Multiscale Testing for Equality of Nonparametric Trend Curves

Marina Khismatullina[1] Michael Vogt[2]
University of Bonn University of Bonn

We develop multiscale methods to test qualitative hypotheses about nonparametric time trends. In many applications, practitioners are interested in whether the observed time series has a time trend at all, that is, whether the trend function is non-constant. Moreover, they would like to get further information about the shape of the trend function. Among other things, they would like to know in which time regions there is an upward/downward movement in the trend. When multiple time series are observed, another important question is whether the observed time series all have the same time trend. We design multiscale tests to formally approach these questions. We derive asymptotic theory for the proposed tests and investigate their finite sample performance by means of simulations. In addition, we illustrate the methods by two applications to temperature data.

Key words: Multiscale statistics; nonparametric regression; time series errors; shape constraints; strong approximations; anti-concentration bounds.
AMS 2010 subject classifications: 62E20; 62G10; 62G20; 62M10.

1 The model

Before we proceed any further, we need to introduce some notation used throughout the paper. For a vector $\mathbf{v} = (v_1, \ldots, v_m) \in \mathbb{R}^m$, we write $|\mathbf{v}| = \big(\sum_{i=1}^m v_i^2\big)^{1/2}$. For a random vector $\mathbf{V}$, we define it's $\mathcal{L}^q, q > 1$ norm as $||\mathbf{V}||_q = \big(\mathbb{E}|\mathbf{V}|^q\big)^{1/q}$. For a patricular case $q = 2$, we write $||\mathbf{V}|| := ||\mathbf{V}||_2$.
Following Wu (2005), we define the *physical dependence measure* for the process $\mathbf{L}(\mathcal{F}_t)$ as the following:

$$\delta_q(\mathbf{L}, t) = ||\mathbf{L}(\mathcal{F}_t) - \mathbf{L}(\mathcal{F}_t')||_q,$$

where $\mathcal{F}_t = (\ldots, \epsilon_{-1}, \epsilon_0, \epsilon_1, \ldots, \epsilon_{t-1}, \epsilon_t)$ and $\mathcal{F}_t' = (\ldots, \epsilon_{-1}, \epsilon_0', \epsilon_1, \ldots, \epsilon_{t-1}, \epsilon_t)$ is a coupled process of $\mathcal{F}_t$ with ϵ_0' being an i.i.d. copy of ϵ_0.
The model setting is as follows. We observe a panel of n time series $\mathcal{Z}_i = \{(Y_{it}, \mathbf{X}_{it}) : 1 \le t \le T\}$ of length T for $1 \le i \le n$. Each time series $\mathcal{Z}_i$ satisfies the model equation

$$Y_{it} = \boldsymbol{\beta}_i^\top \mathbf{X}_{it} + m_i\Big(\frac{t}{T}\Big) + \alpha_i + \varepsilon_{it} \tag{1.1}$$

[1]Address: Bonn Graduate School of Economics, University of Bonn, 53113 Bonn, Germany. Email: marina.k@uni-bonn.de.
[2]Corresponding author. Address: Department of Economics and Hausdorff Center for Mathematics, University of Bonn, 53113 Bonn, Germany. Email: michael.vogt@uni-bonn.de.

for $1 \leq t \leq T$, where $\boldsymbol{\beta}_i$ is a $d \times 1$ vector of unknown parameters, $\mathbf{X}_{it}$ is a $d \times 1$ vector of individual covariates, m_i is an unknown nonparametric trend function defined on $[0, 1]$, α_i is a (deterministic or random) intercept term and $\mathcal{E}_i = \{\varepsilon_{it} : 1 \leq t \leq T\}$ is a zero-mean stationary error process. As usual in nonparametric regression, the trend functions m_i in model (1.1) depend on rescaled time t/T rather than on real time t; cp. Robinson (1989), Dahlhaus (1997) and Vogt and Linton (2014) for the use and some discussion of the rescaled time argument. The functions m_i are only identified up to an additive constant in model (1.1): One can reformulate the model as $Y_{it} = [m_i(t/T) + c_i] + \boldsymbol{\beta}_i^\top \mathbf{X}_{it} + [\alpha_i - c_i] + \varepsilon_{it}$, that is, one can freely shift additive constants c_i between the trend $m_i(t/T)$ and the error component α_i. In order to obtain identification, one may impose different normalization constraints on the trends m_i. One possibility is to normalize them such that $\int_0^1 m_i(u)du = 0$ for all i. In what follows, we take for granted that the trends m_i satisfy this constraint. The term α_i can also be regarded as an additional error component. In the econometrics literature, it is commonly called a fixed effect error term. It can be interpreted as capturing unobserved characteristics of the time series $\mathcal{Z}_i$ which remain constant over time. We allow the error terms α_i to be dependent across i in an arbitrary way. Hence, by including them in model equation (1.1), we allow the n time series $\mathcal{Z}_i$ in our panel to be correlated with each other. Whereas the terms α_i may be correlated, the error processes $\mathcal{E}_i$ are assumed to be independent across i. Technical conditions regarding the model are discussed further in this section.
Finally, note that throughout the paper, we restrict attention to the case where the number of time series n in model (1.1) is fixed. Extending our theoretical results to the case where n slowly grows with the sample size T is a possible topic for further research.

1.1 Assumptions

Each process $\mathcal{E}_i$ is supposed to satisfy the following conditions:

(C1) For each i the variables ε_{it} allow for the representation $\varepsilon_{it} = G_i(\ldots, \eta_{it-1}, \eta_{it})$, where η_{it} are i.i.d. random variables across t and $G_i : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ is a measurable function. Denote $\mathcal{J}_{it} = (\ldots, \eta_{it-2}, \eta_{it-1}, \eta_{it})$.

(C2) For all i it holds that $\mathbb{E}[\varepsilon_{it}] = 0$ and $\|\varepsilon_{it}\|_q < \infty$ for some $q > 4$.

Following Wu (2005), we impose conditions on the dependence structure of the error processes $\mathcal{E}_i$ in terms of the physical dependence measure $\delta_q(G_i, t)$. In particular, we assume the following:

(C3) Define $\Theta_{i,t,q} = \sum_{s \geq t} \delta_q(G_i, s)$ for $t \geq 0$. For each i it holds that $\Theta_{i,t,q} = O(t^{-\tau_q} (\log t)^{-A})$, where $A > \frac{2}{3}(1/q + 1 + \tau_q)$ and $\tau_q = \{q^2 - 4 + (q-2)\sqrt{q^2 + 20q + 4}\}/8q$.

The conditions (C1)–(C3) are fulfilled by a wide range of stationary processes $\mathcal{E}_i$. For a detailed discussion of these properties, see Khismatullina and Vogt (2018).
Regarding the independent variables $\mathbf{X}_{it}$, we need the following additional assumptions for each i:

(C4) The covariates $\mathbf{X}_{it}$ allow for the representation $\mathbf{X}_{it} = \mathbf{H}_i(\ldots, u_{it-1}, u_{it})$ with u_{it} being i.i.d. random variables and $\mathbf{H}_i := (H_{i1}, H_{i2}, \ldots, H_{id})^\top : \mathbb{R}^\mathbb{Z} \to \mathbb{R}^d$ is a measurable function such that $\mathbf{H}_i(\mathcal{U}_{it})$ is well defined. Denote $\mathcal{U}_{it} = (\ldots, u_{it-1}, u_{it})$.

(C5) Let N_i be the $d \times d$ matrix with kl-th entry $n_{i,kl} = \mathbb{E}[H_{ik}(\mathcal{U}_{i0})H_{il}(\mathcal{U}_{i0})]$. We assume that the smallest eigenvalue of N_i is strictly bigger than 0.

(C6) Let $\mathbb{E}[\mathbf{H}_i(\mathcal{U}_{i0})] = \mathbf{0}$ and $||\mathbf{H}_i(\mathcal{U}_{it})||_{q'} < \infty$ for $q' > \theta/2$.

(C7) $\sum_{s=0}^{\infty} \delta_4(\mathbf{H}_i, s) < \infty$.

(C8) $\sup_{m \geq 0}(m+1)^\alpha \sum_{t=m}^{\infty} \delta_{q'}(\mathbf{H}_i, t) < \infty$ for $\alpha > 1/2 - 1/q'$. Dependence adjusted norm!

To be able to prove the main theorems in Section 2, we need additional assumptions on the relationship between the covariates and the error process.

(C9) $\mathbf{X}_{it}$ (elementwise) and ε_{is} are uncorrelated for each $t, s \in \{1, \ldots, T\}$.

(C10) Let $\zeta_{i,t} = (u_{it}, \eta_{it})^\top$. Define $\mathcal{I}_{i,t} = (\ldots, \zeta_{i,t-1}, \zeta_{i,t})$ and $\mathbf{U}_i(\mathcal{I}_{i,t}) = \mathbf{H}_i(\mathcal{U}_{it})G_i(\mathcal{J}_{it})$. Then, $\sum_{s=0}^{\infty} \delta_2(\mathbf{U}_i, s) < \infty$.

2 Testing for equality of time trends

In this section, we adapt the multiscale method developed in Khismatullina and Vogt (2018) to the problem of comparison of the trend curves m_i in model (1.1). As we will see, the proposed multiscale method does not only allow to test whether the null hypothesis is violated. It also provides information on where violations occur. More specifically, it allows to identify, with a pre-specified confidence, (i) trend functions which are different from each other and (ii) time intervals where these trend functions differ.

2.1 Construction of the test statistic

In what follows, we describe the construction of the test statistic that adresses the question of comparing different trend curves. More specifically, we test the null hypothesis $H_0 : m_1 = m_2 = \ldots = m_n$ in model (1.1). We assume that all the trend functions $m_i(\cdot)$ are continuosly differentiable on $[0, 1]$.

It is obvious that if α_i and $\boldsymbol{\beta}_i$ are known, the problem of testing for the common time trend would be greatly simplified. That is, we would test $H_0 : m_1 = m_2 = \ldots = m_n$ in the model

$$\begin{aligned} Y_{it} - \alpha_i - \boldsymbol{\beta}_i^\top \mathbf{X}_{it} &=: Y_{it}^\circ \\ &= m_i\Big(\frac{t}{T}\Big) + \varepsilon_{it}, \end{aligned}$$

which is a standard nonparametric regression equation. The variables Y_{it}° are not observed since the intercept α_i and the coefficients $\boldsymbol{\beta}_i$ are not known. Given appropriate estimators $\widehat{\boldsymbol{\beta}}_i$ and $\widehat{\alpha}_i$, we can then consider

$$\widehat{Y}_{it} := Y_{it} - \widehat{\alpha}_i - \widehat{\boldsymbol{\beta}}_i^\top \mathbf{X}_{it} = (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top \mathbf{X}_{it} + m_i\Big(\frac{t}{T}\Big) + \big(\alpha_i - \widehat{\alpha}_i\big) + \varepsilon_{it}.$$

Then our unobserved variables Y_{it}° can be approximated by $\widehat{Y}_{it}$ and we compute our test statistic based on $\widehat{Y}_{it}$. In what follows, we assume that an estimator with the property that $\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i = O_P(T^{-1/2})$ is given. Details on one of the possible ways to construct $\widehat{\boldsymbol{\beta}}_i$ are deferred to Section 2.4.

Given $\widehat{\boldsymbol{\beta}}_i$, consider an appropriate estimator $\widehat{\alpha}_i$ for the intercept α_i calculated by

$$\begin{aligned} \widehat{\alpha}_i &= \frac{1}{T}\sum_{t=1}^T \big(Y_{it} - \widehat{\boldsymbol{\beta}}_i^\top \mathbf{X}_{it}\big) = \frac{1}{T}\sum_{t=1}^T \big(\boldsymbol{\beta}_i^\top \mathbf{X}_{it} - \widehat{\boldsymbol{\beta}}_i^\top \mathbf{X}_{it} + \alpha_i + m_i(t/T) + \varepsilon_{it}\big) = \qquad (2.1) \\ &= \big(\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i\big)^\top \frac{1}{T}\sum_{t=1}^T \mathbf{X}_{it} + \alpha_i + \frac{1}{T}\sum_{i=1}^T m_i(t/T) + \frac{1}{T}\sum_{i=1}^T \varepsilon_{it}. \end{aligned}$$

Note that $\frac{1}{T}\sum_{i=1}^T \varepsilon_{it} = O_P(T^{-1/2})$ and $\frac{1}{T}\sum_{i=1}^T m_i(t/T) = O(T^{-1})$ due to Lipschitz continuity of m_i and normalization $\int_0^1 m_i(u)du = 0$. Furthermore, $\frac{1}{T}\sum_{t=1}^T \mathbf{X}_{it} = O_P(1)$ by Chebyshev's inequality and $\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i = O_P(T^{-1/2})$. Plugging all these results together in (2.1), we get that $\widehat{\alpha}_i - \alpha_i = O_P(T^{-1/2})$.Thus, the unobserved variable $Y_{it}^\circ := Y_{it} - \boldsymbol{\beta}_i^\top \mathbf{X}_{it} - \alpha_i = m_i(t/T) + \varepsilon_{it}$ can be well approximated by $\widehat{Y}_{it} = Y_{it} - \widehat{\alpha}_i - \widehat{\boldsymbol{\beta}}_i^\top \mathbf{X}_{it} = Y_{it}^\circ + O_P(T^{-1/2})$.

We further let $\widehat{\sigma}_i^2$ be an estimator of the long-run error variance $\sigma_i^2 = \sum_{\ell=-\infty}^{\infty} \mathrm{Cov}(\varepsilon_{i0}, \varepsilon_{i\ell})$ which is computed from the constructed sample $\{\widehat{Y}_{it} : 1 \le t \le T\}$. We thus regard $\widehat{\sigma}_i^2 = \widehat{\sigma}_i^2(\widehat{Y}_{i1}, \ldots, \widehat{Y}_{iT})$ as a function of the variables $\widehat{Y}_{it}$ for $1 \le t \le T$. Throughout the section, we assume that $\widehat{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$ with $\rho_T = o(1/\log T)$. Details on how to construct estimators of σ_i^2 are deferred to Section **??**.

We are now ready to introduce the multiscale statistic for testing the hypothesis $H_0 : m_1 = m_2 = \ldots = m_n$. For any pair of time series i and j, we define the kernel averages

$$\widehat{\psi}_{ij,T}(u,h) = \sum_{t=1}^T w_{t,T}(u,h)(\widehat{Y}_{it} - \widehat{Y}_{jt}),$$

where $w_{t,T}(u,h)$ are the local linear kernel weights calculated by the following formula.

$$w_{t,T}(u,h) = \frac{\Lambda_{t,T}(u,h)}{\{\sum_{t=1}^T \Lambda_{t,T}(u,h)^2\}^{1/2}}, \qquad (2.2)$$

where

$$\Lambda_{t,T}(u,h) = K\Big(\frac{\frac{t}{T}-u}{h}\Big)\Big[S_{T,2}(u,h) - \Big(\frac{\frac{t}{T}-u}{h}\Big)S_{T,1}(u,h)\Big],$$

$S_{T,\ell}(u,h) = (Th)^{-1}\sum_{t=1}^{T} K(\frac{\frac{t}{T}-u}{h})(\frac{\frac{t}{T}-u}{h})^{\ell}$ for $\ell = 0,1,2$ and K is a kernel function with the following properties:

(C11) The kernel K is non-negative, symmetric about zero and integrates to one. Moreover, it has compact support $[-1,1]$ and is Lipschitz continuous, that is, $|K(v) - K(w)| \le C|v-w|$ for any $v,w \in \mathbb{R}$ and some constant $C > 0$.

The kernel average $\widehat{\psi}_{ij,T}(u,h)$ can be regarded as measuring the distance between the two trend curves m_i and m_j on the interval $[u-h, u+h]$.
We now combine the test statistics $\widehat{\psi}_{ij,T}(u,h)$ for a wide range of different locations u and bandwidths or scales h in a following way:

$$\widehat{\Psi}_{ij,T} = \max_{(u,h)\in\mathcal{G}_T} \Big\{\Big|\frac{\widehat{\psi}_{ij,T}(u,h)}{(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2)^{1/2}}\Big| - \lambda(h)\Big\},$$

where $\lambda(h) = \sqrt{2\log\{1/(2h)\}}$ and the set $\mathcal{G}_T$ is the set of points (u,h) that are taken into consideration. The statistic $\widehat{\Psi}_{ij,T}$ can be interpreted as a global distance measure between the two curves m_i and m_j. Thus, the multiscale statistic $\widehat{\Psi}_{ij,T}$ simultaneously takes into account all locations u and bandwidths h with $(u,h) \in \mathcal{G}_T$. Throughout the paper, we suppose that $\mathcal{G}_T$ is some subset of $\mathcal{G}_T^{\text{full}} = \{(u,h) : u = t/T \text{ for some } 1 \le t \le T \text{ and } h \in [h_{\min}, h_{\max}]\}$, where $h_{\min}$ and $h_{\max}$ denote some minimal and maximal bandwidth value, respectively. For our theory to work, we require the following conditions to hold:

(C12) $|\mathcal{G}_T| = O(T^{\theta})$ for some arbitrarily large but fixed constant $\theta > 0$, where $|\mathcal{G}_T|$ denotes the cardinality of $\mathcal{G}_T$.

(C13) $h_{\min} \gg T^{-(1-\frac{2}{q})}\log T$, that is, $h_{\min}/\{T^{-(1-\frac{2}{q})}\log T\} \to \infty$ with $q > 4$ defined in (C2) and $h_{\max} < 1/2$.

We finally define the multiscale statistic for testing the null hypothesis $H_0 : m_1 = m_2 = \ldots = m_n$ as

$$\widehat{\Psi}_{n,T} = \max_{1\le i<j\le n} \widehat{\Psi}_{ij,T},$$

that is, we define it as the maximal distance $\widehat{\Psi}_{ij,T}$ between any pair of curves m_i and m_j with $i \neq j$.

2.2 The test procedure

Let Z_{it} for $1 \le t \le T$ and $1 \le i \le n$ be independent standard normal random variables which are independent of the error terms ε_{it} and the covariates $\mathbf{X}_{it}$. Denote the empirical average of the variables $Z_{i1}, \ldots, Z_{iT}$ by $\bar{Z}_{i,T} = T^{-1} \sum_{t=1}^{T} Z_{it}$. To simplify notation, we write $\bar{Z}_i = \bar{Z}_{i,T}$ in what follows. For each i and j, we introduce the Gaussian statistic

$$\Phi_{ij,T} = \max_{(u,h)\in\mathcal{G}_T} \Big\{ \Big| \frac{\phi_{ij,T}(u,h)}{(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2)^{1/2}} \Big| - \lambda(h) \Big\},$$

where $\phi_{ij,T}(u,h) = \sum_{t=1}^{T} w_{t,T}(u,h) \left\{ \widehat{\sigma}_i (Z_{it} - \bar{Z}_i) - \widehat{\sigma}_j (Z_{jt} - \bar{Z}_j) \right\}$. Moreover, we define the statistic

$$\Phi_{n,T} = \max_{1\le i<j\le n} \Phi_{ij,T} \tag{2.3}$$

and denote its $(1-\alpha)$-quantile by $q_{n,T}(\alpha)$. Our multiscale test of the hypothesis $H_0 : m_1 = m_2 = \ldots = m_n$ is defined as follows: For a given significance level $\alpha \in (0,1)$, we reject H_0 if $\widehat{\Psi}_{n,T} > q_{n,T}(\alpha)$.

2.3 Theoretical properties of the test

To start with, we introduce the auxiliary statistic

$$\widehat{\Phi}_{n,T} = \max_{1\le i<j\le n} \widehat{\Phi}_{ij,T}, \tag{2.4}$$

where

$$\widehat{\Phi}_{ij,T} = \max_{(u,h)\in\mathcal{G}_T} \Big\{ \Big| \frac{\widehat{\phi}_{ij,T}(u,h)}{\{\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\}^{1/2}} \Big| - \lambda(h) \Big\}$$

and $\widehat{\phi}_{ij,T}(u,h) = \sum_{t=1}^{T} w_{t,T}(u,h) \big\{ (\varepsilon_{it} - \bar{\varepsilon}_i) + (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top (\mathbf{X}_{it} - \bar{\mathbf{X}}_i) - (\varepsilon_{jt} - \bar{\varepsilon}_j) - -(\boldsymbol{\beta}_j - \widehat{\boldsymbol{\beta}}_j)^\top (\mathbf{X}_{jt} - \bar{\mathbf{X}}_j) \big\}$ with $\bar{\varepsilon}_i = \bar{\varepsilon}_{i,T} = T^{-1} \sum_{t=1}^{T} \varepsilon_{it}$ and $\bar{\mathbf{X}}_i = \bar{\mathbf{X}}_{i,T} = T^{-1} \sum_{t=1}^{T} \mathbf{X}_{it}$ respectively. Our first theoretical result characterizes the asymptotic behaviour of the statistic $\widehat{\Phi}_{n,T}$.

Theorem 2.1. *Suppose that the error processes $\mathcal{E}_i = \{\varepsilon_{it} : 1 \le t \le T\}$ are independent across i and satisfy (C1)–(C3) for each i. Moreover, let (C4)–(C13) be fulfilled and assume that $\widehat{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$ with $\rho_T = o(1/\log T)$ for each i. Then*

$$\mathbb{P}\big(\widehat{\Phi}_{n,T} \le q_{n,T}(\alpha)\big) = (1-\alpha) + o(1)$$

Theorem 2.1 is the main stepping stone to derive the theoretical properties of our multiscale test. The proof of the theorem is provided in the Section **??**.

2.4 Estimation of the parameters $\boldsymbol{\beta}_i$

We now focus on finding an appropriate estimator $\widehat{\boldsymbol{\beta}}_i$ of $\boldsymbol{\beta}_i$. For that purpose, for each i we consider the time series $\{\Delta Y_{it} : 2 \le t \le T\}$ of the differences $\Delta Y_{it} = Y_{it} - Y_{it-1}$. We then have

$$\Delta Y_{it} = Y_{it} - Y_{it-1} = \boldsymbol{\beta}_i^\top \Delta \mathbf{X}_{it} + \bigg(m_i\Big(\frac{t}{T}\Big) - m_i\Big(\frac{t-1}{T}\Big) \bigg) + \Delta\varepsilon_{it},$$

where $\Delta \mathbf{X}_{it} = \mathbf{X}_{it} - \mathbf{X}_{it-1}$ and $\Delta \varepsilon_{it} = \varepsilon_{it} - \varepsilon_{it-1}$. Since $m_i(\cdot)$ is Lipschitz, we can use the fact that $\left|m_i\left(\frac{t}{T}\right) - m_i\left(\frac{t-1}{T}\right)\right| = O\left(\frac{1}{T}\right)$ and rewrite

$$\Delta Y_{it} = \boldsymbol{\beta}_i^\top \Delta \mathbf{X}_{it} + \Delta \varepsilon_{it} + O\Big(\frac{1}{T}\Big). \tag{2.5}$$

In particular, for each i we employ the least squares estimation method to estimate $\boldsymbol{\beta}_i$ in (2.5), treating $\Delta \mathbf{X}_{it}$ as the regressors and ΔY_{it} as the response variable. That is, we propose the following differencing estimator:

$$\widehat{\boldsymbol{\beta}}_i = \Big(\sum_{t=2}^{T} \Delta \mathbf{X}_{it} \Delta \mathbf{X}_{it}^\top\Big)^{-1} \sum_{t=2}^{T} \Delta \mathbf{X}_{it} \Delta Y_{it} \tag{2.6}$$

Then the asymptotic consistency for this differencing estimator is given by the following theorem:

Theorem 2.2. *Under Assumptions (C1) - (C10), we have*

$$\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i = O_P\Big(\frac{1}{\sqrt{T}}\Big),$$

where $\widehat{\boldsymbol{\beta}}_i$ is the differencing estimator given by (2.6).

3 Appendix

In this section, we prove the theoretical results from Section 2. We use the following notation: The symbol C denotes a universal real constant which may take a different value on each occurrence. For $a, b \in \mathbb{R}$, we write $a_+ = \max\{0, a\}$ and $a \vee b = \max\{a, b\}$. For any set A, the symbol $|A|$ denotes the cardinality of A. The notation $X \stackrel{\mathcal{D}}{=} Y$ means that the two random variables X and Y have the same distribution. Finally, $f_0(\cdot)$ and $F_0(\cdot)$ denote the density and the distribution function of the standard normal distribution, respectively.

3.1 Proof of Theorem 2.1

We will build the proof of Theorem 2.1 on the auxiliary results derived below. The steps of the proof are as follows.

- First, by Proposition A.1, there exist statistics $\widetilde{\Phi}_{n,T}$ for $T = 1, 2, \ldots$ which are distributed as $\widehat{\widehat{\Phi}}_{n,T}$ for any $T \geq 1$ and which have the property that

$$\big|\widetilde{\Phi}_{n,T} - \Phi_{n,T}\big| = o_p\Big(\frac{T^{1/q}}{\sqrt{T h_{\min}}} + \frac{\rho_T}{\sqrt{h_{\min}}}\sqrt{\log T}\Big), \tag{3.1}$$

  where $\Phi_{n,T}$ is a Gaussian statistic as defined in (2.3). The approximation result (3.1) allows us to replace the multiscale statistic $\widehat{\widehat{\Phi}}_{n,T}$ by an identically distributed version $\widetilde{\Phi}_{n,T}$ which is close to the Gaussian statistic $\Phi_{n,T}$.

- Second, by Proposition A.3 we show that $\Phi_{n,T}$ does not concentrate too strongly in small regions of the form $[x-\delta_T, x+\delta_T]$ with δ_T converging to zero. Or, in other words, it holds that

$$\sup_{x\in\mathbb{R}} \mathbb{P}\big(|\Phi_{n,T}-x| \le \delta_T\big) = o(1), \tag{3.2}$$

where $\delta_T = T^{1/q}/\sqrt{Th_{\min}} + \rho_T\sqrt{\log T}/\sqrt{h_{\min}}$.

- In the third step we make use of Lemma A.4 to show that

$$\sup_{x\in\mathbb{R}} \big|\mathbb{P}(\widehat{\widehat{\Phi}}_{n,T} \le x) - \mathbb{P}(\Phi_{n,T} \le x)\big| = o(1). \tag{3.3}$$

This statement directly follows from the previous two steps and the fact that $\widetilde{\Phi}_{n,T}$ is distributed as $\widehat{\widehat{\Phi}}_{n,T}$ for any $n \ge 2$, $T \ge 1$.

- And finally, in Proposition A.7 we prove that

$$\sup_{x\in\mathbb{R}} \big|\mathbb{P}(\widehat{\Phi}_{n,T} \le x) - \mathbb{P}(\Phi_{n,T} \le x)\big| = o(1), \tag{3.4}$$

which immediately implies the statement of Theorem 2.1.

Introducing additional intermediate statistic

The auxiliary statistic $\widehat{\Phi}_{n,T}$ defined in Section 2.3 (which is basically the version of the main statistic $\widehat{\Psi}_{n,T}$ under the null hypothesis) heavily depends on the known covariates $\mathbf{X}_{it}$, whereas the Gaussian version $\Phi_{n,T}$ does not. This is the reason why we need to introduce additional intermediate test statistic that connects $\widehat{\Phi}_{n,T}$ and $\Phi_{n,T}$. Thus, for each i and j, consider the following kernel averages:

$$\widehat{\widehat{\phi}}_{ij,T}(u,h) = \sum_{t=1}^{T} w_{t,T}(u,h)\big\{(\varepsilon_{it}-\bar{\varepsilon}_i) - (\varepsilon_{jt}-\bar{\varepsilon}_j)\big\}.$$

The intermediate statistic is then defined as

$$\begin{aligned} \widehat{\widehat{\Phi}}_{n,T} &= \max_{1\le i<j\le n} \widehat{\widehat{\Phi}}_{ij,T} \quad \text{with} \\ \widehat{\widehat{\Phi}}_{ij,T} &= \max_{(u,h)\in\mathcal{G}_T} \Big\{\Big|\frac{\widehat{\widehat{\phi}}_{ij,T}(u,h)}{(\widehat{\sigma}_i^2+\widehat{\sigma}_j^2)^{1/2}}\Big| - \lambda(h)\Big\}. \end{aligned} \tag{3.5}$$

This statistic can thus be regarded as an approximation of the statistic $\widehat{\Phi}_{n,T}$. Heuristically, the kernel averages $\widehat{\widehat{\phi}}_{ij,T}(u,h)$ are close to the kernel averages $\widehat{\phi}_{ij,T}(u,h)$ because of the properties of our estimators $\widehat{\boldsymbol{\beta}}_i$ and assumptions on $\mathbf{X}_{it}$ but we will prove it later in this Section.

Auxiliary results using strong approximation theory

The main purpose of this section is to prove that there is a version of the multiscale statistic $\widehat{\widehat{\Phi}}_{n,T}$ defined in (3.5) which is close to a Gaussian statistic whose distribution is known. More specifically, we prove the following result.

Proposition A.1. *Under the conditions of Theorem 2.1, there exist statistics* $\widetilde{\Phi}_{n,T}$ *for* $T = 1, 2, \dots$ *with the following two properties: (i)* $\widetilde{\Phi}_{n,T}$ *has the same distribution as* $\widehat{\widehat{\Phi}}_{n,T}$ *for any* T*, and (ii)*

$$\big|\widetilde{\Phi}_{n,T} - \Phi_{n,T}\big| = o_p\Big(\frac{T^{1/q}}{\sqrt{Th_{\min}}} + \frac{\rho_T}{\sqrt{h_{\min}}}\sqrt{\log T}\Big),$$

where $\Phi_{n,T}$ *is a Gaussian statistic as defined in* (2.3)*.*

Proof of Proposition A.1. For the proof, we draw on strong approximation theory for each stationary process $\mathcal{E}_i = \{\varepsilon_{it} : 1 \le t \le T\}$ that fulfill the conditions (C1)–(C3). By Theorem 2.1 and Corollary 2.1 in Berkes et al. (2014), the following strong approximation result holds true: On a richer probability space, there exists a standard Brownian motion $\mathbb{B}$ and a sequence $\{\widetilde{\varepsilon}_t : t \in \mathbb{N}\}$ such that $[\widetilde{\varepsilon}_1, \dots, \widetilde{\varepsilon}_T] \stackrel{\mathcal{D}}{=} [\varepsilon_1, \dots, \varepsilon_T]$ for each T and

$$\max_{1 \le t \le T}\Big|\sum_{s=1}^{t} \widetilde{\varepsilon}_s - \sigma\mathbb{B}(t)\Big| = o\big(T^{1/q}\big) \quad \text{a.s.,} \tag{3.6}$$

where $\sigma^2 = \sum_{k \in \mathbb{Z}} \text{Cov}(\varepsilon_0, \varepsilon_k)$ denotes the long-run error variance.
We apply this result for each stationary process $\mathcal{E}_i = \{\varepsilon_{it} : 1 \le t \le T\}$ separately. In order to do this, we define

$$\widetilde{\Phi}_{n,T} = \max_{1 \le i < j \le n} \widetilde{\Phi}_{ij,T}, \tag{3.7}$$

where

$$\widetilde{\Phi}_{ij,T} = \max_{(u,h) \in \mathcal{G}_T}\Big\{\Big|\frac{\widetilde{\phi}_{ij,T}(u,h)}{\{\widetilde{\sigma}_i^2 + \widetilde{\sigma}_j^2\}^{1/2}}\Big| - \lambda(h)\Big\},$$

where $\widetilde{\phi}_{ij,T}(u,h) = \sum_{t=1}^{T} w_{t,T}(u,h)\big\{(\widetilde{\varepsilon}_{it} - \bar{\widetilde{\varepsilon}}_i) - (\widetilde{\varepsilon}_{jt} - \bar{\widetilde{\varepsilon}}_j)\big\}$ and $\widetilde{\sigma}_i^2$ are the same estimators as $\widehat{\sigma}_i^2$ with $\widehat{Y}_{it} = (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top \mathbf{X}_{it} + m_i(t/T) + \big(\alpha_i - \widehat{\alpha}_i\big) + \varepsilon_{it}$ replaced by $\widetilde{Y}_{it} = = (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top \mathbf{X}_{it} + m_i(t/T) + \big(\alpha_i - \widehat{\alpha}_i\big) + \widetilde{\varepsilon}_{it}$ for $1 \le t \le T$. In addition, we let

$$\Phi_{n,T} = \max_{1 \le i < j \le n} \Phi_{ij,T} = \max_{1 \le i < j \le n} \max_{(u,h) \in \mathcal{G}_T}\Big\{\Big|\frac{\phi_{ij,T}(u,h)}{\{\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\}^{1/2}}\Big| - \lambda(h)\Big\} \text{ as in (2.3),}$$

$$\text{and } \Phi^\diamond_{n,T} = \max_{1 \le i < j \le n} \Phi^\diamond_{ij,T} = \max_{1 \le i < j \le n} \max_{(u,h) \in \mathcal{G}_T}\Big\{\Big|\frac{\phi_{ij,T}(u,h)}{\{\widetilde{\sigma}_i^2 + \widetilde{\sigma}_j^2\}^{1/2}}\Big| - \lambda(h)\Big\}$$

with

$$\phi_{ij,T}(u,h) = \sum\nolimits_{t=1}^{T} w_{t,T}(u,h)\big\{\widehat{\sigma}_i(Z_{it} - \bar{Z}_i) - \widehat{\sigma}_j(Z_{jt} - \bar{Z}_j)\big\}$$

and $Z_{it} = \mathbb{B}_i(t) - \mathbb{B}_i(t-1)$. With this notation, we can write

$$\big|\widetilde{\Phi}_{n,T} - \Phi_{n,T}\big| \le \big|\widetilde{\Phi}_{n,T} - \Phi^{\diamond}_{n,T}\big| + \big|\Phi^{\diamond}_{n,T} - \Phi_{n,T}\big|. \tag{3.8}$$

First consider $|\widetilde{\Phi}_{n,T} - \Phi^{\diamond}_{n,T}|$. Straightforward calculations yield that

$$\big|\widetilde{\Phi}_{n,T} - \Phi^{\diamond}_{n,T}\big| \le \max_{1\le i<j\le n} \Big(\big\{\widetilde{\sigma}_i^2 + \widetilde{\sigma}_j^2\big\}^{-1/2} \max_{(u,h)\in\mathcal{G}_T} \big|\widetilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h)\big| \Big). \tag{3.9}$$

Using summation by parts, ($\sum_{i=1}^n a_i b_i = \sum_{i=1}^{n-1} A_i(b_i - b_{i+1}) + A_n b_n$ with $A_j = \sum_{j=1}^i a_j$) we further obtain that

$$\begin{aligned}
&\big|\widetilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h)\big| = \\
&\qquad = \Big|\sum_{t=1}^{T} w_{t,T}(u,h)\big\{(\widetilde{\varepsilon}_{it} - \bar{\widetilde{\varepsilon}}_i) - (\widetilde{\varepsilon}_{jt} - \bar{\widetilde{\varepsilon}}_j) - \widehat{\sigma}_i(Z_{it} - \bar{Z}_i) + \widehat{\sigma}_j(Z_{jt} - \bar{Z}_j)\big\}\Big| = \\
&\qquad = \Big|\sum_{t=1}^{T-1} A_{ij,t}\big(w_{t,T}(u,h) - w_{t+1,T}(u,h)\big) + A_{ij,T} w_{T,T}(u,h)\Big|,
\end{aligned}$$

where

$$A_{ij,t} = \sum_{s=1}^{t} \big\{(\widetilde{\varepsilon}_{is} - \bar{\widetilde{\varepsilon}}_i) - (\widetilde{\varepsilon}_{js} - \bar{\widetilde{\varepsilon}}_j) - \widehat{\sigma}_i(Z_{it} - \bar{Z}_i) + \widehat{\sigma}_j(Z_{jt} - \bar{Z}_j)\big\}.$$

Note that by construction $A_{ij,T} = 0$ for all pairs (i,j). Denoting

$$W_T(u,h) = \sum_{t=1}^{T-1} |w_{t+1,T}(u,h) - w_{t,T}(u,h)|,$$

we have

$$\big|\widetilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h)\big| = \Big|\sum_{t=1}^{T-1} A_{ij,t}\big(w_{t,T}(u,h) - w_{t+1,T}(u,h)\big)\Big| \le W_T(u,h) \max_{1\le t\le T} |A_{ij,t}|. \tag{3.10}$$

Now consider $\max_{1\le t\le T} |A_{ij,t}|$:

$$\begin{aligned}
\max_{1\le t\le T} |A_{ij,t}| &\le \max_{1\le t\le T} \Big| \sum_{s=1}^{t} (\widetilde{\varepsilon}_{is} - \bar{\widetilde{\varepsilon}}_i) - \widehat{\sigma}_i \sum_{s=1}^{t} \big\{Z_{is} - \bar{Z}_i\big\}\Big| + \\
&\quad + \max_{1\le t\le T} \Big| \sum_{s=1}^{t} (\widetilde{\varepsilon}_{js} - \bar{\widetilde{\varepsilon}}_j) - \widehat{\sigma}_j \sum_{s=1}^{t} \big\{Z_{js} - \bar{Z}_j\big\}\Big| \le \\
&\le \max_{1\le t\le T} \Big| \sum_{s=1}^{t} \widetilde{\varepsilon}_{is} - \widehat{\sigma}_i \sum_{s=1}^{t} Z_{is}\Big| + \max_{1\le t\le T} \Big| t(\bar{\widetilde{\varepsilon}}_i - \widehat{\sigma}_i \bar{Z}_i)\Big| + \\
&\quad + \max_{1\le t\le T} \Big| \sum_{s=1}^{t} \widetilde{\varepsilon}_{js} - \widehat{\sigma}_j \sum_{s=1}^{t} Z_{js}\Big| + \max_{1\le t\le T} \Big| t(\bar{\widetilde{\varepsilon}}_j - \widehat{\sigma}_j \bar{Z}_j)\Big| =
\end{aligned}$$

$$
\begin{aligned}
&= \max_{1\le t\le T}\Big|\sum_{s=1}^{t}\widetilde{\varepsilon}_{is} - \widehat{\sigma}_i\sum_{s=1}^{t} Z_{is}\Big| + T\Big|\frac{1}{T}\Big(\sum_{s=1}^{T}\widetilde{\varepsilon}_{is} - \widehat{\sigma}_i\sum_{s=1}^{T} Z_{is}\Big)\Big| + \\
&\quad + \max_{1\le t\le T}\Big|\sum_{s=1}^{t}\widetilde{\varepsilon}_{js} - \widehat{\sigma}_j\sum_{s=1}^{t} Z_{js}\Big| + T\Big|\frac{1}{T}\Big(\sum_{s=1}^{T}\widetilde{\varepsilon}_{js} - \widehat{\sigma}_j\sum_{s=1}^{T} Z_{js}\Big)\Big| \le \\
&\le 2\max_{1\le t\le T}\Big|\sum_{s=1}^{t}\widetilde{\varepsilon}_{is} - \widehat{\sigma}_i\sum_{s=1}^{t} Z_{is}\Big| + 2\max_{1\le t\le T}\Big|\sum_{s=1}^{t}\widetilde{\varepsilon}_{js} - \widehat{\sigma}_j\sum_{s=1}^{t} Z_{js}\Big| = \\
&= 2\max_{1\le t\le T}\Big|\sum_{s=1}^{t}\widetilde{\varepsilon}_{is} - \widehat{\sigma}_i\sum_{s=1}^{t}\big(\mathbb{B}_i(s) - \mathbb{B}_i(s-1)\big)\Big| + \\
&\quad + 2\max_{1\le t\le T}\Big|\sum_{s=1}^{t}\widetilde{\varepsilon}_{js} - \widehat{\sigma}_j\sum_{s=1}^{t}\big(\mathbb{B}_j(s) - \mathbb{B}_j(s-1)\big)\Big| = \\
&= 2\max_{1\le t\le T}\Big|\sum_{s=1}^{t}\widetilde{\varepsilon}_{is} - \widehat{\sigma}_i\mathbb{B}_i(t)\Big| + 2\max_{1\le t\le T}\Big|\sum_{s=1}^{t}\widetilde{\varepsilon}_{js} - \widehat{\sigma}_j\mathbb{B}_j(t)\Big| \le \\
&\le 2\max_{1\le t\le T}\Big|\sum_{s=1}^{t}\widetilde{\varepsilon}_{is} - \sigma_i\mathbb{B}_i(t)\Big| + 2\max_{1\le t\le T}\Big|\sigma_i\mathbb{B}_i(t) - \widehat{\sigma}_i\mathbb{B}_i(t)\Big| + \\
&\quad + 2\max_{1\le t\le T}\Big|\sum_{s=1}^{t}\widetilde{\varepsilon}_{js} - \sigma_j\mathbb{B}_j(t)\Big| + 2\max_{1\le t\le T}\Big|\sigma_j\mathbb{B}_j(t) - \widehat{\sigma}_j\mathbb{B}_j(t)\Big|
\end{aligned}
$$

Note that for all $1 \le i \le n$ we have $\widehat{\sigma}_i^2 - \sigma_i^2 = o_P(\rho_T)$ which leads to $\widehat{\sigma}_i - \sigma_i = o_P(\rho_T)$. Furthermore, regarding the standard Brownian motion $\mathbb{B}_i(t)$ as a sum of standard normal random variables, we get the following result:

$$
\begin{aligned}
\max_{1\le t\le T}|\mathbb{B}_i(t)| &= \max_{1\le t\le T}\Big|\sum_{s=1}^{t} Z_{is}\Big| = \max_{1\le t\le T}\Big|\sqrt{t}\frac{1}{\sqrt{t}}\sum_{s=1}^{t} Z_{is}\Big| \le \\
&\le \max_{1\le t\le T}\sqrt{t}\max_{1\le t\le T}\Big|\frac{1}{\sqrt{t}}\sum_{s=1}^{t} Z_{is}\Big| = \sqrt{T}\max_{1\le t\le T}\Big|\frac{1}{\sqrt{t}}\sum_{s=1}^{t} Z_{is}\Big| = O_P(\sqrt{T\log T}),
\end{aligned}
$$

where we used the bound for the maximum of T standard normal random variables and the fact that for each t, $1 \le t \le T$, $\frac{1}{\sqrt{t}}\sum_{s=1}^{t} Z_{is}$ is a standard normal random variable. Therefore, applying the strong approximation result (3.6), we can infer that

$$
\begin{aligned}
\max_{1\le t\le T}|A_{ij,t}| \le & 2\max_{1\le t\le T}\Big|\sum_{s=1}^{t}\widetilde{\varepsilon}_{is} - \sigma_i\mathbb{B}_i(t)\Big| + 2\big|\sigma_i - \widehat{\sigma}_i\big|\max_{1\le t\le T}\big|\mathbb{B}_i(t)\big| + \\
& + 2\max_{1\le t\le T}\Big|\sum_{s=1}^{t}\widetilde{\varepsilon}_{js} - \sigma_j\mathbb{B}_j(t)\Big| + 2\big|\sigma_j - \widehat{\sigma}_j\big|\max_{1\le t\le T}\big|\mathbb{B}_j(t)\big| = \\
= & o_P\big(T^{1/q} + \rho_T\sqrt{T\log T}\big). \qquad (3.11)
\end{aligned}
$$

Standard arguments show that $\max_{(u,h)\in\mathcal{G}_T} W_T(u,h) = O(1/\sqrt{Th_{\min}})$. Plugging (3.11) in (3.10) and then in (3.9), we can thus infer that

$$\big|\widetilde{\Phi}_{n,T} - \Phi^{\diamond}_{n,T}\big| \le \{\widetilde{\sigma}_i^2 + \widetilde{\sigma}_j^2\}^{-1/2} \max_{(u,h)\in\mathcal{G}_T} W_T(u,h) \max_{1\le i<j\le n} \max_{1\le t\le T} |A_{ij,t}| =$$

$$= o_P\Big(\frac{T^{1/q}}{\sqrt{Th_{\min}}} + \frac{\rho_T}{\sqrt{h_{\min}}}\sqrt{\log T}\Big). \tag{3.12}$$

Now consider $|\Phi^{\diamond}_{n,T} - \Phi_{n,T}|$. Since $\phi_{ij,T}(u,h)$ is distributed as $N(0, \widehat{\sigma}_i^2 + \widehat{\sigma}_j^2)$ for all $(u,h) \in \mathcal{G}_T$ and all $1 \le i < j \le n$, $|\mathcal{G}_T| = O(T^\theta)$ for some large but fixed constant θ by Assumption (C12), n is fixed and $\widehat{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$ as well as $\widetilde{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$, we can establish that

$$\big|\Phi^{\diamond}_{n,T} - \Phi_{n,T}\big| \le \max_{1\le i<j\le n} \max_{(u,h)\in\mathcal{G}_T} \Big|\frac{\phi_{ij,T}(u,h)}{\{\widetilde{\sigma}_i^2 + \widetilde{\sigma}_j^2\}^{1/2}} - \frac{\phi_{ij,T}(u,h)}{\{\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\}^{1/2}}\Big| = o_P(\rho_T\sqrt{\log T}). \tag{3.13}$$

Plugging (3.12) and (3.13) in (3.8) completes the proof. □

Auxiliary results using anti-concentration bounds

In this section, we establish some properties of the Gaussian statistic $\Phi_{n,T}$ defined in (2.3). We in particular show that $\Phi_{n,T}$ does not concentrate too strongly in small regions of the form $[x - \delta_T, x + \delta_T]$ with δ_T converging to zero.
The main technical tool for proving Proposition A.3 are anti-concentration bounds for Gaussian random vectors. The following proposition slightly generalizes anti-concentration results derived in Chernozhukov et al. (2015), in particular Theorem 3 therein.

Proposition A.2. *Let $(X_1, \ldots, X_p)^\top$ be a Gaussian random vector in $\mathbb{R}^p$ with $\mathbb{E}[X_j] = \mu_j$ and $\mathrm{Var}(X_j) = \sigma_j^2 > 0$ for $1 \le j \le p$. Define $\overline{\mu} = \max_{1\le j\le p}|\mu_j|$ together with $\underline{\sigma} = \min_{1\le j\le p}\sigma_j$ and $\overline{\sigma} = \max_{1\le j\le p}\sigma_j$. Moreover, set $a_p = \mathbb{E}[\max_{1\le j\le p}(X_j - \mu_j)/\sigma_j]$ and $b_p = \mathbb{E}[\max_{1\le j\le p}(X_j - \mu_j)]$. For every $\delta > 0$, it holds that*

$$\sup_{x\in\mathbb{R}} \mathbb{P}\Big(\big|\max_{1\le j\le p} X_j - x\big| \le \delta\Big) \le C\delta\big\{\overline{\mu} + a_p + b_p + \sqrt{1 \vee \log(\underline{\sigma}/\delta)}\big\},$$

where $C > 0$ depends only on $\underline{\sigma}$ and $\overline{\sigma}$.

The proof of Proposition A.2 is provided in Khismatullina and Vogt (2018).

Proposition A.3. *Under the conditions of Theorem 2.1, it holds that*

$$\sup_{x\in\mathbb{R}} \mathbb{P}\big(|\Phi_{n,T} - x| \le \delta_T\big) = o(1),$$

where $\delta_T = T^{1/q}/\sqrt{Th_{\min}} + \rho_T\sqrt{\log T}/\sqrt{h_{\min}}$.

Proof of Proposition A.3. We write $x = (u, h)$ along with $\mathcal{G}_T = \{x : x \in \mathcal{G}_T\} = \{x_1, \ldots, x_p\}$, where $p := |\mathcal{G}_T| \le O(T^\theta)$ for some large but fixed $\theta > 0$ by our assumptions. Moreover, for $k = 1, \ldots, p$, we set

$$U_{ij,2k-1} = \frac{\phi_{ij,T}(x_{k1}, x_{k2})}{\{\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\}^{1/2}} - \lambda(x_{k2})$$
$$U_{ij,2k} = -\frac{\phi_{ij,T}(x_{k1}, x_{k2})}{\{\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\}^{1/2}} - \lambda(x_{k2})$$

with $x_k = (x_{k1}, x_{k2})$. This notation allows us to write

$$\Phi_{n,T} = \max_{1 \le i < j \le n} \max_{1 \le k \le 2p} U_{ij,k}, = \max_{1 \le l \le (n-1)np} U_l'$$

where $(U_1', \ldots, U_{(n-1)np}')^\top \in \mathbb{R}^{n(n-1)p}$ is a Gaussian random vector with the following properties: (i) $\mu_l := \mathbb{E}[U_l'] = \{\mathbb{E}[U_{ij,2k}] \text{ or } \mathbb{E}[U_{ij,2k-1}]\} = -\lambda(x_{k2})$ and thus

$$\overline{\mu} = \max_{1 \le l \le (n-1)np} |\mu_l| \le C\sqrt{\log T},$$

and (ii) $\sigma_l^2 := \text{Var}(U_l') = 1$ for all $1 \le 1 \le (n-1)np$. Hence, $a_{(n-1)np} = b_{(n-1)np}$. Moreover, as the variables $(U_l' - \mu_l)/\sigma_l$ are standard normal, we have that $a_{(n-1)np} = b_{(n-1)np} \le C\sqrt{\log((n-1)np)} \le C\sqrt{\log T}$. With this notation at hand, we can apply Proposition A.2 to obtain that

$$\sup_{x \in \mathbb{R}} \mathbb{P}\Big(\big|\Phi_T - x\big| \le \delta_T\Big) \le C\delta_T\Big[\sqrt{\log T} + \sqrt{\log(1/\delta_T)}\Big] = o(1)$$

with $\delta_T = T^{1/q}/\sqrt{Th_{\min}} + \rho_T\sqrt{\log T}/\sqrt{h_{\min}}$, which is the statement of Proposition A.3. $\square$

Other auxiliary results

Lemma A.4. *Let V_T and W_T be real-valued random variables for $T = 1, 2, \ldots$ such that $V_T - W_T = o_p(\delta_T)$ with some $\delta_T = o(1)$. If*

$$\sup_{x \in \mathbb{R}} \mathbb{P}(|V_T - x| \le \delta_T) = o(1), \tag{3.14}$$

then

$$\sup_{x \in \mathbb{R}} \big|\mathbb{P}(V_T \le x) - \mathbb{P}(W_T \le x)\big| = o(1). \tag{3.15}$$

Proof of this lemma is provided in Khismatullina and Vogt (2018).

Definition 3.1. *Dependence adjusted norm:* $||X.||_{q',\alpha}^{q'} = \sup_{m \ge 0}(m+1)^\alpha \sum_{t=m}^\infty \delta_{q'}(X, t)$.

Theorem A.1. *Wu et al. (2016) Assume that $||X.||_{q',\alpha}^{q'} < \infty$, where $q' > 2$ and $\alpha > 0$, and $\sum_{t=1}^T a_t^2 = T$. Moreover, assume that $\alpha > 1/2 - 1/q'$. Denote $S_T = a_1X_1 + \ldots + a_TX_T$. Then for all $x > 0$,*

$$\mathbb{P}(|S_T| \ge x) \le C_1 \frac{|a|_{q'}^{q'} ||X.||_{q',\alpha}^{q'}}{x^{q'}} + C_2 \exp\left(-\frac{C_3 x^2}{T||X.||_{2,\alpha}^2}\right),$$

wjere C_1, C_2, C_3 are constants that only depend on q' and α.

Auxiliary results concerning the intermediate statistic

In this section, we establish some properties of the intermediate test statistic $\widehat{\widehat{\Phi}}_{n,T}$ defined in (3.5) and its relationship with $\widehat{\Phi}_{n,T}$. First of all, we prove the proposition that relates the distribution of these two statistics.

Proposition A.5. *For any $x \in \mathbb{R}$ and $\gamma_{n,T} > 0$, we have*

$$\begin{aligned}\mathbb{P}\Big(\widehat{\widehat{\Phi}}_{n,T} \leq x - \gamma_{n,T}\Big) - \mathbb{P}\Big(\big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| > \gamma_{n,T}\Big) \leq \mathbb{P}(\widehat{\Phi}_{n,T} \leq x) \leq \\ \leq \mathbb{P}\Big(\widehat{\widehat{\Phi}}_{n,T} \leq x + \gamma_{n,T}\Big) + \mathbb{P}\Big(\big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| > \gamma_{n,T}\Big).\end{aligned}$$

Proof of Proposition A.5. From the law of total probability and the monotonic property of the probability function, we have

$$\begin{aligned}\mathbb{P}(\widehat{\Phi}_{n,T} \leq x) &= \mathbb{P}\Big(\widehat{\Phi}_{n,T} \leq x, \big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| \leq \gamma_{n,T}\Big) + \mathbb{P}\Big(\widehat{\Phi}_{n,T} \leq x, \big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| > \gamma_{n,T}\Big) \leq \\ &\leq \mathbb{P}\Big(\widehat{\Phi}_{n,T} \leq x, \widehat{\Phi}_{n,T} - \gamma_{n,T} \leq \widehat{\widehat{\Phi}}_{n,T} \leq \widehat{\Phi}_{n,T} + \gamma_{n,T}\Big) + \mathbb{P}\Big(\big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| > \gamma_{n,T}\Big) \leq \\ &\leq \mathbb{P}\Big(\widehat{\widehat{\Phi}}_{n,T} \leq x + \gamma_{n,T}\Big) + \mathbb{P}\Big(\big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| > \gamma_{n,T}\Big).\end{aligned}$$

Analogously,

$$\begin{aligned}\mathbb{P}(\widehat{\widehat{\Phi}}_{n,T} \leq x - \gamma_{n,T}) =&\mathbb{P}\Big(\widehat{\widehat{\Phi}}_{n,T} \leq x - \gamma_{n,T}, \big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| \leq \gamma_{n,T}\Big) + \\ &+ \mathbb{P}\Big(\widehat{\widehat{\Phi}}_{n,T} \leq x - \gamma_{n,T}, \big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| > \gamma_{n,T}\Big) \leq \\ \leq&\mathbb{P}\Big(\widehat{\widehat{\Phi}}_{n,T} \leq x - \gamma_{n,T}, \widehat{\widehat{\Phi}}_{n,T} - \gamma_{n,T} \leq \widehat{\Phi}_{n,T} \leq \widehat{\widehat{\Phi}}_{n,T} + \gamma_{n,T}\Big) + \\ &+ \mathbb{P}\Big(\big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| > \gamma_{n,T}\Big) \leq \\ \leq&\mathbb{P}\Big(\widehat{\Phi}_{n,T} \leq x\Big) + \mathbb{P}\Big(\big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| > \gamma_{n,T}\Big),\end{aligned}$$

which completes the proof. $\square$

The aim of the next proposition is to determine the value of $\gamma_{n,T}$ such that the difference between the distributions of $\widehat{\Phi}_{n,T}$ and $\widehat{\widehat{\Phi}}_{n,T}$ is not too big. In other words,

Proposition A.6. *There exists a sequnce of random numbers $\{\gamma_{n,T}\}_T$, that converges to 0 as $T \to \infty$, such that*

$$\mathbb{P}\Big(\big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| > \gamma_{n,T}\Big) = o(1).$$

Proof. Consider $\big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big|$. Analogously as in the proof of Proposition A.1, we can establish that

$$\big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| \leq \max_{1 \leq i < j \leq n} \Big(\{\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\}^{-1/2} \max_{(u,h) \in \mathcal{G}_T} \Big|\widehat{\widehat{\phi}}_{ij,T}(u,h) - \widehat{\phi}_{ij,T}(u,h)\Big|\Big),$$

where

$$\widehat{\widehat{\phi}}_{ij,T}(u,h) = \sum_{t=1}^{T} w_{t,T}(u,h)\big\{(\varepsilon_{it} - \bar{\varepsilon}_i) - (\varepsilon_{jt} - \bar{\varepsilon}_j)\big\}$$

and

$$\widehat{\phi}_{ij,T}(u,h) = \sum_{t=1}^{T} w_{t,T}(u,h)\big\{(\varepsilon_{it}-\bar{\varepsilon}_i)+(\boldsymbol{\beta}_i-\widehat{\boldsymbol{\beta}}_i)^\top(\mathbf{X}_{it}-\bar{\mathbf{X}}_i)-(\varepsilon_{jt}-\bar{\varepsilon}_j)-(\boldsymbol{\beta}_j-\widehat{\boldsymbol{\beta}}_j)^\top(\mathbf{X}_{jt}-\bar{\mathbf{X}}_j)\big\}.$$

Thus,

$$\begin{aligned}
\Big|\widehat{\widehat{\phi}}_{ij,T}(u,h) - \widehat{\phi}_{ij,T}(u,h)\Big| &= \Big|\sum_{t=1}^{T} w_{t,T}(u,h)\big\{(\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top(\mathbf{X}_{it} - \bar{\mathbf{X}}_i) - (\boldsymbol{\beta}_j - \widehat{\boldsymbol{\beta}}_j)^\top(\mathbf{X}_{jt} - \bar{\mathbf{X}}_j)\big\}\Big| \leq \\
&\leq \Big|\sum_{t=1}^{T} w_{t,T}(u,h)(\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top(\mathbf{X}_{it} - \bar{\mathbf{X}}_i)\Big| + \Big|\sum_{t=1}^{T} w_{t,T}(u,h)(\boldsymbol{\beta}_j - \widehat{\boldsymbol{\beta}}_j)^\top(\mathbf{X}_{jt} - \bar{\mathbf{X}}_j)\Big| = \\
&= |\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i|^\top\Big|\sum_{t=1}^{T} w_{t,T}(u,h)(\mathbf{X}_{it} - \bar{\mathbf{X}}_i)\Big| + |\boldsymbol{\beta}_j - \widehat{\boldsymbol{\beta}}_j|^\top\Big|\sum_{t=1}^{T} w_{t,T}(u,h)(\mathbf{X}_{jt} - \bar{\mathbf{X}}_j)\Big| = \\
&= |\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i|^\top\Big|\sum_{t=1}^{T} w_{t,T}(u,h)(\mathbf{X}_{it} - \bar{\mathbf{X}}_i)\Big| + |\boldsymbol{\beta}_j - \widehat{\boldsymbol{\beta}}_j|^\top\Big|\sum_{t=1}^{T} w_{t,T}(u,h)(\mathbf{X}_{jt} - \bar{\mathbf{X}}_j)\Big| \leq \\
&\leq |\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i|^\top\Big|\sum_{t=1}^{T} w_{t,T}(u,h)\mathbf{X}_{it}\Big| + \big|(\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top\bar{\mathbf{X}}_i\big|\Big|\sum_{t=1}^{T} w_{t,T}(u,h)\Big| + \\
&\quad + |\boldsymbol{\beta}_j - \widehat{\boldsymbol{\beta}}_j|^\top\Big|\sum_{t=1}^{T} w_{t,T}(u,h)\mathbf{X}_{jt}\Big| + \big|(\boldsymbol{\beta}_j - \widehat{\boldsymbol{\beta}}_j)^\top\bar{\mathbf{X}}_j\big|\Big|\sum_{t=1}^{T} w_{t,T}(u,h)\Big|
\end{aligned}$$

Hence,

$$\begin{aligned}
\big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| \leq& 2\max_{1\leq i<j\leq n}\Bigg(\{\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\}^{-1/2}|\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i|^\top \max_{(u,h)\in\mathcal{G}_T}\Big|\sum_{t=1}^{T} w_{t,T}(u,h)\mathbf{X}_{it}\Big|\Bigg) + \\
&+ 2\max_{1\leq i<j\leq n}\Bigg(\{\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\}^{-1/2}\big|(\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top\bar{\mathbf{X}}_i\big| \max_{(u,h)\in\mathcal{G}_T}\Big|\sum_{t=1}^{T} w_{t,T}(u,h)\Big|\Bigg)
\end{aligned} \tag{3.16}$$

To begin with, we consider the second summand in (3.16).
First, by our assumptions $\widehat{\sigma}_i^2 = \sigma_i^2 + o_P(\rho_T)$. Hence, $\max_{1\leq i<j\leq n}\{\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\}^{-1/2} = O_P(1)$.
Then, by Theorem 2.2, we know that $|\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i| = O_P(1/\sqrt{T})$.
Next, we want to show that $\bar{\mathbf{X}}_i = O_P(1)$.
For every element $\bar{X}_{i,j}, j = 1, \ldots, d$ of a vector $\bar{\mathbf{X}}_i$ we have

$$\mathbb{E}[\bar{X}_{i,j}] = \mathbb{E}\Big[\frac{1}{T}\sum_{t=1}^{T} H_{i,j}(\mathcal{U}_{it})\Big] = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[H_{i,j}(\mathcal{U}_{it})] = 0$$

and

$$\begin{aligned}
\mathrm{Var}[\bar{X}_{i,j}] &= \mathrm{Var}\Big[\frac{1}{T}\sum_{t=1}^{T} H_{i,j}(\mathcal{U}_{it})\Big] = \frac{1}{T^2}\mathbb{E}\Big[\Big(\sum_{t=1}^{T} H_{i,j}(\mathcal{U}_{it})\Big)^2\Big] = \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}[H_{i,j}^2(\mathcal{U}_{it})] + \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\mathbb{E}[H_{i,j}(\mathcal{U}_{it})H_{i,j}(\mathcal{U}_{is})] = \\
&= \frac{1}{T}\mathbb{E}[H_{i,j}^2(\mathcal{U}_{i0})] + \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\mathbb{E}[H_{i,j}(\mathcal{U}_{it})H_{i,j}(\mathcal{U}_{is})] \leq \\
&\leq \frac{1}{T}\mathbb{E}[H_{i,j}^2(\mathcal{U}_{i0})] + \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\sqrt{\mathbb{E}[H_{i,j}^2(\mathcal{U}_{it})]}\sqrt{\mathbb{E}[H_{i,j}^2(\mathcal{U}_{is})]} = \\
&= \frac{1}{T}\mathbb{E}[H_{i,j}^2(\mathcal{U}_{i0})] + \frac{T-1}{T}\mathbb{E}[H_{i,j}^2(\mathcal{U}_{i0})] = \\
&= \mathbb{E}\big[H_{i,j}^2(\mathcal{U}_{i0})\big],
\end{aligned}$$

by Chebyshev's inequality we have that $|\bar{X}_{i,j}| = \Big|\frac{1}{T}\sum_{t=1}^{T} H_{i,j}(\mathcal{U}_{it})\Big| = O_P(1)$ for each $j \in \{1,\dots,d\}$. And this in turn implies that

$$\big|\bar{\mathbf{X}}_i\big| = \Big|\frac{1}{T}\sum_{t=1}^{T}\mathbf{H}_i(\mathcal{U}_{it})\Big| = O_P(1).$$

Prove here that the value above is $o_P(1)$!!!

Finally, consider the local linear kernel weights $w_{t,T}(u,h)$ defined in (2.2). By construction the weights $w_{t,T}(u,h)$ are not equal to 0 if and only if $T(u-h) \leq t \leq T(u+h)$. We can use this fact to bound $\max_{(u,h)\in\mathcal{G}_T}\Big|\sum_{t=1}^{T} w_{t,T}(u,h)\Big|$ using the Cacuhy Schwarz inequality:

$$\begin{aligned}
\Big|\sum_{t=1}^{T} w_{t,T}(u,h)\Big| &= \Big|\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} w_{t,T}(u,h)\Big| \leq \sqrt{\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} w_{t,T}^2(u,h)}\sqrt{\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} 1^2} = \\
&= 1\cdot\sqrt{\lceil T(u+h)\rceil - \lfloor T(u-h)\rfloor + 1} \leq \sqrt{2Th+2} \leq \sqrt{2Th_{\max}+2}.
\end{aligned}$$

Hence, $\max_{(u,h)\in\mathcal{G}_T}\Big|\sum_{t=1}^{T} w_{t,T}(u,h)\Big| = O(\sqrt{T})$.

Now, back to the first summand in (3.16).

Similarly as before, by our assumptions $\widehat{\sigma}_i^2 = \sigma_i^2 + o_P(\rho_T)$. Hence,

$$\max_{1\leq i<j\leq n}\{\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\}^{-1/2} = O_P(1).$$

Then, by Theorem 2.2, we know that $|\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i| = O_P(1/\sqrt{T})$.

Consider the term $\big|\sum_{t=1}^{T} w_{t,T}(u,h)\mathbf{X}_{it}\big|$. Without loss of generality, we can regard the covariates $\mathbf{X}_{it}$ to be scalars X_{it}, not vectors. The proof in case of vectors proceeds analogously.

By construction the weights $w_{t,T}(u,h)$ are not equal to 0 if and only if $T(u-h) \le t \le T(u+h)$. We can use this fact to rewrite

$$\Big|\sum_{t=1}^{T} w_{t,T}(u,h)X_{it}\Big| = \Big|\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} w_{t,T}(u,h)X_{it}\Big|.$$

We want to show that

$$\max_{(u,h)\in\mathcal{G}_T}\Big|\sum_{t=1}^{T} w_{t,T}(u,h)X_{it}\Big| = \max_{(u,h)\in\mathcal{G}_T}\Big|\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} w_{t,T}(u,h)X_{it}\Big| = o_P(\sqrt{T}).$$

Note that

$$\begin{aligned}
\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} w^2_{t,T}(u,h) &= \sum_{t=1}^{T} w^2_{t,T}(u,h) = \\
&= \sum_{t=1}^{T} \frac{K^2\Big(\frac{\frac{t}{T}-u}{h}\Big)\Big[S_{T,2}(u,h) - \Big(\frac{\frac{t}{T}-u}{h}\Big)S_{T,1}(u,h)\Big]^2}{\Big\{\sum_{s=1}^{T} K^2\Big(\frac{\frac{s}{T}-u}{h}\Big)\Big[S_{T,2}(u,h) - \Big(\frac{\frac{s}{T}-u}{h}\Big)S_{T,1}(u,h)\Big]^2\Big\}} = \\
&= 1.
\end{aligned}$$

Denoting by $D_{T,u,h}$ the number of integers between $\lfloor T(u-h)\rfloor$ and $\lceil T(u+h)\rceil$ incl. (with obvious bounds $2Th \le D_{T,u,h} \le 2Th+2$), we can rewrite the weights as follows:

$$\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} \big(\sqrt{D_{T,u,h}}\cdot w_{t,T}(u,h)\big)^2 = D_{T,u,h}.$$

According to Theorem A.1 (Theorem 2 in Wu et al. (2016)), using the weights $a_t = \sqrt{D_{T,u,h}}\cdot w_{t,T}(u,h)$ we can bound the following probability:

$$\begin{aligned}
&\mathbb{P}\Bigg(\Big|\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} \sqrt{D_{T,u,h}}\cdot w_{t,T}(u,h)X_{it}\Big| \ge x\Bigg) \le \\
&\le C_1\frac{\big(\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil}|\sqrt{D_{T,u,h}}\cdot w_{t,T}(u,h)|^{q'}\big)||X_{i\cdot}||^{q'}_{q',\alpha}}{x^{q'}} + C_2\exp\Bigg(-\frac{C_3x^2}{D_{T,u,h}||X_{i\cdot}||^2_{2,\alpha}}\Bigg) = \\
&= C_1\frac{(\sqrt{D_{T,u,h}})^{q'}\big(\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil}|w_{t,T}(u,h)|^{q'}\big)||X_{i\cdot}||^{q'}_{q',\alpha}}{x^{q'}} + C_2\exp\Bigg(-\frac{C_3x^2}{D_{T,u,h}||X_{i\cdot}||^2_{2,\alpha}}\Bigg)
\end{aligned}$$

We want to prove that $\max_{(u,h)\in\mathcal{G}_T}\Big|\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} w_{t,T}(u,h)X_{it}\Big| = o_P(\sqrt{T})$. For that, take any $\delta > 0$:

$$\mathbb{P}\left(\frac{\max_{(u,h)\in\mathcal{G}_T}\left|\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} w_{t,T}(u,h)X_{it}\right|}{\sqrt{T}} \geq \delta\right) =$$

$$= \mathbb{P}\left(\max_{(u,h)\in\mathcal{G}_T}\left|\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} w_{t,T}(u,h)X_{it}\right| \geq \delta\sqrt{T}\right) \leq$$

$$\overset{\text{Boole's inequality}}{\leq} \sum_{(u,h)\in\mathcal{G}_T}\mathbb{P}\left(\left|\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} w_{t,T}(u,h)X_{it}\right| \geq \delta\sqrt{T}\right) =$$

$$\overset{\text{"normalisation"}}{=} \sum_{(u,h)\in\mathcal{G}_T}\mathbb{P}\left(\left|\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} \sqrt{D_{T,u,h}}\cdot w_{t,T}(u,h)X_{it}\right| \geq \delta\sqrt{D_{T,u,h}T}\right) \leq$$

$$\overset{\text{Wu's Theorem}}{\leq} \sum_{(u,h)\in\mathcal{G}_T}\left[C_1\frac{(\sqrt{D_{T,u,h}})^{q'}\left(\sum|w_{t,T}(u,h)|^{q'}\right)||X_{i\cdot}||_{q',\alpha}^{q'}}{\left(\delta\sqrt{D_{T,u,h}T}\right)^{q'}} + C_2\exp\left(-\frac{C_3\left(\delta\sqrt{D_{T,u,h}T}\right)^2}{D_{T,u,h}||X_{i\cdot}||_{2,\alpha}^2}\right)\right] =$$

$$\overset{\text{simplification}}{=} \sum_{(u,h)\in\mathcal{G}_T}\left[C_1\frac{\left(\sum|w_{t,T}(u,h)|^{q'}\right)||X_{i\cdot}||_{q',\alpha}^{q'}}{\left(\delta\sqrt{T}\right)^{q'}} + C_2\exp\left(-\frac{C_3\delta^2T}{||X_{i\cdot}||_{2,\alpha}^2}\right)\right] \leq$$

$$\leq C_1\frac{T^\theta||X_{i\cdot}||_{q',\alpha}^{q'}}{T^{q'/2}\cdot\delta^{q'}}\max_{(u,h)\in\mathcal{G}_T}\left(\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil}|w_{t,T}(u,h)|^{q'}\right) + C_2T^\theta\exp\left(-\frac{C_3\delta^2T}{D_{T,u,h}||X_{i\cdot}||_{2,\alpha}^2}\right) =$$

$$\leq C\frac{T^{\theta-q'/2}}{\delta^{q'}} + CT^\theta\exp\left(-CT\delta^2\right).$$

where in the last equality we used the following facts:

1. $||X_{i\cdot}||_{q',\alpha}^{q'} < \infty$ by the assumption (C8);
2. $\max_{(u,h)\in\mathcal{G}_T}\left(\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil}|w_{t,T}(u,h)|^{q'}\right) < \infty$; (still need to fomally write down)
3. $||X_{i\cdot}||_{2,\alpha}^2 < \infty$ (follows from (1)).

If we take $\theta - q'/2 < 0$, then the term on the RHS is converging to can 0 as $T \to \infty$ for any fixed $\delta > 0$. Hence,

$$\max_{(u,h)\in\mathcal{G}_T}\left|\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} w_{t,T}(u,h)X_{it}\right| = o_P(\sqrt{T}).$$

Plugging all the equations derived above together, we can establish that the first summand is also $o_P(1)$. Hence, $\left|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\right| = o_P(1)$ and the statement of the theorem follows. □

Proposition A.7. *Under the conditions of Theorem 2.1, it holds that*

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}(\widehat{\Phi}_{n,T} \leq x) - \mathbb{P}(\Phi_{n,T} \leq x)\right| = o(1).$$

Proof. First, we consider those $x \in \mathbb{R}$ such that $\mathbb{P}(\widehat{\Phi}_{n,T} \le x) \ge \mathbb{P}(\Phi_{n,T} \le x)$. Then by Proposition A.5 and taking $\gamma_{n,T}$ from the Proposition A.6 we have

$$
\begin{aligned}
&0 \le \left|\mathbb{P}(\widehat{\Phi}_{n,T} \le x) - \mathbb{P}(\Phi_{n,T} \le x)\right| = \mathbb{P}(\widehat{\Phi}_{n,T} \le x) - \mathbb{P}(\Phi_{n,T} \le x) \le \\
&\quad \le \mathbb{P}\Big(\widehat{\widehat{\Phi}}_{n,T} \le x + \gamma_{n,T}\Big) + \mathbb{P}\Big(\big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| > \gamma_{n,T}\Big) - \mathbb{P}\Big(\Phi_{n,T} \le x\Big) = \\
&\quad = \mathbb{P}\Big(\widehat{\widehat{\Phi}}_{n,T} \le x + \gamma_{n,T}\Big) - \mathbb{P}\Big(\Phi_{n,T} \le x\Big) + o(1) = \\
&\quad = \mathbb{P}\Big(\widehat{\widehat{\Phi}}_{n,T} \le x + \gamma_{n,T}\Big) - \mathbb{P}\Big(\Phi_{n,T} \le x + \gamma_{n,T}\Big) + \mathbb{P}\Big(\Phi_{n,T} \le x + \gamma_{n,T}\Big) - \mathbb{P}\Big(\Phi_{n,T} \le x\Big) + o(1) = \\
&\quad = \mathbb{P}\Big(\widehat{\widehat{\Phi}}_{n,T} \le x + \gamma_{n,T}\Big) - \mathbb{P}\Big(\Phi_{n,T} \le x + \gamma_{n,T}\Big) + \mathbb{P}\Big(x \le \Phi_{n,T} \le x + \gamma_{n,T}\Big) + o(1) \le \\
&\quad \le \mathbb{P}\Big(\widehat{\widehat{\Phi}}_{n,T} \le x + \gamma_{n,T}\Big) - \mathbb{P}\Big(\Phi_{n,T} \le x + \gamma_{n,T}\Big) + \mathbb{P}\Big(\big|\Phi_{n,T} - x\big| \le \gamma_{n,T}\Big) + o(1).
\end{aligned}
$$

Now consider such $x \in \mathbb{R}$ that $\mathbb{P}(\widehat{\Phi}_{n,T} \le x) < \mathbb{P}(\Phi_{n,T} \le x)$. Then again by Proposition A.5 for $\gamma_{n,T}$ from Proposition A.6 we have

$$
\begin{aligned}
&0 \le \left|\mathbb{P}(\widehat{\Phi}_{n,T} \le x) - \mathbb{P}(\Phi_{n,T} \le x)\right| = \mathbb{P}(\Phi_{n,T} \le x) - \mathbb{P}(\widehat{\Phi}_{n,T} \le x) \le \\
&\quad \le \mathbb{P}\Big(\Phi_{n,T} \le x\Big) + \mathbb{P}\Big(\big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| > \gamma_{n,T}\Big) - \mathbb{P}\Big(\widehat{\widehat{\Phi}}_{n,T} \le x - \gamma_{n,T}\Big) = \\
&\quad = \mathbb{P}\Big(\Phi_{n,T} \le x\Big) - \mathbb{P}\Big(\widehat{\widehat{\Phi}}_{n,T} \le x - \gamma_{n,T}\Big) + o(1) = \\
&\quad = \mathbb{P}\Big(\Phi_{n,T} \le x\Big) - \mathbb{P}\Big(\Phi_{n,T} \le x - \gamma_{n,T}\Big) + \mathbb{P}\Big(\Phi_{n,T} \le x - \gamma_{n,T}\Big) - \mathbb{P}\Big(\widehat{\widehat{\Phi}}_{n,T} \le x - \gamma_{n,T}\Big) + o(1) = \\
&\quad = \mathbb{P}\Big(x - \gamma_{n,T} \le \Phi_{n,T} \le x\Big) + \mathbb{P}\Big(\Phi_{n,T} \le x - \gamma_{n,T}\Big) - \mathbb{P}\Big(\widehat{\widehat{\Phi}}_{n,T} \le x - \gamma_{n,T}\Big) + o(1) \le \\
&\quad \le \mathbb{P}\Big(\big|\Phi_{n,T} - x\big| \le \gamma_{n,T}\Big) + \mathbb{P}\Big(\Phi_{n,T} \le x - \gamma_{n,T}\Big) - \mathbb{P}\Big(\widehat{\widehat{\Phi}}_{n,T} \le x - \gamma_{n,T}\Big) + o(1).
\end{aligned}
$$

Now note that since $\gamma_{n,T} \to 0$, we can use the anticoncentration results for the Gaussian statistic $\Phi_{n,T}$ to get $\mathbb{P}\Big(\big|\Phi_{n,T} - x\big| \le \gamma_{n,T}\Big) = o(1)$. Thus,

$$
\begin{aligned}
&\sup_{x\in\mathbb{R}} \left|\mathbb{P}(\widehat{\Phi}_{n,T} \le x) - \mathbb{P}(\Phi_{n,T} \le x)\right| \le \\
&\qquad \le \max\Bigg\{ \sup_{x\in\mathbb{R}} \left|\mathbb{P}\Big(\Phi_{n,T} \le x - \gamma_{n,T}\Big) - \mathbb{P}\Big(\widehat{\widehat{\Phi}}_{n,T} \le x - \gamma_{n,T}\Big)\right|, \\
&\qquad\qquad \sup_{x\in\mathbb{R}} \left|\mathbb{P}\Big(\Phi_{n,T} \le x + \gamma_{n,T}\Big) - \mathbb{P}\Big(\widehat{\widehat{\Phi}}_{n,T} \le x + \gamma_{n,T}\Big)\right| \Bigg\} + o(1) = \\
&\qquad = \sup_{y\in\mathbb{R}} \left|\mathbb{P}\Big(\Phi_{n,T} \le y\Big) - \mathbb{P}\Big(\widehat{\widehat{\Phi}}_{n,T} \le y\Big)\right| + o(1) = o(1).
\end{aligned}
$$

□

3.2 Proof of Theorem 2.2

We define the first-differenced regressors as follows.

$$
\Delta \mathbf{X}_{it} = \mathbf{H}_i(\mathcal{U}_{it}) - \mathbf{H}_i(\mathcal{U}_{it-1}) := \Delta \mathbf{H}_i(\mathcal{U}_{it}).
$$

Similarly,

$$\Delta\varepsilon_{it} = \varepsilon_{it} - \varepsilon_{it-1} = G_i(\mathcal{J}_{it}) - G_i(\mathcal{J}_{it-1}) = \Delta G_i(\mathcal{J}_{it}).$$

With these assumptions we can prove the following propositions.

Proposition 3.1. *Under Assumptions (C4) and (C6), $||\Delta\mathbf{H}_i(\mathcal{U}_{it})||_4 < \infty$.*

Proof of Proposition 3.1. By Assumption (C6),

$$||\Delta\mathbf{H}_i(\mathcal{U}_{it})||_4 \leq ||\mathbf{H}_i(\mathcal{U}_{it})||_4 + ||\mathbf{H}_i(\mathcal{U}_{it-1})||_4 < \infty.$$

□

Proposition 3.2. *Under Assumption (C7), $\Delta\mathbf{X}_{it}$ (elementwise) and $\Delta\varepsilon_{it}$ are uncorrelated for each $t \in \{1, \dots, T\}$.*

Proof of Proposition 3.2. By Assumption (C7),

$$\begin{aligned}\mathbb{E}[\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}] = \mathbb{E}\big[(\mathbf{X}_{it} - \mathbf{X}_{it-1})(\varepsilon_{it} - \varepsilon_{it-1})\big] = \\ =&\mathbb{E}[\mathbf{X}_{it}\varepsilon_{it}] - \mathbb{E}[\mathbf{X}_{it-1}\varepsilon_{it}] - \mathbb{E}[\mathbf{X}_{it}\varepsilon_{it-1}] + \mathbb{E}[\mathbf{X}_{it-1}\varepsilon_{it-1}] = \\ =&\mathbb{E}[\mathbf{X}_{it}]\mathbb{E}[\varepsilon_{it}] - \mathbb{E}[\mathbf{X}_{it-1}]\mathbb{E}[\varepsilon_{it}] - \mathbb{E}[\mathbf{X}_{it}]\mathbb{E}[\varepsilon_{it-1}] + \mathbb{E}[\mathbf{X}_{it-1}]\mathbb{E}[\varepsilon_{it-1}] = \\ =&\big(\mathbb{E}[\mathbf{X}_{it}] - \mathbb{E}[\mathbf{X}_{it-1}]\big)\big(\mathbb{E}[\varepsilon_{it}] - \mathbb{E}[\varepsilon_{it-1}]\big) = \mathbb{E}[\Delta\mathbf{X}_{it}]\mathbb{E}[\Delta\varepsilon_{it}]\end{aligned}$$

□

Proposition 3.3. *Define*

$$\Delta\mathbf{U}_i(\mathcal{I}_{i,t}) := \Delta\mathbf{H}_i(\mathcal{U}_{it})\Delta G_i(\mathcal{J}_{it}).$$

Under Assumptions (C2) - (C3), (C8) - **??**, *we have that $\sum_{s=0}^{\infty}\delta_2(\Delta\mathbf{U}_i, s) < \infty$.*

Proof of Proposition 3.3. Note the following

$$\begin{aligned}&\delta_2(\Delta\mathbf{U}_i, t) = ||\Delta\mathbf{U}_i(\mathcal{I}_{i,t}) - \Delta\mathbf{U}_i(\mathcal{I}'_{i,t})||_2 = \\ &= ||\Delta\mathbf{H}_i(\mathcal{U}_{it})\Delta G_i(\mathcal{J}_{it}) - \Delta\mathbf{H}_i(\mathcal{U}'_{it})\Delta G_i(\mathcal{J}'_{it})||_2 = \\ &= ||\big(\mathbf{H}_i(\mathcal{U}_{it}) - \mathbf{H}_i(\mathcal{U}_{it-1})\big)\big(G_i(\mathcal{J}_{it}) - G_i(\mathcal{J}_{it-1})\big) - \big(\mathbf{H}_i(\mathcal{U}'_{it}) - \mathbf{H}_i(\mathcal{U}'_{it-1})\big)\big(G_i(\mathcal{J}'_{it}) - G_i(\mathcal{J}'_{it-1})\big)||_2 = \\ &= ||\mathbf{H}_i(\mathcal{U}_{it})G_i(\mathcal{J}_{it}) - \mathbf{H}_i(\mathcal{U}_{it-1})G_i(\mathcal{J}_{it}) - \mathbf{H}_i(\mathcal{U}_{it})G_i(\mathcal{J}_{it-1}) + \mathbf{H}_i(\mathcal{U}_{it-1})G_i(\mathcal{J}_{it-1}) - \\ &\quad - \mathbf{H}_i(\mathcal{U}'_{it})G_i(\mathcal{J}'_{it}) + \mathbf{H}_i(\mathcal{U}'_{it-1})G_i(\mathcal{J}'_{it}) + \mathbf{H}_i(\mathcal{U}'_{it})G_i(\mathcal{J}'_{it-1}) - \mathbf{H}_i(\mathcal{U}'_{it-1})G_i(\mathcal{J}'_{it-1})||_2 \leq \\ &\leq ||\mathbf{H}_i(\mathcal{U}_{it})G_i(\mathcal{J}_{it}) - \mathbf{H}_i(\mathcal{U}'_{it})G_i(\mathcal{J}'_{it})||_2 + ||\mathbf{H}_i(\mathcal{U}_{it-1})G_i(\mathcal{J}_{it-1}) - \mathbf{H}_i(\mathcal{U}'_{it-1})G_i(\mathcal{J}'_{it-1})||_2 + \\ &\quad + ||\mathbf{H}_i(\mathcal{U}_{it-1})G_i(\mathcal{J}_{it}) - \mathbf{H}_i(\mathcal{U}'_{it-1})G_i(\mathcal{J}'_{it})||_2 + ||\mathbf{H}_i(\mathcal{U}_{it})G_i(\mathcal{J}_{it-1}) - \mathbf{H}_i(\mathcal{U}'_{it})G_i(\mathcal{J}'_{it-1})||_2 = \\ &= \delta_2(\mathbf{U}_i, t) + \delta_2(\mathbf{U}_i, t-1) + \\ &\quad + ||\mathbf{H}_i(\mathcal{U}_{it-1})G_i(\mathcal{J}_{it}) - \mathbf{H}_i(\mathcal{U}'_{it-1})G_i(\mathcal{J}_{it}) + \mathbf{H}_i(\mathcal{U}'_{it-1})G_i(\mathcal{J}_{it}) - \mathbf{H}_i(\mathcal{U}'_{it-1})G_i(\mathcal{J}'_{it})||_2 + \\ &\quad + ||\mathbf{H}_i(\mathcal{U}_{it})G_i(\mathcal{J}_{it-1}) - \mathbf{H}_i(\mathcal{U}'_{it})G_i(\mathcal{J}_{it-1}) + \mathbf{H}_i(\mathcal{U}'_{it})G_i(\mathcal{J}_{it-1}) - \mathbf{H}_i(\mathcal{U}'_{it})G_i(\mathcal{J}'_{it-1})||_2 \leq \\ &\leq \delta_2(\mathbf{U}_i, t) + \delta_2(\mathbf{U}_i, t-1) + \\ &\quad + ||\big(\mathbf{H}_i(\mathcal{U}_{it-1}) - \mathbf{H}_i(\mathcal{U}'_{it-1})\big)G_i(\mathcal{J}_{it})||_2 + ||\mathbf{H}_i(\mathcal{U}'_{it-1})\big(G_i(\mathcal{J}_{it}) - G_i(\mathcal{J}'_{it})\big)||_2 + \\ &\quad + ||\big(\mathbf{H}_i(\mathcal{U}_{it}) - \mathbf{H}_i(\mathcal{U}'_{it})\big)G_i(\mathcal{J}_{it-1})||_2 + ||\mathbf{H}_i(\mathcal{U}'_{it})\big(G_i(\mathcal{J}_{it-1}) - G_i(\mathcal{J}'_{it-1})\big)||_2 \leq \\ &\leq \delta_2(\mathbf{U}_i, t) + \delta_2(\mathbf{U}_i, t-1) + \big(\delta_2(\mathbf{H}_i, t-1) + \delta_2(\mathbf{H}_i, t)\big)||G_i||_2 + \big(\delta_2(G_i, t-1) + \delta_2(G_i, t)\big)||\mathbf{H}_i||_2\end{aligned}$$

Here $\mathcal{U}'_{it} = (\ldots, u_{i(-1)}, u'_{i0}, u_{i1}, \ldots, u_{it-1}, u_{it})$, $\mathcal{U}'_{it-1} = (\ldots, u_{i(-1)}, u'_{i0}, u_{i1}, \ldots, u_{it-1})$, $\mathcal{J}'_{it} = (\ldots, \eta_{i(-1)}, \eta'_{i0}, \eta_{i1}, \ldots, \eta_{it-1}, \eta_{it})$, $\mathcal{J}'_{it-1} = (\ldots, \eta_{i(-1)}, \eta'_{i0}, \eta_{i1}, \ldots, \eta_{it-1})$ are coupled processes with u'_{i0} being an i.i.d. copy of u_{i0} and η'_{i0} being an i.i.d. copy of η_{i0}. This leads us to

$$\sum_{s=0}^{\infty} \delta_2(\Delta \mathbf{U}_i, s) \leq \sum_{s=0}^{\infty} \delta_2(\mathbf{U}_i, s) + \sum_{s=1}^{\infty} \delta_2(\mathbf{U}_i, s-1) +$$
$$+ \sum_{s=1}^{\infty} \big(\delta_2(\mathbf{H}_i, s-1) + \delta_2(\mathbf{H}_i, s)\big) ||G_i||_2 + \sum_{s=1}^{\infty} \big(\delta_2(G_i, s-1) + \delta_2(G_i, s)\big) ||\mathbf{H}_i||_2 < \infty$$

□

Proposition 3.4. *Under Assumptions (C1) -* **??**,

$$\Big|\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Delta \mathbf{X}_{it} \Delta \varepsilon_{it}\Big| = O_P(1).$$

Proof of Proposition 3.4. We need the following notation:

$$\begin{aligned}
\mathcal{P}_{i,t}(\cdot) &:= \mathbb{E}[\cdot|\mathcal{I}_{i,t}] - \mathbb{E}[\cdot|\mathcal{I}_{i,t-1}], \\
\kappa_i &:= \frac{1}{T} \sum_{t=1}^{T} \Delta \mathbf{X}_{it} \Delta \varepsilon_{it}, \\
\kappa_{i,s}^{\mathcal{P}} &:= \frac{1}{T} \sum_{t=1}^{T} \mathcal{P}_{i,t-s}\big(\Delta \mathbf{X}_{it} \Delta \varepsilon_{it}\big).
\end{aligned}$$

Then,

$$\begin{aligned}
||\kappa_{i,s}^{\mathcal{P}}||^2 &= \Big|\Big|\frac{1}{T} \sum_{t=1}^{T} \mathcal{P}_{i,t-s}\big(\Delta \mathbf{X}_{it} \Delta \varepsilon_{it}\big)\Big|\Big|^2 \leq \\
&\leq \frac{1}{T^2} \sum_{t=1}^{T} \Big|\Big|\mathbb{E}(\Delta \mathbf{X}_{it} \Delta \varepsilon_{it} | \mathcal{I}_{i,t-s}) - \mathbb{E}(\Delta \mathbf{X}_{it} \Delta \varepsilon_{it} | \mathcal{I}_{i,t-s-1})\Big|\Big|^2 = \\
&= \frac{1}{T^2} \sum_{t=1}^{T} \Big|\Big|\mathbb{E}(\Delta \mathbf{X}_{it} \Delta \varepsilon_{it} | \mathcal{I}_{i,t-s}) - \mathbb{E}(\Delta \mathbf{X}'_{it,s} \Delta \varepsilon'_{it,s} | \mathcal{I}_{i,t-s})\Big|\Big|^2,
\end{aligned}$$

where $\Delta \mathbf{X}'_{it,s} \Delta \varepsilon'_{it,s}$ denotes $\Delta \mathbf{X}_{it} \Delta \varepsilon_{it}$ with $\{\zeta_{i,t-s}\}$ replaced by its i.i.d. copy $\{\zeta'_{i,t-s}\}$. In this case $\mathbb{E}(\Delta \mathbf{X}'_{it,s} \Delta \varepsilon'_{it,s} | \mathcal{I}_{i,t-s-1}) = \mathbb{E}(\Delta \mathbf{X}'_{it,s} \Delta \varepsilon'_{it,s} | \mathcal{I}_{i,t-s})$. Furthermore, by linearity of the expectation and Jensen's inequality, we have

$$||\kappa_{i,s}^{\mathcal{P}}||^2 \leq \frac{1}{T^2}\sum_{t=1}^{T}\Big|\Big|\mathbb{E}(\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}|\mathcal{I}_{i,t-s}) - \mathbb{E}(\Delta\mathbf{X}'_{it,s}\Delta\varepsilon'_{it,s}|\mathcal{I}_{i,t-s})\Big|\Big|^2 \leq$$

$$\leq \frac{1}{T^2}\sum_{t=1}^{T}\Big|\Big|\Delta\mathbf{X}_{it}\Delta\varepsilon_{it} - \Delta\mathbf{X}'_{it,s}\Delta\varepsilon'_{it,s}\Big|\Big|^2 =$$

$$= \frac{1}{T^2}\sum_{t=1}^{T}\Big|\Big|\Delta\mathbf{H}_i(\mathcal{U}_{it})\Delta G_i(\mathcal{J}_{it}) - \Delta\mathbf{H}_i(\mathcal{U}'_{it,s})\Delta G_i(\mathcal{J}'_{it,s})\Big|\Big|^2 =$$

$$= \frac{1}{T^2}\sum_{t=1}^{T}\Big|\Big|\Delta\mathbf{U}_i(\mathcal{I}_{i,t}) - \Delta\mathbf{U}_i(\mathcal{I}'_{i,t,s})\Big|\Big|^2 \leq \frac{1}{T^2}\sum_{t=1}^{T}\delta_2^2(\Delta\mathbf{U}_i, s) = \frac{1}{T}\delta_2^2(\Delta\mathbf{U}_i, s)$$

with $\mathcal{U}'_{it,s} = (\ldots, u_{it-s-1}, u'_{it-s}, u_{it-s+1}, \ldots, u_{it})$, $\mathcal{J}'_{it,s} = (\ldots, \eta_{it-s-1}, \eta'_{it-s}, \eta_{it-s+1}, \ldots, \eta_{it})$, $\zeta'_{it} = (u'_{it}, \eta'_{it})^\top$ and $\mathcal{I}'_{i,t,s} = (\ldots, \zeta_{it-s-1}, \zeta'_{it-s}, \zeta_{it-s+1}, \ldots, \zeta_{it})$.
Moreover,

$$\kappa_i - \mathbb{E}\kappa_i = \frac{1}{T}\sum_{t=1}^{T}\Delta\mathbf{X}_{it}\Delta\varepsilon_{it} - \mathbb{E}\kappa_i = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}(\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}|\mathcal{I}_{i,t}) - \mathbb{E}\kappa_i =$$

$$= \frac{1}{T}\sum_{t=1}^{T}\big(\mathbb{E}(\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}|\mathcal{I}_{i,t}) - \mathbb{E}(\mathbf{X}_{it}\Delta\varepsilon_{it})\big) =$$

$$= \frac{1}{T}\sum_{t=1}^{T}\sum_{s=0}^{\infty}\big(\mathbb{E}(\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}|\mathcal{I}_{i,t-s}) - \mathbb{E}(\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}|\mathcal{I}_{i,t-s-1})\big) =$$

$$= \frac{1}{T}\sum_{t=1}^{T}\sum_{s=0}^{\infty}\mathcal{P}_{i,t-s}(\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}) = \sum_{s=0}^{\infty}\kappa_{i,s}^{\mathcal{P}}.$$

Thus, by Proposition 3.3,

$$||\kappa_i - \mathbb{E}\kappa_i|| \leq \sum_{s=0}^{\infty}||\kappa_{i,s}^{\mathcal{P}}|| \leq \frac{1}{\sqrt{T}}\sum_{s=0}^{\infty}\delta_2(\Delta\mathbf{U}_i, s) = O\Big(\frac{1}{\sqrt{T}}\Big)$$

Since $\mathbb{E}\kappa_i = 0$ by Proposition 3.2, we conclude that

$$\Big|\Big|\frac{1}{T}\sum_{t=1}^{T}\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}\Big|\Big| = O\Big(\frac{1}{\sqrt{T}}\Big).$$

Therefore, the proposition follows. □

Proof of Theorem 2.2. Recall the differencing estimator $\widehat{\boldsymbol{\beta}}_i$:

$$\widehat{\boldsymbol{\beta}}_i = \Big(\sum_{t=1}^{T}\Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^\top\Big)^{-1}\sum_{t=1}^{T}\Delta\mathbf{X}_{it}\Delta Y_{it} =$$

$$= \Big(\sum_{t=1}^{T}\Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^\top\Big)^{-1}\sum_{t=1}^{T}\Delta\mathbf{X}_{it}\Big(\Delta\mathbf{X}_{it}^\top\boldsymbol{\beta}_i + \Delta\varepsilon_{it} + O\Big(\frac{1}{T}\Big)\Big) =$$

$$= \boldsymbol{\beta}_i + \Big(\sum_{t=1}^{T}\Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^\top\Big)^{-1}\sum_{t=1}^{T}\Delta\mathbf{X}_{it}\Delta\varepsilon_{it} + O\Big(\frac{1}{T}\Big)\Big(\sum_{t=1}^{T}\Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^\top\Big)^{-1}\sum_{t=1}^{T}\Delta\mathbf{X}_{it}.$$

This leads to

$$\sqrt{T}(\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i) = \Big(\frac{1}{T}\sum_{t=1}^{T}\Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^{\top}\Big)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}+$$
$$+\,O\Big(\frac{1}{\sqrt{T}}\Big)\Big(\frac{1}{T}\sum_{t=1}^{T}\Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^{\top}\Big)^{-1}\frac{1}{T}\sum_{t=1}^{T}\Delta\mathbf{X}_{it}.$$

Since

$$\mathbb{E}\Big[\frac{1}{T}\sum_{t=1}^{T}\Delta H_{ij}(\mathcal{U}_{it})\Big] = 0$$

and

$$\mathrm{Var}\Big[\frac{1}{T}\sum_{t=1}^{T}\Delta H_{ij}(\mathcal{U}_{it})\Big] \le \frac{4}{T^2}\mathbb{E}\big[H_{ij}^2(\mathcal{U}_{it})\big],$$

by Chebyshev's inequality we have that $\Big|\frac{1}{T}\sum_{t=1}^{T}\Delta H_{ij}(\mathcal{U}_{it})\Big| = O_P(1)$ for each $j \in \{1,\dots,d\}$. And this in turn implies that

$$\Big|\frac{1}{T}\sum_{t=1}^{T}\Delta\mathbf{H}_i(\mathcal{U}_{it})\Big| = \Big|\frac{1}{T}\sum_{t=1}^{T}\Delta\mathbf{X}_{it}\Big| = O_P(1). \tag{3.17}$$

Similarly, by Proposition 3.1 and Chebyshev's inequality, we have that for each $j,k \in \{1,\dots,d\}$

$$\Big|\frac{1}{T}\sum_{t=1}^{T}\Delta H_{ij}(\mathcal{U}_{it})\Delta H_{ik}(\mathcal{U}_{it})\Big| = O_P(1),$$

which leads to

$$\Big|\Big|\frac{1}{T}\sum_{t=1}^{T}\Delta\mathbf{H}_i(\mathcal{U}_{it})\Delta\mathbf{H}_i(\mathcal{U}_{it})^{\top}\Big|\Big| = \Big|\Big|\frac{1}{T}\sum_{t=1}^{T}\Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^{\top}\Big|\Big| = O_P(1), \tag{3.18}$$

where $||A||$ with A being a matrix is any matrix norm.
By Assumption (C5), we know that $\mathbb{E}[\Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^{\top}] = \mathbb{E}[\Delta\mathbf{X}_{i0}\Delta\mathbf{X}_{i0}^{\top}]$ is invertible, thus,

$$\Bigg|\Bigg|\Big(\frac{1}{T}\sum_{t=1}^{T}\Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^{\top}\Big)^{-1}\Bigg|\Bigg| = O_P(1).$$

By applying Proposition 3.4, (3.17) and (3.18), the statement of the theorem follows. □