-1/2}+T^{-1/2}+N^{-1}T^{1/2}\right) \end{aligned} \tag{C.2.12}$$

From above, the desired result follows directly.

Proofs of Lemma C.1.2(c)

Recall that $\widehat{\boldsymbol{\lambda}}_i = (\widehat{\boldsymbol{B}}'\widehat{\boldsymbol{B}})^{-1}\widehat{\boldsymbol{B}}'\boldsymbol{Z}_i$ where $\boldsymbol{Z}_i = \boldsymbol{B}^0\boldsymbol{\lambda}_i^0 + \boldsymbol{W}_i$. By the identity that $\boldsymbol{B}^0 = (\boldsymbol{B}^0 - \widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1}) + \widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1}$ and $\widehat{\boldsymbol{B}} = (\widehat{\boldsymbol{B}} - \boldsymbol{B}^0\boldsymbol{H}') + \boldsymbol{B}^0\boldsymbol{H}'$, and the normalization condition that $T^{-1}\widehat{\boldsymbol{B}}'\widehat{\boldsymbol{B}} = \boldsymbol{I}_{R_0}$, we have

$$\begin{aligned}\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{H}'^{-1}\boldsymbol{\lambda}_i^0 &= T^{-1}\boldsymbol{H}(\widehat{\boldsymbol{B}}'\boldsymbol{H}^{-1} - \boldsymbol{B}^0)'(\boldsymbol{B}^0 - \widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1})\boldsymbol{\lambda}_i^0 + T^{-1}\boldsymbol{H}\boldsymbol{B}^{0\prime}(\boldsymbol{B}^0 - \widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1})\boldsymbol{\lambda}_i^0 \\ &\quad + T^{-1}\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{W}_i + T^{-1}(\widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1} - \boldsymbol{B}^0)'\boldsymbol{W}_i \\ &\equiv \mathbb{D}_{1i} + \mathbb{D}_{2i} + \mathbb{D}_{3i} + \mathbb{D}_{4i}. \end{aligned} \tag{C.2.13}$$

We bound each term in the last display in turn.

First, for $\mathbb{D}_{1i}$ we have

$$\begin{aligned}\|\mathbb{D}_{1i}\| &\leq T^{-1}\left\|\widehat{\boldsymbol{B}} - \boldsymbol{B}^0\boldsymbol{H}'\right\|\left\|\widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1} - \boldsymbol{B}^0\right\|\left\|\boldsymbol{\lambda}_i^0\right\| \\ &= T^{-1}O_p\left(T^{1/2}N^{-1/2}\right)O_p\left(T^{1/2}N^{-1/2}\right)O_p(1) \\ &= O_p\left(N^{-1}\right), \end{aligned}$$

where the second line follows by (C.2.7), (C.2.9) and the Markov's inequality under Assumption A3.2.

For $\mathbb{D}_{2i}$, by substituting the expansions of $(\widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1} - \boldsymbol{B}^0)$ as shown in (C.2.3), we make the following decomposition,

$$\begin{aligned}
\mathbb{D}_{2i} &= T^{-1}\boldsymbol{H}\boldsymbol{B}^{0\prime}(\widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1} - \boldsymbol{B}^0)\boldsymbol{\lambda}_i \\
&= N^{-1}T^{-2}\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{W}'\boldsymbol{W}\widehat{\boldsymbol{B}}\boldsymbol{\lambda}_i^0 + N^{-1}T^{-2}\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{W}\widehat{\boldsymbol{B}}\boldsymbol{\lambda}_i^0 \\
&\quad + N^{-1}T^{-2}\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{W}'\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\boldsymbol{\lambda}_i^0 \\
&= \mathbb{D}_{2ia} + \mathbb{D}_{2ib} + \mathbb{D}_{2ic}.
\end{aligned}$$

Recall that $\boldsymbol{W} = (\boldsymbol{\rho}_N - \boldsymbol{I}_N)\,\boldsymbol{e}_{-1} + \boldsymbol{\epsilon}$. For $\mathbb{D}_{2ib}$, we have

$$\begin{aligned}
\mathbb{D}_{2ib} &= N^{-1}T^{-2}\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{W}\widehat{\boldsymbol{B}}\boldsymbol{\lambda}_i^0 \\
&= N^{-1}T^{-2}\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}(\boldsymbol{\rho}_N - \boldsymbol{I}_N)\,\boldsymbol{e}_{-1}\widehat{\boldsymbol{B}}\boldsymbol{\lambda}_i^0 + N^{-1}T^{-2}\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}\widehat{\boldsymbol{B}}\boldsymbol{\lambda}_i^0 \\
&= N^{-1}T^{-2}\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}(\boldsymbol{\rho}_N - \boldsymbol{I}_N)\,\boldsymbol{e}_{-1}\boldsymbol{B}^0\boldsymbol{H}'\boldsymbol{\lambda}_i^0 + N^{-1}T^{-2}\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}\boldsymbol{B}^0\boldsymbol{H}'\boldsymbol{\lambda}_i^0 \\
&\quad + N^{-1}T^{-2}\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}(\boldsymbol{\rho}_N - \boldsymbol{I}_N)\,\boldsymbol{e}_{-1}(\widehat{\boldsymbol{B}} - \boldsymbol{B}^0\boldsymbol{H}')\boldsymbol{\lambda}_i^0 \\
&\quad + N^{-1}T^{-2}\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}(\widehat{\boldsymbol{B}} - \boldsymbol{B}^0\boldsymbol{H}')\boldsymbol{\lambda}_i^0 \\
&\equiv \mathbb{D}_{2ib,1} + \mathbb{D}_{2ib,2} + \mathbb{D}_{2ib,3} + +\mathbb{D}_{2ib,4},
\end{aligned}$$

From above, it is readily to show that $\mathbb{D}_{2ib,3}$ and $\mathbb{D}_{2ib,4}$ are dominated by $\mathbb{D}_{2ib,1}$ and $\mathbb{D}_{2ib,2}$ due to the the term $(\widehat{\boldsymbol{B}} - \boldsymbol{B}^0\boldsymbol{H}')$. Following the analysis of $\|\mathbb{A}_{2t1}\|$ in the proofs of Lemma C.1.2(b),

we have

$$
\begin{aligned}
\left\|\mathbb{D}_{2ib,2}\right\| &= \left\|N^{-1}T^{-2}\boldsymbol{\lambda}_i^{0\prime}\boldsymbol{H}\sum_{j=1}^{N}\sum_{s=2}^{T}\boldsymbol{B}_s^0\epsilon_{js}\boldsymbol{\lambda}_j^{0\prime}\sum_{t=2}^{T}\boldsymbol{B}_t^0\boldsymbol{B}_t^{0\prime}\boldsymbol{H}'\right\| \\
&\leq \left\|\boldsymbol{\lambda}_i^0\right\|\left\|\boldsymbol{H}\right\|^2\left\|N^{-1}T^{-1}\sum_{j=1}^{N}\sum_{s=2}^{T}\boldsymbol{B}_s^0\epsilon_{js}\boldsymbol{\lambda}_j^{0\prime}\right\|\left\|T^{-1}\sum_{t=2}^{T}\boldsymbol{B}_t^0\boldsymbol{B}_t^{0\prime}\right\| \\
&= O_p(1)O_p\left(N^{-1/2}T^{-1/2}\right)O_p\left(1\right) \\
&= O_p\left(N^{-1/2}T^{-1/2}\right),
\end{aligned}
$$

where the third line follows by the Markov's inequity under Assumption A3.2, (C.2.6), the result given in Section C.2.2 and Lemma C.3.2 1(e). Similarly, we have

$$
\begin{aligned}
\left\|\mathbb{D}_{2ib,1}\right\| &= \left\|N^{-1}T^{-2}\boldsymbol{\lambda}_i'\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{e}_{-1}'\left(\boldsymbol{\rho}_N-\boldsymbol{I}_N\right)\boldsymbol{\Lambda}^0\sum_{t=2}^{T}\boldsymbol{B}_t^0\boldsymbol{B}_t^{0\prime}\boldsymbol{H}'\right\| \\
&\leq N^{-1}T^{-1}\left\|\boldsymbol{\lambda}_i\right\|\left\|\boldsymbol{H}\right\|^2\left\|\boldsymbol{B}^0\right\|\left\|\boldsymbol{e}_{-1}'\left(\boldsymbol{\rho}_N-\boldsymbol{I}_N\right)\boldsymbol{\Lambda}^0\right\|\left\|T^{-1}\sum_{t=2}^{T}\boldsymbol{B}_t^0\boldsymbol{B}_t^{0\prime}\right\| \\
&= N^{-1}T^{-1}O_p(1)O_p\left(T^{1/2}\right)O_p\left(N^{1/2}\right)O_p\left(1\right) \\
&= O_p\left(N^{-1/2}T^{-1/2}\right),
\end{aligned}
$$

In sum, it follows that $\mathbb{D}_{2ib} = O_p\left(N^{-1/2}T^{-1/2}\right)$. For $\mathbb{D}_{2ic}$,

$$
\begin{aligned}
\mathbb{D}_{2ic} &= N^{-1}T^{-2}\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{W}'\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\boldsymbol{\lambda}_i \\
&= N^{-1}T^{-2}\left\{\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{W}'\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\boldsymbol{H}'\boldsymbol{\lambda}_i+\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{W}'\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}(\widehat{\boldsymbol{B}}-\boldsymbol{B}^0\boldsymbol{H}')\boldsymbol{\lambda}_i\right\} \\
&= N^{-1}T^{-2}\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{\epsilon}'\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\boldsymbol{H}'\boldsymbol{\lambda}_i+N^{-1}T^{-2}\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{e}_{-1}'\left(\boldsymbol{\rho}_N-\boldsymbol{I}_N\right)\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\boldsymbol{H}'\boldsymbol{\lambda}_i \\
&\quad +N^{-1}T^{-2}\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{\epsilon}'\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}(\widehat{\boldsymbol{B}}-\boldsymbol{B}^0\boldsymbol{H}')\boldsymbol{\lambda}_i \\
&\quad +N^{-1}T^{-2}\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{e}_{-1}'\left(\boldsymbol{\rho}_N-\boldsymbol{I}_N\right)\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}(\widehat{\boldsymbol{B}}-\boldsymbol{B}^0\boldsymbol{H}')\boldsymbol{\lambda}_i \\
&\equiv \mathbb{D}_{2ic,1}+\mathbb{D}_{2ic,2}+\mathbb{D}_{2ic,3}+\mathbb{D}_{2ic,4},
\end{aligned}
$$

Again, it is straightforward to see $\mathbb{D}_{2ic,3}$ and $\mathbb{D}_{2ic,4}$ are dominated by $\mathbb{D}_{2ic,1}$ and $\mathbb{D}_{2ic,2}$ due to the the term $(\widehat{\boldsymbol{B}}-\boldsymbol{B}^0\boldsymbol{H}')$. Note that

$$\begin{aligned}
\|\mathbb{D}_{2ic,1}\| &= \left\|N^{-1}T{-}2\boldsymbol{\lambda}_i'\sum_{s=2}^{T}\boldsymbol{H}\boldsymbol{B}_s^0\boldsymbol{B}_s^{0\prime}\sum_{j=1}^{N}\boldsymbol{\lambda}_j^0\sum_t\epsilon_{jt}\boldsymbol{B}_t^{0\prime}\boldsymbol{H}'\right\| \\
&\leq \|\boldsymbol{\lambda}_i\|\,\|\boldsymbol{H}\|^2\left\|T^{-1}\sum_{s=2}^{T}\boldsymbol{B}_s^0\boldsymbol{B}_s^{0\prime}\right\|\left\|N^{-1}T^{-1}\sum_{j=1}^{N}\boldsymbol{\lambda}_j^0\sum_t\epsilon_{jt}\boldsymbol{B}_t^{0\prime}\right\| \\
&= O_p(1)\,O_p\left(N^{-1/2}T^{-1/2}\right)=O_p\left(N^{-1}T^{-1/2}\right),
\end{aligned}$$

where the third line follows by the Markov's inequity under Assumption A3.2, (C.2.6), the result given in Section C.2.2 and Lemma C.3.2 1(e). Similarly,

$$\begin{aligned}
\|\mathbb{D}_{2ic,2}\| &= \left\|N^{-1}T^{-2}\boldsymbol{\lambda}_i'\sum_{s=2}^{T}\boldsymbol{H}\boldsymbol{B}_s^0\boldsymbol{B}_s^{0\prime}\boldsymbol{\Lambda}^{0\prime}(\boldsymbol{\rho}_N-\boldsymbol{I}_N)\,\boldsymbol{e}_{-1}\boldsymbol{B}^0\boldsymbol{H}'\right\| \\
&\leq N^{-1}T^{-1}\|\boldsymbol{\lambda}_i\|\,\|\boldsymbol{H}\|^2\left\|T^{-1}\sum_{s=2}^{T}\boldsymbol{B}_s^0\boldsymbol{B}_s^{0\prime}\right\|\left\|\boldsymbol{\Lambda}^{0\prime}(\boldsymbol{\rho}_N-\boldsymbol{I}_N)\,\boldsymbol{e}_{-1}\right\|\left\|\boldsymbol{B}^0\right\| \\
&= N^{-1}T^{-1}O_p(1)\,O_p\left(N^{1/2}\right)O_p\left(\left\|\boldsymbol{B}^0\right\|\right) \\
&= O_p\left(N^{-1/2}T^{-1/2}\right),
\end{aligned}$$

In sum, it follows that $\mathbb{D}_{2ic}=O_p\left(N^{-1/2}T^{-1/2}\right)$.

For $\mathbb{D}_{2ia}$, similar to arguments for bounding $\mathbb{D}_{2ib}$ and $\mathbb{D}_{2ic}$ just above, we can readily show that the stochastic bound of the leading term $N^{-1}T^{-2}\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{W}'\boldsymbol{W}\boldsymbol{B}^0\boldsymbol{\lambda}_i$ is $O_p\left(T^{-1}\right)$ by using immediate results for bounding $\mathbb{D}_{2ib}$ and $\mathbb{D}_{2ic}$.

It follows that $\mathbb{D}_{2i}=O_p\left(T^{-1}+N^{-1/2}T^{-1/2}\right)$

For $\mathbb{D}_{3i}$, we have $\mathbb{D}_{3i}=T^{-1}\boldsymbol{H}\left\{\boldsymbol{B}^{0\prime}\boldsymbol{\epsilon}_i+\boldsymbol{B}^{0\prime}\left(\rho_i^0-1\right)\boldsymbol{e}_{i,-1}\right\}\equiv T^{-1}\boldsymbol{H}\left(\mathbb{D}_{3ia}+\mathbb{D}_{3ib}\right)$. Noting that $E\left(\mathbb{D}_{3ia}\right)=\mathbf{0}$ under Assumptions A3.1(c) and A3.4, we can readily show that $\left\|\sum_{s=2}^{T}\boldsymbol{B}_s^0\epsilon_{is}\right\|=O_p\left(T^{1/2}\right)$ by Davydov's inequality and Lemma C.3.2 1(c) under Assumptions A3.1 and A3.4 by noting that $\boldsymbol{B}_s^0=T^{-1}\boldsymbol{\nu}\boldsymbol{F}_{t-1}+\boldsymbol{u}_t$. So, $\|\mathbb{D}_{3ia}\|=O_p\left(T^{1/2}\right)$. Similarly, we can readily show that $\|\mathbb{D}_{3ib}\| = O_p(1) = o_p\left(T^{1/2}\right)$. Then, use the fact that $\|\boldsymbol{H}\| = O_p(1)$, we have

$$\|\mathbb{D}_{3i}\| \leq T^{-1}\|\boldsymbol{H}\|\left(\|\mathbb{D}_{3ia}\| + \|\mathbb{D}_{3ib}\|\right) = O_p\left(T^{-1}T^{1/2}\right) = O_p\left(T^{-1/2}\right)$$

For $\mathbb{D}_{4i}$, we make the following decomposition

$$\begin{aligned}
\mathbb{D}_{4i} &= N^{-1}T^{-2}\widehat{\boldsymbol{B}}'\boldsymbol{W}'\boldsymbol{W}\boldsymbol{W}_i + N^{-1}T^{-2}\widehat{\boldsymbol{B}}'\boldsymbol{W}'\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}\boldsymbol{W}_i + N^{-1}T^{-2}\widehat{\boldsymbol{B}}'\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{W}\boldsymbol{W}_i \\
&\equiv \mathbb{D}_{4ia} + \mathbb{D}_{4ib} + \mathbb{D}_{4ic}.
\end{aligned}$$

Recall that $\boldsymbol{W} = (\boldsymbol{\rho}_N - \boldsymbol{I}_N)\boldsymbol{e}_{-1} + \boldsymbol{\epsilon}$. For $\mathbb{D}_{4ib}$, we have

$$\begin{aligned}
\mathbb{D}_{4ia} &= N^{-1}T^{-2}\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{W}'\boldsymbol{W}\boldsymbol{W}_i + N^{-1}T^{-2}(\widehat{\boldsymbol{B}} - \boldsymbol{B}^0\boldsymbol{H}')'\boldsymbol{W}'\boldsymbol{W}\boldsymbol{W}_i \\
&= N^{-1}T^{-2}\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{W}'\boldsymbol{W}\boldsymbol{\epsilon}_i + N^{-1}T^{-2}\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{W}'\boldsymbol{W}\boldsymbol{e}_{i,-1}\left(\rho_i^0 - 1\right) \\
&\quad N^{-1}T^{-2}(\widehat{\boldsymbol{B}} - \boldsymbol{B}^0\boldsymbol{H}')\boldsymbol{W}'\boldsymbol{W}\boldsymbol{\epsilon}_i + N^{-1}T^{-2}(\widehat{\boldsymbol{B}} - \boldsymbol{B}^0\boldsymbol{H}')\boldsymbol{W}'\boldsymbol{W}\boldsymbol{e}_{i,-1}\left(\rho_i^0 - 1\right) \\
&\equiv \mathbb{D}_{4ia1} + \mathbb{D}_{4ia2} + \mathbb{D}_{4ia3} + \mathbb{D}_{4ia4}.
\end{aligned}$$

Obviously, $\mathbb{D}_{4ia3}$ and $\mathbb{D}_{4ia4}$ are dominated by $\mathbb{D}_{4ia1}$ and $\mathbb{D}_{4ia2}$ due to the term $(\widehat{\boldsymbol{B}} - \boldsymbol{B}^0\boldsymbol{H}')$. It then suffices to $\mathbb{D}_{4ia1}$ and $\mathbb{D}_{4ia2}$ as follows.

Note that $\mathbb{D}_{4ia1} = N^{-1}T^{-2}\boldsymbol{H}\boldsymbol{B}^{0\prime}\left[(\boldsymbol{\rho}_N - \boldsymbol{I}_N)\boldsymbol{e}_{-1} + \boldsymbol{\epsilon}\right]'\left[(\boldsymbol{\rho}_N - \boldsymbol{I}_N)\boldsymbol{e}_{-1} + \boldsymbol{\epsilon}\right]\boldsymbol{\epsilon}_i \equiv \mathbb{D}_{4ia1} + \mathbb{D}_{4ia11} + \mathbb{D}_{4ia12} + \mathbb{D}_{4ia13} + \mathbb{D}_{4ia14}$ with

$$\mathbb{D}_{4ia11} = N^{-1}T^{-2}\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{e}_{-1}'(\boldsymbol{\rho}_N - \boldsymbol{I}_N)^2\boldsymbol{e}_{-1}\boldsymbol{\epsilon}_i,$$

$$\mathbb{D}_{4ia12} = N^{-1}T^{-2}\boldsymbol{H}\boldsymbol{B}^{0\prime}\left[(\boldsymbol{\rho}_N - \boldsymbol{I}_N)\boldsymbol{e}_{-1}\right]'\boldsymbol{\epsilon}\boldsymbol{\epsilon}_i,$$

$$\mathbb{D}_{4ia13} = N^{-1}T^{-2}\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{\epsilon}'\left[(\boldsymbol{\rho}_N - \boldsymbol{I}_N)\boldsymbol{e}_{-1}\right]\boldsymbol{\epsilon}_i,$$

$$\mathbb{D}_{4ia14} = N^{-1}T^{-2}\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{\epsilon}'\boldsymbol{\epsilon}\boldsymbol{\epsilon}_i.$$

Note that for $\mathbb{D}_{4ia1}$,

$$\begin{aligned}\mathbb{D}_{4ia11} &\leq N^{-1}T^{-2}\left\|\boldsymbol{H}\right\|\left\|\boldsymbol{B}^{0}\right\|\left\|\left(\boldsymbol{\rho}_{N}-\boldsymbol{I}_{N}\right)\boldsymbol{e}_{-1}\right\|^{2}\left\|\boldsymbol{\epsilon}_{i}\right\| \\ &= N^{-1}T^{-2}O_{p}\left(1\right)O_{p}\left(T^{1/2}\right)O_{p}\left(N\right)O_{p}\left(T^{1/2}\right)=O_{p}\left(T^{-1}\right)\end{aligned}$$

where the second line above follows by the fact that $\left\|\boldsymbol{H}\right\| = O_{p}\left(1\right)$, Lemma C.3.2 1(e), the bound given in Section C.2.2, and the fact that $\left\|\boldsymbol{\epsilon}_{i}\right\|^{2} = T\left(T^{-1}\sum_{t}\epsilon_{it}^{2}\right) = O_{p}\left(T\right)$ under Assumption A3.1(c). Similarly, we can obtain that $\mathbb{D}_{4ia12} = O_{p}\left(N^{-1/2}T^{-1}\right)$, $\mathbb{D}_{4ia13} = O_{p}\left(N^{-1/2}T^{-1}\right)$ and $\mathbb{D}_{4ia14} = O_{p}\left(N^{-1}+T^{-1}\right)$ by Lemma C.3.2, the results given in C.2.2, and Assumptions A3.1 and A3.3. Thus, $\mathbb{D}_{4ia1} = O_{p}\left(N^{1/2}T^{-3/2}+N^{-1/2}T^{-1/2}+N^{-1}+T^{-1}\right)$. By similar decomposition and arguments above, $\mathbb{D}_{4ia2} = O_{p}\left(T^{-3/2}+N^{-1}T^{-1/2}+N^{-1/2}T^{-1}\right)$ holds.

Thus, collecting above immediate results, we have $\mathbb{D}_{4ia} = O_{p}\left(N^{-1/2}T^{-1/2}+N^{-1}+T^{-1}\right)$. We would like to address that some of bounds for the terms in the decomposition of $\mathbb{D}_{4ia}$ can be improved, though they are still of smaller of order than other terms here.

For $\mathbb{D}_{4ib}$, we have

$$\begin{aligned}\mathbb{D}_{4ib} &= N^{-1}T^{-2}\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{W}^{\prime}\boldsymbol{\Lambda}^{0}\boldsymbol{B}^{0\prime}\boldsymbol{W}_{i}+N^{-1}T^{-2}(\widehat{\boldsymbol{B}}-\boldsymbol{B}^{0}\boldsymbol{H}^{\prime})^{\prime}\boldsymbol{W}^{\prime}\boldsymbol{\Lambda}^{0}\boldsymbol{B}^{0\prime}\boldsymbol{W}_{i} \\ &= N^{-1}T^{-2}\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{\epsilon}^{\prime}\boldsymbol{\Lambda}^{0}\boldsymbol{B}^{0\prime}\boldsymbol{W}_{i}+N^{-1}T^{-2}\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{e}_{-1}^{\prime}\left(\boldsymbol{\rho}_{N}-\boldsymbol{I}_{N}\right)\boldsymbol{\Lambda}^{0}\boldsymbol{B}^{0\prime}\boldsymbol{W}_{i} \\ &\quad N^{-1}T^{-2}(\widehat{\boldsymbol{B}}-\boldsymbol{B}^{0}\boldsymbol{H}^{\prime})^{\prime}\boldsymbol{\epsilon}^{\prime}\boldsymbol{\Lambda}^{0}\boldsymbol{B}^{0\prime}\boldsymbol{W}_{i}+N^{-1}T^{-2}(\widehat{\boldsymbol{B}}-\boldsymbol{B}^{0}\boldsymbol{H}^{\prime})^{\prime}\boldsymbol{e}_{-1}^{\prime}\left(\boldsymbol{\rho}_{N}-\boldsymbol{I}_{N}\right)\boldsymbol{\Lambda}^{0}\boldsymbol{B}^{0\prime}\boldsymbol{W}_{i} \\ &\equiv \mathbb{D}_{4ib1}+\mathbb{D}_{4ib2}+\mathbb{D}_{4ib3}+\mathbb{D}_{4ib4}.\end{aligned}$$

It is easy to see that $\mathbb{D}_{4ib3}$ and $\mathbb{D}_{4ib4}$ are dominated by $\mathbb{D}_{4ib1}$ and $\mathbb{D}_{4ib2}$ due to the term $(\widehat{\boldsymbol{B}}-\boldsymbol{B}^{0}\boldsymbol{H}^{\prime})$. So we focus on $\mathbb{D}_{4ib1}$ and $\mathbb{D}_{4ib2}$ below. Because $\boldsymbol{W}_{i} = \left(\rho_{i}^{0}-1\right)\boldsymbol{e}_{i,-1}+\boldsymbol{\epsilon}_{i}$. For $\mathbb{D}_{4ib1}$,

$$\begin{aligned}\mathbb{D}_{4ib1} &= N^{-1}T^{-2}\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{\epsilon}^{\prime}\boldsymbol{\Lambda}^{0}\boldsymbol{B}^{0\prime}\left(\rho_{i}^{0}-1\right)\boldsymbol{e}_{i,-1}+N^{-1}T^{-2}\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{\epsilon}^{\prime}\boldsymbol{\Lambda}^{0}\boldsymbol{B}^{0\prime}\boldsymbol{\epsilon}_{i} \\ &\equiv \mathbb{D}_{4ib11}+\mathbb{D}_{4ib12}\end{aligned}$$

We bound $\mathbb{D}_{4ib12}$ below firstly,

$$\begin{aligned}\left\|\mathbb{D}_{4ib12}\right\| &= \left\|N^{-1}T^{-2}\boldsymbol{H}\left(\boldsymbol{\rho}_0^0\right)^T\sum_{j=1}^{N}\sum_{s=2}^{T}\boldsymbol{B}_s^0\epsilon_{js}\boldsymbol{\lambda}_j^{0\prime}\sum_{t=2}^{T}\boldsymbol{B}_t^0\epsilon_{it}\right\| \\ &\leq \|\boldsymbol{H}\|\left\|N^{-1}T^{-1}\sum_{j=1}^{N}\sum_{s=2}^{T}\boldsymbol{B}_s^0\epsilon_{js}\boldsymbol{\lambda}_j^{0\prime}\right\|\left\|T^{-1}\sum_{t=2}^{T}\boldsymbol{B}_t^0\epsilon_{it}\right\| \\ &= O_p\left(N^{-1/2}T^{-1/2}\right)O_p\left(T^{-1/2}\right) = O_p\left(N^{-1/2}T^{-1}\right),\end{aligned}$$

where the third line above follows by (C.2.6), and the results given in Section C.2.2. Then, we can similarly obtain that $\left\|\mathbb{D}_{4ib11}\right\| = O_p\left(N^{-1/2}T^{-3/2}\right)$ because $\left\|\boldsymbol{B}^{0\prime}\boldsymbol{\epsilon}'\boldsymbol{\Lambda}^0\right\| = O_p\left(N^{1/2}T^{1/2}\right)$ and $\left\|\boldsymbol{B}^{0\prime}\left(\rho_i^0 - 1\right)\boldsymbol{e}_{i,-1}\right\| = O_p(1)$ hold according to Chebyshev's inequality by direct moments calculations that take advantages of Lemma C.3.2 and Assumptions A3.1 and A3.4. It follows that $\left\|\mathbb{D}_{4ib1}\right\| = O_p\left(N^{-1/2}T^{-1}\right)$.

For $\mathbb{D}_{4ib2}$,

$$\begin{aligned}\mathbb{D}_{4ib2} &= N^{-1}T^{-2}\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{e}_{-1}'\left(\boldsymbol{\rho}_N - \boldsymbol{I}_N\right)\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}\left(\rho_i^0 - 1\right)\boldsymbol{e}_{i,-1} \\ &\quad + N^{-1}T^{-2}\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{e}_{-1}'\left(\boldsymbol{\rho}_N - \boldsymbol{I}_N\right)\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}\boldsymbol{\epsilon}_i \\ &\equiv \mathbb{D}_{4ib21} + \mathbb{D}_{4ib22}\end{aligned}$$

Following similar arguments for bounding $\mathbb{D}_{4ib1}$, it is readily to show that $\left\|\mathbb{D}_{4ib2}\right\| = O_p\left(N^{-1/2}T^{-1}\right)$. Thus, $\left\|\mathbb{D}_{4ib}\right\| = O_p\left(N^{-1/2}T^{-1}\right)$ holds. And for $\mathbb{D}_{4ic}$, we have

$$\begin{aligned}\mathbb{D}_{4ic} &= N^{-1}T^{-2}\widehat{\boldsymbol{B}}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}\boldsymbol{\epsilon}_i + N^{-1}T^{-2}\widehat{\boldsymbol{B}}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\left(\boldsymbol{\rho}_N - \boldsymbol{I}_N\right)\boldsymbol{e}_{-1}\boldsymbol{e}_{i,-1}\left(\rho_i^0 - 1\right) \\ &\quad + N^{-1}T^{-2}\widehat{\boldsymbol{B}}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\left(\boldsymbol{\rho}_N - \boldsymbol{I}_N\right)\boldsymbol{e}_{-1}\boldsymbol{\epsilon}_i + N^{-1}T^{-2}\widehat{\boldsymbol{B}}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}\boldsymbol{e}_{i,-1}\left(\rho_i^0 - 1\right) \\ &\equiv \mathbb{D}_{4ic1} + \mathbb{D}_{4ic2} + \mathbb{D}_{4ic3} + \mathbb{D}_{4ic4}.\end{aligned}$$

We then bound $\mathbb{D}_{4ic1}$ below. Note that

$$\mathbb{D}_{4ic1} = N^{-1}T^{-2}\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{B}^{0}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}\boldsymbol{\epsilon}_i + N^{-1}T^{-2}\left(\widehat{\boldsymbol{B}} - \boldsymbol{B}^0\boldsymbol{H}'\right)'\boldsymbol{B}^{0}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}\boldsymbol{\epsilon}_i \equiv \mathbb{D}_{4ic11} + \mathbb{D}_{4ic12}.$$

It is easy to see that $\mathbb{D}_{4ic12}$ is dominated by $\mathbb{D}_{4ic11}$ due to the term $\left(\widehat{\boldsymbol{B}} - \boldsymbol{B}^0\boldsymbol{H}'\right)$. Thus, it suffices to focus on bounding $\mathbb{D}_{4ic11}$. To this end, we have

$$\begin{aligned}
\|\mathbb{D}_{4ic11}\| &= \left\|N^{-1}T^{-2}\sum_{s=1}^{T}\boldsymbol{H}\boldsymbol{B}_s^0\boldsymbol{B}_s^{0\prime}\sum_{j=1}^{N}\boldsymbol{\lambda}_j^0\sum_t \epsilon_{jt}\epsilon_{it}\right\| \\
&\leq \|\boldsymbol{H}\|\left\|T^{-1}\sum_{s=1}^{T}\boldsymbol{B}_s^0\boldsymbol{B}_s^{0\prime}\right\|\left\|N^{-1}T^{-1}\sum_{j=1}^{N}\boldsymbol{\lambda}_j^0\sum_t \epsilon_{jt}\epsilon_{it}\right\| \\
&= O_p(1)\, O_p\left(N^{-1/2}T^{-1/2}\right) = O_p\left(N^{-1/2}T^{-1/2}\right),
\end{aligned}$$

where the third line above follows from (C.2.6), Lemma C.3.2 1(e), and the bound is given in Section C.2.2.

It follows that $\|\mathbb{D}_{4ic1}\| = O_p\left(N^{-1/2}T^{-1/2}\right)$. Similarly, we can readily show that $\|\mathbb{D}_{4ic2}\|$, $\|\mathbb{D}_{4ic3}\|$ and $\|\mathbb{D}_{4ic4}\|$ are both $O_p\left(N^{-1/2}T^{-1/2}\right)$. Then, $\|\mathbb{D}_{4ic}\| = O_p\left(N^{-1/2}T^{-1/2}\right)$.

From above, $\|\mathbb{D}_{4i}\| = O_p\left(N^{-1} + T^{-1} + N^{-1/2}T^{-1/2}\right)$.

Collecting above all leading terms, namely, $\mathbb{D}_{1i}$, $\mathbb{D}_{2i}$, $\mathbb{D}_{3i}$, and $\mathbb{D}_{4i}$, it follows that:

$$\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{H}'^{-1}\boldsymbol{\lambda}_i = O_p\left(T^{-1/2} + N^{-1}\right).$$

C.2.4 Proofs of Lemma C.1.3

By (C.2.8) and (C.2.5), we have

$$
\begin{aligned}
& \left\{T^{-1}\boldsymbol{V}_{Z,N}\right\}'\sum_{s=2}^{q}\widehat{\boldsymbol{B}}_s - \left[\left(N^{-1}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^{0}\right)\left(T^{-1}\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\right)\right]'\sum_{s=2}^{q}\boldsymbol{B}_s \\
= & \left\{T^{-1}\boldsymbol{V}_{Z,N}\right\}'\sum_{s=2}^{q}\widehat{\boldsymbol{B}}_s - \boldsymbol{H}_B'\sum_{s=2}^{q}\boldsymbol{B}_s^0 \\
= & \left(N^{-1}T^{-1}\right)\left\{\widehat{\boldsymbol{B}}'\boldsymbol{W}'\sum_{s=2}^{q}\boldsymbol{W}_s + \widehat{\boldsymbol{B}}'\boldsymbol{W}'\boldsymbol{\Lambda}^0\sum_{s=2}^{q}\boldsymbol{B}_s^0 + \widehat{\boldsymbol{B}}'\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\sum_{s=2}^{q}\boldsymbol{W}_s\right\} \\
= & \left(N^{-1}T^{-1}\right)\left(\mathbb{C}_{1q}+\mathbb{C}_{2q}+\mathbb{C}_{3q}\right), \qquad \text{(C.2.14)}
\end{aligned}
$$

where $\mathbb{C}_{1q}$ to $\mathbb{C}_{3q}$ are defined as follows given that $\boldsymbol{W} = (\boldsymbol{\rho}_N - \boldsymbol{I}_N)\,\boldsymbol{e}_{-1} + \boldsymbol{\epsilon}$ and $\widehat{\boldsymbol{B}} = (\widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1} - \boldsymbol{B}^0)\boldsymbol{H}' + \boldsymbol{B}^0\boldsymbol{H}'$,

$$
\begin{aligned}
\mathbb{C}_{1q} \;=\;& \boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{\epsilon}'\sum_{s=2}^{q}\boldsymbol{\epsilon}_s + \left(\widehat{\boldsymbol{B}} - \boldsymbol{B}^0\boldsymbol{H}'\right)'\boldsymbol{\epsilon}'\sum_{s=2}^{q}\boldsymbol{\epsilon}_s \\
&+\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{e}_{-1}'\left(\boldsymbol{\rho}_N - \boldsymbol{I}_N\right)^2\sum_{s=2}^{q}\boldsymbol{e}_{s-1} + \left(\widehat{\boldsymbol{B}} - \boldsymbol{B}^0\boldsymbol{H}'\right)'\boldsymbol{e}_{-1}'\left(\boldsymbol{\rho}_N - \boldsymbol{I}_N\right)^2\sum_{s=2}^{q}\boldsymbol{e}_{s-1} \\
&+\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{e}_{-1}'\left(\boldsymbol{\rho}_N - \boldsymbol{I}_N\right)\sum_{s=2}^{q}\boldsymbol{\epsilon}_s + \left(\widehat{\boldsymbol{B}} - \boldsymbol{B}^0\boldsymbol{H}'\right)'\boldsymbol{e}_{-1}'\left(\boldsymbol{\rho}_N - \boldsymbol{I}_N\right)\sum_{s=2}^{q}\boldsymbol{\epsilon}_s \\
&+\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{\epsilon}'\left(\boldsymbol{\rho}_N - \boldsymbol{I}_N\right)\sum_{s=2}^{q}\boldsymbol{e}_{s-1} + \left(\widehat{\boldsymbol{B}} - \boldsymbol{B}^0\boldsymbol{H}'\right)'\boldsymbol{\epsilon}'\left(\boldsymbol{\rho}_N - \boldsymbol{I}_N\right)\sum_{s=2}^{q}\boldsymbol{e}_{s-1} \\
\equiv\;& \mathbb{C}_{1q1} + \mathbb{C}_{1q2} + \mathbb{C}_{1q3} + \mathbb{C}_{1q4} + \mathbb{C}_{1q5} + \mathbb{C}_{1q6} + \mathbb{C}_{1q7} + \mathbb{C}_{1q8}, \\
\mathbb{C}_{2q} \;=\;& \boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{\epsilon}'\boldsymbol{\Lambda}^0\sum_{s=2}^{q}\boldsymbol{B}_s^0 + \left(\widehat{\boldsymbol{B}} - \boldsymbol{B}^0\boldsymbol{H}'\right)'\boldsymbol{\epsilon}'\boldsymbol{\Lambda}^0\sum_{s=2}^{q}\boldsymbol{B}_s^0 \\
&+\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{e}_{-1}'\left(\boldsymbol{\rho}_N - \boldsymbol{I}_N\right)\boldsymbol{\Lambda}^0\sum_{s=2}^{q}\boldsymbol{B}_s^0 + \left(\widehat{\boldsymbol{B}} - \boldsymbol{B}^0\boldsymbol{H}'\right)'\boldsymbol{e}_{-1}'\left(\boldsymbol{\rho}_N - \boldsymbol{I}_N\right)\boldsymbol{\Lambda}^0\sum_{s=2}^{q}\boldsymbol{B}_s^0 \\
\equiv\;& \mathbb{C}_{2q1} + \mathbb{C}_{2q2} + \mathbb{C}_{2q3} + \mathbb{C}_{2q4}, \\
\mathbb{C}_{3q} \;=\;& \boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\sum_{s=2}^{q}\boldsymbol{\epsilon}_s + \left(\widehat{\boldsymbol{B}} - \boldsymbol{B}^0\boldsymbol{H}'\right)'\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\sum_{s=1}^{q}\boldsymbol{\epsilon}_s \\
&+\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\left(\boldsymbol{\rho}_N - \boldsymbol{I}_N\right)\sum_{s=2}^{q}\boldsymbol{e}_{s-1} + \left(\widehat{\boldsymbol{B}} - \boldsymbol{B}^0\boldsymbol{H}'\right)'\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\left(\boldsymbol{\rho}_N - \boldsymbol{I}_N\right)\sum_{s=2}^{q}\boldsymbol{e}_{s-1} \\
\equiv\;& \mathbb{C}_{3q1} + \mathbb{C}_{3q2} + \mathbb{C}_{3q3} + \mathbb{C}_{3q4}.
\end{aligned}
\tag{C.2.15}
$$

We bound $\max_{1\le q\le T}\|\mathbb{C}_{lq}\|$ for $l = 1, \dots, 3$ below. The arguments are similar to those in the proofs of Lemma 2 in Bai and Ng (2004).

For $\max_{1\le q\le T}\|C_{1q}\|$, it suffices to bound $\max_{1\le q\le T}\|C_{1q1}\|$ to $\max_{1\le q\le T}\|C_{1q8}\|$ separately. It is straightforward to see that $\max_{1\le q\le T}\|C_{1q2}\|$, $\max_{1\le q\le T}\|C_{1q4}\|$, $\max_{1\le q\le T}\|C_{1q6}\|$, and $\max_{1\le q\le T}\|C_{1q8}\|$ are dominated by $\max_{1\le q\le T}\|C_{1q1}\|$, $\max_{1\le q\le T}\|C_{1q3}\|$, $\max_{1\le q\le T}\|C_{1q5}\|$, and $\max_{1\le q\le T}\|C_{1q7}\|$ due to the term $\left(\widehat{\boldsymbol{B}} - \boldsymbol{B}^0\boldsymbol{H}'\right)$.

Hence, we focus on bounding $\max_{1\le q\le T}\|C_{1q1}\|$, $\max_{1\le q\le T}\|C_{1q3}\|$, $\max_{1\le q\le T}\|C_{1q5}\|$, and

$\max_{1\leq q\leq T}\|C_{1q7}\|$ below. Note that

$$\mathbb{C}_{1q1} = \boldsymbol{H}\boldsymbol{B}^{0\prime}E\left(\boldsymbol{\epsilon}'\sum_{s=1}^{q}\boldsymbol{\epsilon}_s\right)+\boldsymbol{H}\boldsymbol{B}^{0\prime}\left[\boldsymbol{\epsilon}'\sum_{s=1}^{q}\boldsymbol{\epsilon}_s-E\left(\boldsymbol{\epsilon}'\sum_{s=1}^{q}\boldsymbol{\epsilon}_s\right)\right]\equiv\mathbb{C}_{1q1a}+\mathbb{C}_{1q1b}.$$

Note that $\mathbb{C}_{1q1a}=\boldsymbol{H}\bar{\mathbb{C}}_{1q1a}$ where $\bar{\mathbb{C}}_{1q1a}=\boldsymbol{B}^{0\prime}E\left(\boldsymbol{\epsilon}'\sum_{s=1}^{q}\boldsymbol{\epsilon}_s\right)$. Note that

$$\begin{aligned}\left\|N^{-1}T^{-1/2}\bar{\mathbb{C}}_{1q1a}\right\| &= \left\|T^{-1/2}\sum_{t=2}^{T}\boldsymbol{B}_t^0\sum_{s=1}^{q}\left[N^{-1}\sum_{i=1}^{N}E(\epsilon_{it}\epsilon_{is})\right]\right\| \\ &\leq \left\|T^{-1/2}\sum_{t=2}^{T}\boldsymbol{B}_t^0\right\|\sum_{s=1}^{T}\left\|N^{-1}\sum_{i=1}^{N}E(\epsilon_{it}\epsilon_{is})\right\| \\ &\leq MO_p(1)\,O_p(1)=O_p(1)\ \text{ uniformly in } q,\end{aligned}$$

where the third line above follows from the fact that $\max_{1\leq s\leq T}\sum_{t=1}^{T}\left|N^{-1}\sum_{i=1}^{N}E(\epsilon_{it}\epsilon_{is})\right|\leq M$ by Assumption A3.3 and $\left\|T^{-1/2}\sum_{t=2}^{T}\boldsymbol{B}_t^0\right\|=O_p(1)$by Lemma C.3.2 given that $\boldsymbol{B}_t^0=T^{-1}\boldsymbol{\nu}\boldsymbol{F}_{t-1}+\boldsymbol{u}_t$ for $t=2,\ldots,T$. This result, along with the fact that $\|\boldsymbol{H}\|=O_p(1)$, implies that

$$\max_{1\leq q\leq T}\|\mathbb{C}_{1q1a}\|\leq\|\boldsymbol{H}\|\max_{1\leq q\leq T}\left\|\bar{\mathbb{C}}_{1q1a}\right\|=O_p\left(NT^{1/2}\right). \qquad \text{(C.2.16)}$$

Let $\Phi_{q,s}=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{q}[\epsilon_{is}\epsilon_{it}-E\left(\epsilon_{is}\epsilon_{it}\right)]$. Then

$$\mathbb{C}_{1q1b}=\sqrt{NT}\boldsymbol{H}\sum_{s=1}^{T}\boldsymbol{B}_s^0\Phi_{q,s}.$$

Following proofs of Lemma C.1.2(b) in Bai and Ng (2004), it suffices to show that for every $s=1,\ldots,T$ $\max_{1\leq q\leq T}|\Phi_{q,s}|=O_p(1)$. Equivalently, we are going to prove that

$$P\left(\max_{1\leq q\leq T}|\Phi_{q,s}|\geq K\right)=o(1)$$

where K is finite and chosen to be large enough. Let $\varphi_{it}^{s}=\epsilon_{is}\epsilon_{it}-E\left(\epsilon_{is}\epsilon_{it}\right)$ and $\vartheta_{Nq}=N^{2/(4+\delta)}q^{(2+4m)/(4+\delta)}$, where $m>0$ and can be small enough, δ is defined in Assumption

A3.1. Let $\mathbf{1}_{it} = \mathbf{1}\{|\varphi_{it}^s| \leq \vartheta_{Nq}\}$, and $\bar{\mathbf{1}}_{it} = 1 - \mathbf{1}_{it}$. Define

$$\varphi_{1it}^s = \varphi_{it}^s \mathbf{1}_{it} - E\left(\varphi_{it}^s \mathbf{1}_{it}\right), \varphi_{2it}^s = \varphi_{it}^s \bar{\mathbf{1}}_{it}, \text{ and } \varphi_{3it}^s = E\left(\varphi_{it}^s \bar{\mathbf{1}}_{it}\right).$$

Apparently $\varphi_{1it}^s + \varphi_{2it}^s - \varphi_{3it}^s = \varphi_{it}^s$ as $\mathbb{E}\left(\varphi_{it}^s\right) = 0$. We prove the claim by showing that

(i1) $P\left(\max_{1\leq q\leq T}\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^q \varphi_{1it}^s\right| \geq K\right) = o(1)$

(i2) $P\left(\max_{1\leq q\leq T}\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^q \varphi_{2it}^s\right| \geq K\right) = o(1)$, and (i3) $\max_{1\leq q\leq T}\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^q \varphi_{3it}^s\right| = o(1)$.

First, we prove (i3). Note that

$$\begin{aligned}
\max_{1\leq q\leq T}\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^q \varphi_{3it}^s\right| &\leq \frac{1}{\sqrt{NT}}\max_{1\leq q\leq T}\left\{\sum_{i=1}^N\sum_{t=1}^q E|\varphi_{it}^s|^2\right\}^{1/2}\left\{\sum_{i=1}^N\sum_{t=1}^q P\left(|\varphi_{it}^s| > \vartheta_{Nq}\right)\right\}^{1/2} \\
&\leq \frac{1}{\sqrt{NT}}\max_{1\leq q\leq T}(\sqrt{Nq})\left\{\sum_{i=1}^N\sum_{t=1}^q P\left(|\varphi_{it}^s| > \vartheta_{Nq}\right)\right\}^{1/2} \\
&\leq \frac{1}{\sqrt{NT}}\max_{1\leq q\leq T}(\sqrt{Nq})\left\{\sum_{i=1}^N\sum_{t=1}^q \vartheta_{Nq}^{-(4+\delta)/2}E|\varphi_{it}^s|^{(4+\delta)/2}\right\}^{1/2} \\
&\leq \frac{1}{\sqrt{NT}}\max_{1\leq q\leq T} Nq\vartheta_{Nq}^{-(4+\delta)/4} = O\left(T^{-m}\right) = o\left(1\right).
\end{aligned}$$

where the first inequality holds due to Holder's inequality, the second and fourth inequalities hold because $E|\varphi_{it}^s|^{(4+\delta)/2} < M$ by the construction of φ_{it}^s under Assumption A3.1. The third inequality holds because of Markov inequality.

Next, we prove (i2). Noting that $\max_{1\leq q\leq T}\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^q \varphi_{2it}^s\right| \geq K$ implies that

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T |\varphi_{it}^s|\,\bar{\mathbf{1}}_{it} \geq K,$$

using Holder and Markov inequalities, we have,

$$\begin{aligned}
& P\left(\max_{1\leq q\leq T}\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{q}\varphi_{2it}^{s}\right|\geq K\right) \\
\leq\ & P\left[\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}|\varphi_{it}^{s}|\,\bar{\mathbf{1}}_{it}\geq K\right] \\
\leq\ & \frac{(NT)^{-1/2}\left\{\sum_{i=1}^{N}\sum_{t=1}^{T}E|\varphi_{it}^{s}|^{2}\right\}^{1/2}\left\{\sum_{i=1}^{N}\sum_{t=1}^{T}P\left(|\varphi_{it}^{s}|>\vartheta_{Nq}\right)\right\}^{1/2}}{K} \\
\leq\ & (NT)^{-1/2}(NT)^{1/2}\left\{\sum_{i=1}^{N}\sum_{t=1}^{T}\vartheta_{Nq}^{-(4+\delta)/2}E|\varphi_{it}^{s}|^{(4+\delta)/2}\right\}^{1/2} \\
\leq\ & (NT)^{1/2}\vartheta_{Nq}^{-(4+\delta)/4}=O\left(T^{-m}\right)=o\left(1\right).
\end{aligned}$$

where the third and fourth inequalities hold because $E|\varphi_{it}^{s}|^{(4+\delta)/2}<M$ by the construction of φ_{it}^{s} under Assumption A3.1.

To prove (i1), we consider two typical cases for q, i.e., (i1a) $q\asymp T$, (i1b) q is finite and. We first prove (i1) when $q\asymp T$. Without loss of generality, let $\{a_T\}$ be a sequence of integers such that $0<a_T<T$, $a_T\to\infty$ as $T\to\infty$, and $T-a_T=o(\sqrt{T})$. We have

$$\begin{aligned}
& P\left(\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{q}\varphi_{1it}^{s}\right|\geq K\right) \\
\leq\ & P\left(\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{a_T}\varphi_{1it}^{s}\right|+\max_{1\leq q\leq T}\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=a_T+1}^{q}\varphi_{1it}^{s}\right|\geq K\right) \\
\leq\ & P\left(\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{a_T}\varphi_{1it}^{s}\right|\geq K/2\right)+P\left(\max_{1\leq q\leq T}\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=a_T+1}^{q}\varphi_{1it}^{s}\right|\geq K/2\right).
\end{aligned}$$

Using Markov inequality, we bound the second term in the above display as follows,

$$
\begin{aligned}
& P\left(\max_{1\leq q\leq T}\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=a_T+1}^{q}\varphi_{1it}^{s}\right|\geq K/2\right) \\
\leq\ & P\left(\max_{1\leq q\leq T}\frac{1}{\sqrt{T}}\sum_{t=a_T+1}^{q}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\varphi_{1it}^{s}\right|\geq K/2\right) \\
\leq\ & P\left(\frac{1}{\sqrt{T}}\sum_{t=a_T+1}^{T}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\varphi_{1it}^{s}\right|\geq K/2\right) \\
\leq\ & T^{-1/2}\sum_{t=a_T+1}^{T}E\left|N^{-1/2}\sum_{i=1}^{N}\varphi_{1it}^{s}\right| \\
\leq\ & T^{-1/2}\sum_{t=a_T+1}^{T}E\left|N^{-1/2}\sum_{i=1}^{N}\varphi_{it}^{s}\right| = O\left((T-a_T)T^{-1/2}\right) = o\left(1\right),
\end{aligned}
$$

where the fourth line above holds by constructions of φ_{1it}^{s}, and $\left|N^{-1/2}\sum_{i=1}^{N}\varphi_{it}^{s}\right| = O_p\left(1\right)$ by Markov's inequity under Assumption A3.3(b).

Now, we are in the position to show $\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{a_T}\varphi_{1it}^{s}\right| = O_p\left(1\right)$, to this end, by Chebyshev's inequality, it suffices to show $E\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{a_T}\varphi_{1it}^{s}\right)^2 = O_p\left(1\right)$. Recall $\varphi_{1it}^{s} = \xi_{it}\mathbf{1}_{it}-E\left(\xi_{it}\mathbf{1}_{it}\right)$, and $\varphi_{it}^{s} = \epsilon_{is}\epsilon_{it}-E\left(\epsilon_{is}\epsilon_{it}\right)$. Therefore, under Assumption A3.1, $\{N^{-1/2}\sum_{i=1}^{N}\varphi_{1it}^{s}\}$ are still mixing sequence with zero mean. We have,

$$
\begin{aligned}
E\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{a_T}\varphi_{1it}^{s}\right)^2 &= \frac{1}{T}\sum_{t=1}^{a_T}\sum_{q=1}^{a_T}E\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\varphi_{1it}^{s}\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\varphi_{1jq}^{s}\right] \\
&\leq \frac{1}{T}\sum_{t=1}^{a_T}\sum_{q=1}^{a_T}(\alpha(|t-q|))^{\delta/(4+\delta)}\left(E\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\varphi_{1it}^{s}\right|^{(4+\delta)/2}\right)^{4/(4+\delta)} \\
&\leq \frac{1}{T}\sum_{t=1}^{a_T}\sum_{q=1}^{a_T}(\alpha(|t-q|))^{\delta/(4+\delta)}\left(E\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\varphi_{it}^{s}\right|^{(4+\delta)/2}\right)^{4/(4+\delta)} \\
&= O\left(a_T T^{-1}\right) = O\left(1\right),
\end{aligned}
$$

where the second line follows by the covariance inequality for the mixing sequences like (28) in Gao and Hong (2008), and $E\left|N^{-1/2}\sum_{i=1}^{N}\varphi_{it}^{s}\right|^{(4+\delta)/2} < M$ by Assumption A3.3(b) for some $\delta > 0$ defined Assumption A3.1. Collecting all above proofs for the claims (i1), (i2) and (i3), $\max_{1\leq q\leq T}|\Phi_{q,s}| = O_p(1)$ holds for every $s = 1,\ldots,T$. Besides, it is the trivial case to obtain $\max_{1\leq q\leq T}|\Phi_{q,s}| = O_p(1)$ when q is fixed and finite by similar arguments for the case $q \asymp T$. For other values of q, corresponding proofs can follow above two typical cases with slight modifications. Then, we have

$$\begin{aligned}
\max_{1\leq q\leq T}\|\mathbb{C}_{1q1b}\| &= \sqrt{NT}\max_{1\leq q\leq T}\left\|\boldsymbol{H}\sum_{s=2}^{T}\boldsymbol{B}_s^0\Phi_{q,s}\right\| \\
&\leq \sqrt{NT}\|\boldsymbol{H}\|\left\|\sum_{s=2}^{T}\boldsymbol{B}_s^0\right\|\max_{1\leq q\leq T}\|\Phi_{q,s}\| \\
&= \sqrt{NT}O_p(1)\,O_p\left(T^{1/2}\right)O_p(1) = O_p\left(N^{1/2}T\right), \qquad \text{(C.2.17)}
\end{aligned}$$

where the third line above follows because (C.2.6), and Lemma C.3.2 given that $\boldsymbol{B}_t^0 = T^{-1}\boldsymbol{\nu}\boldsymbol{F}_{t-1} + \boldsymbol{u}_t$ for $t = 2,\ldots,T$. In sum, we have

$$\max_{1\leq q\leq T}\|\mathbb{C}_{1q1}\| = O_p\left(N^{1/2}T\right) + O_p\left(NT^{1/2}\right). \qquad \text{(C.2.18)}$$

Next, let $\bar{\mathbb{C}}_{1q3} = \boldsymbol{e}_{-1}'(\boldsymbol{\rho}_N - \boldsymbol{I}_N)^2\sum_{s=2}^{q}\boldsymbol{e}_{s-1}$, we have

$$\begin{aligned}
\max_{1\leq q\leq T}\|\mathbb{C}_{1q3}\| &= \max_{1\leq q\leq T}\left\|\boldsymbol{H}\boldsymbol{B}^{0\prime}\bar{\mathbb{C}}_{1q3}\right\| \\
&\leq \|\boldsymbol{H}\|\left\|\boldsymbol{B}^0\right\|\max_{1\leq q\leq T}\left\|\bar{\mathbb{C}}_{1q3}\right\| \\
&= O_p(1)\,O_p\left(T^{1/2}\right)O_p\left(N^{1/2}T^{1/2}\right) = O_p\left(N^{1/2}T\right),
\end{aligned}$$

where the third line above follows by (C.2.6), $\|\boldsymbol{B}^0\| = O_p\left(T^{1/2}\right)$ due to Lemma C.3.2 1(e), and we can readily show $\left\|\bar{\mathbb{C}}_{1q3}\right\| = O_p\left(N^{1/2}T^{1/2}\right)$ holds uniformly in q via moment inequalities for the strong-mixing random sequences, which is similar to arguments for the result (10) in Section C.2.2. Similarly, we can readily show that $\max_{1\leq q\leq T}\|\mathbb{C}_{1q5}\| = O_p\left(N^{1/2}T^{1/2}\right)$ and

$\max_{1 \leq q \leq T} \|\mathbb{C}_{1q7}\| = O_p\left(N^{1/2}T\right)$. It follows that

$$\max_{1 \leq q \leq T} \|\mathbb{C}_{1q}\| = O_p\left(N^{1/2}T + NT^{1/2}\right). \tag{C.2.19}$$

For $\mathbb{C}_{2q}$, we can see that $\max_{1 \leq q \leq T} \|\mathbb{C}_{2q2}\|$ and $\max_{1 \leq q \leq T} \|\mathbb{C}_{2q4}\|$ in (C.2.15) will be dominated by $\max_{1 \leq q \leq T} \|\mathbb{C}_{2q1}\|$ and $\max_{1 \leq q \leq T} \|\mathbb{C}_{2q3}\|$ due to the term $\left(\widehat{\boldsymbol{B}} - \boldsymbol{B}^0\boldsymbol{H}'\right)$. Then, it is enough to bound $\max_{1 \leq q \leq T} \|\mathbb{C}_{2q1}\|$ and $\max_{1 \leq q \leq T} \|\mathbb{C}_{2q3}\|$. Note that

$$\max_{1 \leq q \leq T} \|\mathbb{C}_{2q1}\| \leq \|\boldsymbol{H}\| \left\|\sum_i \sum_r \boldsymbol{B}_r^0 \epsilon_{ir} \boldsymbol{\lambda}_i^{0\prime}\right\| \max_{1 \leq q \leq T} \left\|\sum_{s=1}^{q} \boldsymbol{B}_s^0\right\|.$$

Since the term $\|\boldsymbol{H}\|$ does not involve q and is of order $O_p(1)$ according to (C.2.6), and the term $\|\sum_i \sum_r \boldsymbol{B}_r^0 \epsilon_{ir}' \boldsymbol{\lambda}_i^{0\prime}\|$ does not involve q and is of order $O_p\left(N^{1/2}T^{1/2}\right)$ in Frobenius norm according to the result in Section C.2.2. In addition, $\|\sum_{s=2}^{q} \boldsymbol{B}_s^0\| = O_p\left(T^{1/2}\right)$ holds uniformly in q by the same arguments for Lemma C.3.2 1(b). Then, based on these facts, it follows that $\max_{1 \leq q \leq T} \|\mathbb{C}_{2q1}\| = O_p\left(N^{1/2}T\right)$.

Similarly, for $\mathbb{C}_{2q3}$, $\max_{1 \leq q \leq T} \|\mathbb{C}_{2q3}\| \leq \|\boldsymbol{H}\| \|\boldsymbol{B}^0\| \left\|\boldsymbol{e}_{-1}' (\boldsymbol{\rho}_N - \boldsymbol{I}_N) \boldsymbol{\Lambda}^0\right\| \max_{1 \leq q \leq T} \|\sum_{s=2}^{q} \boldsymbol{B}_s^0\| = O_p\left(N^{1/2}T\right)$ given that $\|\boldsymbol{B}^0\| = O_p\left(T^{1/2}\right)$, and $\|\sum_{s=2}^{q} \boldsymbol{B}_s^0\| = O_p\left(T^{1/2}\right)$ holds uniformly in q by Lemma C.3.2 given that $\boldsymbol{B}_t^0 = T^{-1}\boldsymbol{\nu}\boldsymbol{F}_{t-1} + \boldsymbol{u}_t$ for $t = 2, \ldots, T$, meanwhile, because of the fact $\left\|\boldsymbol{e}_{-1}' (\boldsymbol{\rho}_N - \boldsymbol{I}_N) \boldsymbol{\Lambda}^0\right\| = O_p\left(N^{1/2}\right)$ as shown in Section C.2.2. So we obtain that

$$\max_{1 \leq q \leq T} \|\mathbb{C}_{2q}\| = O_p\left(N^{1/2}T\right). \tag{C.2.20}$$

.

Similarly, for $\mathbb{C}_{3q}$, it suffices to bound the dominant term $\max_{1 \leq q \leq T} \|\mathbb{C}_{3q1}\|$ and $\max_{1 \leq q \leq T} \|\mathbb{C}_{3q3}\|$. Note that

$$
\begin{aligned}
\max_{1\le q\le T}\|\mathbb{C}_{3q1}\| &= \left\|\boldsymbol{H}\boldsymbol{B}^{0\prime}\boldsymbol{B}^{0}\boldsymbol{\Lambda}^{\prime}\sum_{s=1}^{q}\boldsymbol{\epsilon}_s\right\| \\
&\le \|\boldsymbol{H}\|\left\|\sum_{r}\boldsymbol{B}_r^{0}\boldsymbol{B}_r^{0\prime}\right\|\max_{1\le q\le T}\left\|\sum_{i}\sum_{s=1}^{q}\boldsymbol{\lambda}_i^{0}\epsilon_{is}\right\| \\
&= O_p\left(N^{1/2}T^{3/2}\right)
\end{aligned}
$$

where $\|\boldsymbol{H}\| = O_p(1)$ by (C.2.6), $\|\sum_r \boldsymbol{B}_r^0\boldsymbol{B}_r^{0\prime}\| = O_p(T)$ by Lemma C.3.2 1(e). These terms does not involve q. Now, we are in the position to bound $\max_{1\le q\le T}\|\sum_i\sum_{s=1}^q \boldsymbol{\lambda}_i^0\epsilon_{is}\|$, to this end, we can readily obtain that, uniformly in q, $E\left\|N^{-1/2}T^{-1/2}\sum_i\sum_{s=1}^q\boldsymbol{\lambda}_i^0\epsilon_{is}\right\|\le M$ holds under Assumption A3.1 by trading $\{N^{-1/2}\sum_i\boldsymbol{\lambda}_i\epsilon_{is}\}$ as the mixing sequences over s, which is the similar to the bound shown in Lemma C.1.1(4) stated in Bai and Ng (2004). Then, we have $\max_{1\le q\le T}\|\sum_i\sum_{s=1}^q\boldsymbol{\lambda}_i\epsilon_{is}\| = O_p\left(N^{1/2}T^{1/2}\right)$.

Similarly, because $\max_{1\le q\le T}\|\mathbb{C}_{3q3}\| \le \|\boldsymbol{H}\|\,\|\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\|\max_{1\le q\le T}\left\|\boldsymbol{\Lambda}^{0\prime}(\boldsymbol{\rho}_N-\boldsymbol{I}_N)\sum_{s=2}^q\boldsymbol{e}_{s-1}\right\|$, then, $\max_{1\le q\le T}\|\mathbb{C}_{3q3}\| = O_p\left(N^{1/2}T^{3/2}\right)$ holds by using the fact that $\|\boldsymbol{B}^0\|^2 = O_p(T)$ and noting that $\left\|\boldsymbol{\Lambda}^{0\prime}(\boldsymbol{\rho}_N-\boldsymbol{I}_N)\sum_{s=2}^q\boldsymbol{e}_{s-1}\right\| = O_p\left(N^{1/2}T^{1/2}\right)$ holds uniformly in q by direct moments calculations under Assumptions A3.1, A3.4. We have

$$
\max_{1\le q\le T}\|\mathbb{C}_{3q}\| = O_p\left(N^{1/2}T^{3/2}\right). \tag{C.2.21}
$$

Neglecting those dominated terms, from above immediate results, we then can conclude that

$$
\begin{aligned}
&\max_{1\le q\le T}\left\|\sum_{s=1}^{q}\left(\boldsymbol{H}^{-1}\widehat{\boldsymbol{B}}_s-\boldsymbol{B}_s^0-\boldsymbol{u}_s\right)\right\| \\
\le\ & N^{-1}T^{-1}\left(\max_{1\le q\le T}\|\mathbb{C}_{1q}\|+\max_{1\le q\le T}\|\mathbb{C}_{2q}\|+\max_{1\le q\le T}\|\mathbb{C}_{3q}\|\right) \\
=\ & O_p\left(N^{-1}T^{-1}\right)\left[O_p\left(N^{1/2}T\right)+O_p\left(NT^{1/2}\right)+O_p\left(N^{1/2}T^{3/2}\right)\right] \\
=\ & O_p\left(N^{-1/2}T^{1/2}+N^{-1/2}+T^{-1/2}\right).
\end{aligned} \tag{C.2.22}
$$

Then, it follows $\max\limits_{1\le t\le T}\frac{1}{\sqrt{T}}\left\|\sum_{s=2}^{t}\boldsymbol{H}^{-1}\widehat{\boldsymbol{B}}_s-\boldsymbol{B}_s\right\|=O_p\left(N^{-1/2}\right)$ directly.

C.2.5 Proofs of Lemma C.1.4

Recall that $\widehat{\boldsymbol{\Lambda}}_i=(\widehat{\boldsymbol{B}}\widehat{\boldsymbol{B}}')^{-1}\widehat{\boldsymbol{B}}'\boldsymbol{Z}_i$ with $\boldsymbol{Z}_i=\boldsymbol{B}^0\boldsymbol{\Lambda}_i^0+\boldsymbol{W}_i$, where $\boldsymbol{Z}_i=(Z_{i1},\ldots,Z_{iT})'$ and $\boldsymbol{W}_i=(W_{i1},\ldots,W_{iT})'$. Then

$$
\begin{aligned}
\left\|\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}-\widehat{\boldsymbol{\Lambda}}^0\widehat{\boldsymbol{B}}'\right\|^2 &= \sum_{i=1}^{N}\left\|\boldsymbol{B}^0\boldsymbol{\Lambda}_i^0-\widehat{\boldsymbol{B}}\widehat{\boldsymbol{\Lambda}}_i\right\|^2=\sum_{i=1}^{N}\left\|P_{\widehat{\boldsymbol{B}}}\boldsymbol{Z}_i-\boldsymbol{f}^0\boldsymbol{\Lambda}_i^0\right\|^2\\
&= \sum_{i=1}^{N}\left\|P_{\widehat{\boldsymbol{B}}}\left(\boldsymbol{B}^0\boldsymbol{\Lambda}_i^0+\boldsymbol{W}_i\right)-\boldsymbol{B}^0\boldsymbol{\Lambda}_i^0\right\|^2=\sum_{i=1}^{N}\left\|M_{\widehat{\boldsymbol{B}}}\boldsymbol{B}^0\boldsymbol{\Lambda}_i^0-P_{\widehat{\boldsymbol{B}}}\boldsymbol{W}_i\right\|^2\\
&\le 2\sum_{i=1}^{N}\left\|M_{\widehat{\boldsymbol{B}}}\boldsymbol{B}^0\boldsymbol{\Lambda}_i^0\right\|^2+2\sum_{i=1}^{N}\left\|P_{\widehat{\boldsymbol{B}}}\boldsymbol{W}_i\right\|^2\equiv 2\mathcal{B}_1+2\mathcal{B}_2.
\end{aligned}
\tag{C.2.23}
$$

It suffices to bound $\mathcal{B}_1$ and $\mathcal{B}_2$. For $\mathcal{B}_2$, we have $\mathcal{B}_2\le 2\sum_{i=1}^{N}\boldsymbol{\epsilon}_i'P_{\widehat{\boldsymbol{B}}}\boldsymbol{\epsilon}_i+2\sum_{i=1}^{N}\left(\rho_i^0-1\right)^2\boldsymbol{e}_i'P_{\widehat{\boldsymbol{B}}}\boldsymbol{e}_i=O(N+T)$ by Lemma C.3.1(a). For $\mathcal{B}_1$, we apply Lemma C.1.2(a) and Assumption A3.2 to obtain

$$
\begin{aligned}
\mathcal{B}_1 &= \sum_{i=1}^{N}\left\|M_{\widehat{\boldsymbol{B}}}\left(\boldsymbol{B}^0-\widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1}\right)\boldsymbol{\Lambda}_i^0\right\|^2\le\left\|M_{\widehat{\boldsymbol{B}}}\right\|_{sp}^2\left\|\boldsymbol{B}^0-\widehat{\boldsymbol{B}}\boldsymbol{H}'^{-1}\right\|^2\sum_{i=1}^{N}\left\|\boldsymbol{\Lambda}_i^0\right\|^2\\
&= O(1)\,O_p\left(TN^{-1}\right)O_p(N)=O_p(T).
\end{aligned}
$$

Consequently, $||\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}-\widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{B}}'||^2=O_p(N+T)$.

C.3 Some Useful Lemmas

Lemma C.3.1. *Suppose Assumption A3.1 to A3.3 hold. Then*

(a) $\sup_{\boldsymbol{F}\in\mathbb{D}_P}(NT)^{-1}\sum_{i=1}^{N}\boldsymbol{\epsilon}_i'\boldsymbol{P}_F\boldsymbol{\epsilon}_i=O_P\left(N^{-1}+T^{-1}\right)$

(b) $(NT)^{-1}\left\|\boldsymbol{\epsilon}\boldsymbol{\epsilon}'\right\|=O_P\left(N^{-1/2}+T^{-1/2}\right)$ *and* $(NT)^{-1}\left\|\boldsymbol{\epsilon}'\boldsymbol{\epsilon}\right\|=O_P\left(N^{-1/2}+T^{-1/2}\right)$

where $\mathbb{D}_P = \left\{ \boldsymbol{F} \in \mathbb{R}^{T \times R} \right\}$*, and* (a)*,* (b) *hold under Assumptions A3.3.*

Proof. The derivations are exactly same as those in Peng et al. (2020). ■

Lemma C.3.2. *Suppose Assumption A3.1 holds, it follows that*

(a) *Under the alternative of local-to-unity such that* $\rho^0_{0,r} = 1 + \nu_r/T$*, then,*

   (a) $T^{-1/2}\boldsymbol{F}^0_{[Tr]} \Rightarrow \Sigma_u^{1/2}\mathbf{J}_\nu(r)$*;*

   (b) $T^{-3/2}\sum_{t=2}^T \boldsymbol{F}^0_t \Rightarrow \Sigma_u^{1/2}\int_0^1 \mathbf{J}_\nu(r)\mathrm{d}r$*;*

   (c) $T^{-2}\sum_{t=2}^T \boldsymbol{F}^0_{t-1}\boldsymbol{F}^{0\prime}_{t-1} \Rightarrow \Sigma_u^{1/2}\int_0^1 \mathbf{J}_\nu(r)\mathbf{J}_\nu(r)'\mathrm{d}r\Sigma_u^{1/2}$*;*

   (d) $T^{-1}\sum_{t=2}^T \boldsymbol{u}_t\boldsymbol{F}^{0\prime}_{t-1} \Rightarrow \Sigma_u^{1/2}\int_0^1 \mathrm{d}\mathbf{W}(r)'\mathbf{J}_\nu(r)\Sigma_u^{1/2} + \Omega_u$*, where* $\Omega_u \equiv \sum_{k=1}^\infty E\left(\boldsymbol{u}_t\boldsymbol{u}'_{t+k}\right)$*;*

   (e) $\left\|T^{-1}\sum_{t=2}^T \boldsymbol{B}^0_t\boldsymbol{B}^{0\prime}_t\right\| = O_p(1)$*;*

   where $\mathbf{J}_\nu(r) \equiv (\mathrm{J}_{\nu_1}(r), \ldots, \mathrm{J}_{\nu_R}(r))$ *is a Ornstein-Uhlenbeck process such that* $\mathbf{J}_\nu(0) = \mathbf{0}$ *and* $\mathbf{J}_\nu(r) = \boldsymbol{W}(r) + \boldsymbol{\nu}\int_0^r e^{(r-s)c}\boldsymbol{W}(s)\mathrm{d}s$*; besides,* $\boldsymbol{W}(r)$ *is the* R_0*-vector standard Brownian motion on* $\mathcal{C}[0,1]$ *that is given by the weak limit of the partial sum* $\Sigma_u^{-1/2}T^{-1/2}\sum_{t=1}^{[Tr]}\boldsymbol{u}_t$*.*

(b) *Further, if Assumption A3.5 also holds, then, for each* i

   (a) $T^{-1/2}e_{i,[Tr]} \Rightarrow \sigma_i\mathrm{J}_{c_i}(r)$*;*

   (b) $T^{-3/2}\sum_{t=2}^T e_{it} \Rightarrow \sigma_i\int_0^1 \mathrm{J}_{c_i}(r)\mathrm{d}r$*;*

   (c) $T^{-2}\sum_{t=2}^T e_{it-1}e'_{it-1} \Rightarrow \sigma_i^2\int_0^1 \mathrm{J}_{c_i}(r)\mathrm{J}_{c_i}(r)'\mathrm{d}r$*;*

   (d) $T^{-1}\sum_{t=2}^T e_{it}\epsilon'_{it-1} \Rightarrow \sigma_i^2\int_0^1 \mathrm{J}_{c_i}(r)\mathrm{d}\mathrm{W}(r)' + \Omega_{\epsilon i}$*, where* $\Omega_{\epsilon i} = \sum_{k=1}^\infty E\left(\epsilon_{it}\epsilon_{it+k}\right)$*;*

   where $\mathrm{J}_{c_i}(r) \equiv (\mathrm{J}_{\nu_1}(r), \ldots, \mathrm{J}_{\nu_R}(r))$ *is a Ornstein-Uhlenbeck process such that* $\mathrm{J}_{c_i}(0) = 0$ *and* $\mathrm{J}_{c_i}(r) = \mathrm{W}(r) + c_i\int_0^r e^{(r-s)c}\mathrm{W}(s)\mathrm{d}s$*; besides,* $\mathrm{W}(r)$ *is the standard Brownian motion on* $\mathcal{C}[0,1]$ *that is given by the weak limit of the partial sum* $\sigma_i^{-1}T^{-1/2}\sum_{t=1}^{[Tr]}\epsilon_{it}$*.*

Proof. According to lines developed in Phillips (1987) and Phillips (1988), proofs follow directly. ■

The next lemma studies the asymptotic property of $\boldsymbol{V}_{\boldsymbol{Z},\boldsymbol{N}}$.

Lemma C.3.3. *Under Assumption A3.1-A3.5, as $N, T \to +\infty$ jointly,*

$$T^{-1}\boldsymbol{V}_{Z,N} \to \Upsilon_1 \quad \text{(C.3.1)}$$

where $\Upsilon_1 \equiv \lim_{N,T\to\infty} \left(T^{-1}\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\right)^{-1} \left(T^{-1}\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\right) \left(N^{-1}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\right) \left(T^{-1}\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\right)$*, which is a positively definite matrix.*

Proof. Premultiplying $B^{0\prime}$ and T^{-1} on both sides of (C.2.2), recall that $\boldsymbol{W} = (\boldsymbol{\rho}_N - \boldsymbol{I}_N)\,\boldsymbol{e}_{-1} + \boldsymbol{\epsilon}$ and (C.2.1), we have

$$\begin{aligned}
& T^{-2}\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\boldsymbol{V}_{Z,N} - T^{-1}\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\left(N^{-1}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}\right)\left(T^{-1}\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\right) \\
= \; & N^{-1}T^{-2}\boldsymbol{B}^{0\prime}\boldsymbol{W}^{\prime}\boldsymbol{W}\widehat{\boldsymbol{B}} + N^{-1}T^{-2}\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{W}\widehat{\boldsymbol{B}} \\
& + N^{-1}T^{-2}\boldsymbol{B}^{0\prime}\boldsymbol{W}^{\prime}\boldsymbol{\Lambda}^0\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}} \\
\equiv \; & \mathcal{A}_1 + \mathcal{A}_2 + \mathcal{A}_3.
\end{aligned}$$

We bound each term below. First, note that

$$\begin{aligned}
\mathcal{A}_1 = \; & N^{-1}T^{-2}\boldsymbol{B}^{0\prime}\boldsymbol{e}_{-1}^{\prime}\left(\boldsymbol{\rho}_N - \boldsymbol{I}_N\right)^2\boldsymbol{e}_{-1}\widehat{\boldsymbol{B}} + N^{-1}T^{-2}\left(\boldsymbol{\epsilon}\boldsymbol{B}^0\right)^{\prime}\boldsymbol{\epsilon}\widehat{\boldsymbol{B}} \\
& + N^{-1}T^{-2}\boldsymbol{B}^{0\prime}\boldsymbol{e}_{-1}^{\prime}\left(\boldsymbol{\rho}_N - \boldsymbol{I}_N\right)\boldsymbol{\epsilon}\widehat{\boldsymbol{B}} + N^{-1}T^{-2}\left(\boldsymbol{\epsilon}\boldsymbol{B}^0\right)^{\prime}\left(\boldsymbol{\rho}_N - \boldsymbol{I}_N\right)\boldsymbol{e}_{-1}\widehat{\boldsymbol{B}} \\
\equiv \; & \mathcal{A}_{11} + \mathcal{A}_{12} + \mathcal{A}_{13} + \mathcal{A}_{14}.
\end{aligned}$$

For $\mathcal{A}_{11}$, we have

$$\begin{aligned}
\|\mathcal{A}_{11}\| = \; & N^{-1}T^{-2}\left\|\boldsymbol{B}^{0\prime}\boldsymbol{e}_{-1}^{\prime}\left(\boldsymbol{\rho}_N - \boldsymbol{I}_N\right)^2\boldsymbol{e}_{-1}\widehat{\boldsymbol{B}}\right\| \\
\leq \; & N^{-1}T^{-1}\left\|T^{-1/2}\boldsymbol{B}^{0\prime}\right\|\left\|\boldsymbol{e}_{-1}^{\prime}\left(\boldsymbol{\rho}_N - \boldsymbol{I}_N\right)^2\boldsymbol{e}_{-1}\right\|\left\|T^{-1/2}\widehat{\boldsymbol{B}}\right\| \\
= \; & N^{-1}T^{-1}O_p(1)\,O_p(N)\,O_p(1) \\
= \; & O_p\left(T^{-1}\right),
\end{aligned}$$

where the third line above follows from Lemma C.3.2 1(e), the bound given in Section C.2.2 and the normalization condition. Similarly, we have

$$\begin{aligned}\|\mathcal{A}_{12}\| &\leq N^{-1}T^{-1}\left\|T^{-1/2}\boldsymbol{B}^0\right\|\|\boldsymbol{\epsilon}'\boldsymbol{\epsilon}\|_{sp}\left\|T^{-1/2}\widehat{\boldsymbol{B}}\right\| \\ &= N^{-1}T^{-1}O_p(1)\,O_p(N+T)\,O_p(1) \\ &= O_p\left(N^{-1}+T^{-1}\right),\end{aligned}$$

where the second line above follows from Lemma C.3.2, Assumption A3.3, and the normalization condition. By Cauchy-Schwarz inequality, it is easy to see $\|\mathcal{A}_{13}\|$ and $\|\mathcal{A}_{14}\|$ will not be the dominant term in comparison with $\|\mathcal{A}_{11}\|$ and $\|\mathcal{A}_{12}\|$. It follows that

$$\|\mathcal{A}_1\| = O_p\left(N^{-1}+T^{-1}\right). \tag{C.3.2}$$

Next, we study $\mathcal{A}_2$. Note that

$$\begin{aligned}\mathcal{A}_2 &= N^{-1}T^{-2}\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}(\boldsymbol{\rho}_N-\boldsymbol{I}_N)\,\boldsymbol{e}_{-1}\widehat{\boldsymbol{B}}+N^{-1}T^{-2}\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}\widehat{\boldsymbol{B}} \\ &\equiv \mathcal{A}_{21}+\mathcal{A}_{22}.\end{aligned}$$

We then bound $\mathcal{A}_{21}$ below,

$$\begin{aligned}\|\mathcal{A}_{21}\| &\leq N^{-1}T^{-3/2}\left\|\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\right\|\left\|\boldsymbol{\Lambda}^{0\prime}(\boldsymbol{\rho}_N-\boldsymbol{I}_N)\,\boldsymbol{e}_{-1}\right\|\left\|T^{-1/2}\widehat{\boldsymbol{B}}\right\| \\ &= N^{-1}T^{-3/2}O_p(T)\,O_p\left(N^{1/2}\right)O_p(1) = O_p\left(N^{-1/2}T^{-1/2}\right),\end{aligned}$$

where the first line above holds by Lemma C.3.2 1(e), the result given in Section C.2.2, and the normalization condition. Similarly,

$$\begin{aligned}\|\mathcal{A}_{22}\| &\leq N^{-1/2}T^{-3/2}\left\|\boldsymbol{B}^{0\prime}\boldsymbol{B}^{0}\right\|\left\|\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\epsilon}}{N^{1/2}T^{1/2}}\right\|\left\|T^{-1/2}\widehat{\boldsymbol{B}}\right\| \\ &= O_p\left(N^{-1/2}\right),\end{aligned}$$

where the first line above holds by Lemma C.3.2 1(e), the bound given in Section C.2.2 and the normalization condition. In sum, we have

$$\|\mathcal{A}_{2}\| = O_p\left(N^{-1/2}\right). \tag{C.3.3}$$

Now, we study $\mathcal{A}_3$. Note that

$$\begin{aligned}\mathcal{A}_{3} &= N^{-1}T^{-2}\boldsymbol{B}^{0\prime}\boldsymbol{e}_{-1}'\left(\boldsymbol{\rho}_N-\boldsymbol{I}_N\right)\boldsymbol{\Lambda}^{0}\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}} + N^{-1}T^{-2}\boldsymbol{B}^{0\prime}\boldsymbol{\epsilon}'\boldsymbol{\Lambda}^{0}\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}} \\ &\equiv \mathcal{A}_{31}+\mathcal{A}_{32}.\end{aligned}$$

Below, we bound $\mathcal{A}_{31}$ and $\mathcal{A}_{32}$ in turns. For $\mathcal{A}_{31}$,

$$\begin{aligned}\|\mathcal{A}_{31}\| &\leq N^{-1}T^{-1}\left\|\boldsymbol{B}^{0\prime}\right\|\left\|\boldsymbol{e}_{-1}'\left(\boldsymbol{\rho}_N-\boldsymbol{I}_N\right)\boldsymbol{\Lambda}^{0}\right\|\left\|T^{-1}\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\right\| \\ &\leq N^{-1}T^{-1}\left\|\boldsymbol{B}^{0}\right\|O_p\left(N^{1/2}\right)O_p\left(1\right) = O_p\left(N^{-1/2}T^{-1/2}\right),\end{aligned}$$

where the first line above follows by Lemma C.3.2 1(e), the bound given in Section C.2.2, and Lemma C.1.1. Similarly,

$$\begin{aligned}\|\mathcal{A}_{32}\| &\leq N^{-1/2}T^{-1}\left\|\boldsymbol{\epsilon}\boldsymbol{B}^{0}\right\|\left\|\frac{\boldsymbol{\Lambda}^{0}}{N^{1/2}}\right\|\left\|T^{-1}\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\right\| \\ &= N^{-1/2}T^{-1}O_p\left(N^{1/2}T^{1/2}\right)O_p\left(1\right) = O_p\left(T^{-1/2}\right),\end{aligned}$$

where the second line above holds by Lemma C.3.2 1(e), Assumption A3.2, $\left\|\boldsymbol{\epsilon}\boldsymbol{B}^{0}\right\| = O_p\left(N^{1/2}T^{1/2}\right)$ according to the result given in Section C.2.2 by direct calculations, and Lemma C.1.1. In sum,

we have

$$\|\mathcal{A}_3\| = O_p\left(T^{-1/2}\right). \tag{C.3.4}$$

Combining (C.3.2), (C.3.3), and (C.3.4), and recall the fact that $\|\boldsymbol{B}^0\| = O_p\left(T^{1/2}\right)$, we have

$$\left\|\left(T^{-1}\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\right)\left(T^{-1}\boldsymbol{V}_{Z,N}\right) - T^{-1}\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\left(N^{-1}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\right)\left(T^{-1}\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\right)\right\| = o_p(1) \tag{C.3.5}$$

Thus, (C.3.5) implies that

$$\left(T^{-1}\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}\right)\left(T^{-1}\boldsymbol{V}_{Z,N}\right) = \left(T^{-1}\boldsymbol{B}^{0\prime}\boldsymbol{B}^0\right)\frac{\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0}{N}(T^{-1}\boldsymbol{B}^{0\prime}\widehat{\boldsymbol{B}}) + o_p(1)$$

and $T^{-1}\boldsymbol{V}_{Z,N} \xrightarrow{p} \Upsilon_1$ follows directly. ■