Multiscale Testing for Equality of Nonparametric Trend Curves

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

We develop multiscale methods to test qualitative hypotheses about nonparametric time trends in the presence of covariates. In many applications, practitioners are interested whether the observed time series all have the same time trend. Moreover, when some of the trends are different, there may still be groups of time series with the same trend. In this case, it is often of interest to estimate the unknown groups from the data. In addition, when two trends are not the same, it may also be relevant to know in which time regions they differ from each other. 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.

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 Introduction

Comparison of several regression curves is a classical topic in econometrics and statistics. In many cases of practical interest, the objective regression curves are of unknown functional form and the parametric approach is not applicable. In this paper, we are interested in performing the comparison of several regression curves in a nonparametric context. Specifically, we present a new testing procedure for detecting differences in the nonparametric trends curves.

In what follows, we consider a general panel framework with heterogeneous trends. Suppose we observe a panel of n time series $\mathcal{Z}_i = \{(Y_{it}, \mathbf{X}_{it}) : 1 \leq t \leq T\}$ for $1 \leq i \leq n$, where Y_{it} are real-valued random variables and $\mathbf{X}_{it} = (X_{it,1}, \dots, X_{it,d})^\top$ are d-dimensional random vectors. Each time series $\mathcal{Z}_i$ is modelled by the equation

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

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 or controls, m_i is an unknown nonparametric (deterministic) trend function defined on $[0, 1]$, α_i are so-called fixed effect error terms and $\mathcal{E}_i = \{\varepsilon_{it} : 1 \leq t \leq T\}$ is a zero-mean stationary error process.

[1]Corresponding author. Address: Bonn Graduate School of Economics, University of Bonn, 53113 Bonn, Germany. Email: `marina.k@uni-bonn.de`.
[2]Address: Institute of Statistics, Department of Mathematics and Economics, Ulm University, 89081 Ulm, Germany. Email: `m.vogt@uni-ulm.de`.

An important question in many applications is whether the observed time series have the common trend. In other words, the researchers would like to know if m_i are the same for all i. Moreover, when some of the trends are different, there may still be groups of time series with the same trend. In this case, it is often of interest to estimate the unknown groups from the data. In addition, when two trends m_i and m_j are not the same, it may also be relevant to know in which time regions they differ from each other. In this paper, we introduce new statistical methods to approach these questions. In particular, we develop a test of the hypothesis that all time trends in model (1.1) are the same. In this setting, the null hypothesis is formulated as

$$H_0 : m_1 = m_2 = \ldots = m_n, \tag{1.2}$$

whereas the alternative hypothesis is

$$H_1 : \text{ there exists } x \in [0,1] \text{ such that } m_i(x) \neq m_j(x) \text{ for some } 1 \leq i < j \leq n.$$

The method that we propose does not only allow to test whether the null hypothesis is violated. It also allows to detect, with a given statistical confidence, which time trends are different and in which time regions they differ. More specifically, for any given interval $[u-h, u+h] \subseteq [0,1]$, consider the hypothesis

$$H_0^{[i,j]}(u,h) : m_i(w) = m_j(w) \text{ for all } w \in [u-h, u+h].$$

Here, we can regard h as a bandwidth, a common tuning parameter in nonparametric estimation. The given interval $\mathcal{I}_{(u,h)} = [u-h, u+h] \subseteq [0,1]$ is then fully characterized by u, its center (a location parameter), and h, the bandwidth. In order to determine the regions where the time trends are different, we consider a broad range of pairs (u,h) with the property that they fully cover the unit interval $[0,1]$. Formally, let $\mathcal{G} := \{(u,h) : \mathcal{I}_{(u,h)} = [u-h, u+h] \subseteq [0,1]\}$ be a grid of location-bandwidth points such that

$$\bigcup_{(u,h)\in\mathcal{G}} \mathcal{I}_{(u,h)} = [0,1].$$

We then reformulate our null hypothesis (1.2) as

$$\begin{aligned} H_0 : {} & \text{The hypothesis } H_0^{[i,j]}(u,h) \text{ holds true for all intervals } \mathcal{I}_{(u,h)}, (u,h) \in \mathcal{G}, \\ & \text{and for all } 1 \leq i < j \leq n. \end{aligned}$$

In this paper, we introduce a method that allows to test the hypotheses $H_0^{[i,j]}(u,h)$ simultaneously for all pairs (i,j) and for all intervals $\mathcal{I}_{(u,h)}$ under consideration. Specifically, we develop a multiscale test for the model (1.1). The underlying idea of any multiscale test is to consider a number of test statistics (each of which corresponds to different values of some tuning parameters) all at once rather than to perform a separate test for each single test statistics. In our case, this means testing many different null hypotheses $H_0^{[i,j]}(u,h)$ simultaneously. In the paper, we show how to derive appropriate critical values and prove that the proposed multiscale test has the correct (asymptotic) level, which constitutes the main theoretical result of the paper.

Trend comparison is a common statistical problem that arises in various contexts. For example, in economics the researchers are interested in comparing trends in real gross domestic product across several countries (Grier and Tullock, 1989), in yield over time of US Treasury bills at different maturities (Park et al., 2009), or the evolution of long-term interest rates in a number of countries (Christiansen and Pigott, 1997). In finance, comparison and subsequent classification of the trends of market fragmentation can be used to assess the market quality in the European stock market (Vogt and Linton, 2017, 2020). In climatology, comparing the temperature time series in different areas is investigated in the context of the regional and global warming trends (Karoly and Wu, 2005). Finally, in industry, mobile phone providers are interested in comparison of the cell phone download activity in different locations (Degras et al., 2012).

In the statistical literature, the problem of testing whether the observed time series all have the same trend has been widely studied and tests for equality of trend or regression curves have been developed in Härdle and Marron (1990), Hall and Hart (1990), Delgado (1993) and Degras et al. (2012) among many others. Versions of model (1.1) with a parametric trend are considered in Vogelsang and Franses (2005), Sun (2011) and Xu (2012) among others. In the nonparametric context, Li et al. (2010), Atak et al. (2011), Robinson (2012) and Chen et al. (2012) studied panel models where the observed time series have a common time trend. However, in many applications the assumption of a common time trend is questionable at best. For example, when we observe a large number of time series, it is reasonable to expect that at least some of the time trends are different from the others.

This leads us to more flexible panel settings with heterogeneous trends which have been studied, for example, in Degras et al. (2012), Zhang et al. (2012) and Hidalgo and Lee (2014). Degras et al. (2012) consider the problem of testing H_0 in a model that is a special case of (1.1) which does not include additional regressors. Chen and Wu (2018) develop theory for a very similar model framework but under more general conditions on the error terms. Zhang et al. (2012) investigate the problem of testing the hypothesis H_0 in a slightly restricted version of model (1.1), where $\beta_i = \beta$ for all i. These test have an important drawback: they involve classical nonparametric estimation of the trend functions that depends on one or several bandwidth parameters. This is a very important limitation of the applicability of such tests since in most cases it is far from clear how to choose such parameters in an appropriate way. On the contrary, our multiscale method allows us to consider a large collection of bandwidths simultaneously, thus, avoiding the problem of choosing only one bandwidth.

Recently, Khismatullina and Vogt (2021) proposed a new inference method that allows to detect differences between epidemic time trends in the context of the COVID-19 pandemic. They presented a statistically rigorous procedure that not only allows to compare trends across different countries, but to pinpoint the time intervals where the differences occur as well. Moreover, they also circumvented the need to pick a bandwidth parameter by using a multiscale testing procedure. However, the model that the authors considered is only a special case of the model (1.1) which does not include neither the covariates $\mathbf{X}_{it}$, nor the fixed effects α_i, and they restricted the error terms ε_{it} to be independent across t. Our model (1.1), which can be regarded as a generalization of the one that was studied in Khismatullina and Vogt (2021), allows for a wider range of economic and financial applications.

The main theoretical contribution of the current paper is the multiscale method that allows to make simultaneous confidence statements about the regions where the time trends differ. We believe that currently there are no equivalent statistical methods. Even though tests for equality of the trends have been developed already for a while, most existing procedures allow only to test whether the trend curves are all the same or not, but they almost never allow to infer which curves are different and where. To the best of our knowledge, the only two exceptions are Khismatullina and Vogt (2021) whose contribution is briefly discussed above and Park et al. (2009) who developed SiZer methods for the comparison of nonparametric trend curves in a strongly simplified version of model (1.1). Moreover, Park et al. (2009) derive theoretical results for their analysis only for the special case $n = 2$, that is, when only two time series are observed. In case of $n > 2$, the algorithm is provided without proof.

The structure of the paper is as follows. Section 2 introduces the model setting and the necessary technical assumptions that are required for the theory. The multiscale test is developed step by step in Section 3. The main theoretical results are presented in Section 4. To keep the discussion as clear as possible, we include in the main text of the paper only the essential parts of the theoretical arguments, whereas the technical details and extended proofs are deferred to the Appendix. Section 6 concludes.

2 The model

Throughout the paper, we adopt the following notation. For a vector $\mathbf{v} = (v_1, \ldots, v_m) \in \mathbb{R}^m$, we write $|\mathbf{v}| = \big(\sum_{i=1}^m v_i^2\big)^{1/2}$ and $|\mathbf{v}|_q = \big(\sum_{i=1}^m v_i^q\big)^{1/q}$ respectively. 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 the particular 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:

$$\delta_q(\mathbf{L}, t) = ||\mathbf{L}(\mathcal{F}_t) - \mathbf{L}(\mathcal{F}_t')||_q, \tag{2.1}$$

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. Intuitively, $\delta_q(\mathbf{L}, t)$ measures the dependency of $\mathbf{L}(\mathcal{F}_t)$ on ϵ_0, i.e., how replacing ϵ_0 by an i.i.d. copy while keeping all other innovations in place affects the output $\mathbf{L}(\mathcal{F}_t)$.

2.1 Setting

As was already briefly discussed in the Introduction, the model setting is as follows. We observe a panel of n time series $\mathcal{Z}_i = \{(Y_{it}, \mathbf{X}_{it}) : 1 \leq t \leq T\}$ of length T for $1 \leq i \leq n$. Each time series $\mathcal{Z}_i$ satisfies the model equation

$$Y_{it} = \boldsymbol{\beta}_i^\top \mathbf{X}_{it} + m_i\Big(\frac{t}{T}\Big) + \alpha_i + \varepsilon_{it} \tag{2.2}$$

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 $[0, 1]$ with $\int_0^1 m_i(u)du = 0$ for all i, 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 common in nonparametric regression, the trend functions m_i in model (2.2) depend on rescaled time t/T rather than on real time t. Using rescaled time is equivalent to restricting the domain of the functions to the unit interval which in turn allows us to apply the usual asymptotic arguments. Discussion about the application of the rescaled time in the context of nonparametric estimation can be found in Robinson (1989), Dahlhaus (1997) and Vogt and Linton (2014). The condition $\int_0^1 m_i(u)du = 0$ for all i is necessary identification condition due the presence of α_i. Without imposing this condition, we can freely increase or decrease the functions m_i by any constant c_i while simultaneously subtract or add the same constant to the intercept term α_i:

$$Y_{it} = [m_i(t/T) + c_i] + \boldsymbol{\beta}_i^\top \mathbf{X}_{it} + [\alpha_i - c_i] + \varepsilon_{it}.$$

We also assume that all the trend functions $m_i(\cdot)$ are continuously differentiable on $[0, 1]$. The term α_i can be regarded as an additional error component. In the econometrics literature, it is commonly called a fixed effect. It is often interpreted as the term that captures 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 (2.2), 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, throughout the paper we restrict attention to the case where the number of time series n in model (2.2) 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.

2.2 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 $q > 4$.

The condition (C1) can be translated as the restriction on the error process $\mathcal{E}_i$ to be stationary and causal (in a sense that ε_{it} does not depend on the future innovations η_{is}, $s > t$). The class of error processes that satisfies the condition (C1) is massive, and includes linear processes, their nonlinear transformation, as well as a large variety of nonlinear processes such as Markov chain models and nonlinear autoregressive models (Wu and Wu, 2016).
Following Wu (2005), we impose conditions on the dependence structure of the error processes $\mathcal{E}_i$ in terms of the physical dependence measure $\delta_q(G_i, t)$ definedin (2.1). 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 $\tau_q = \{q^2 - 4 + (q-2)\sqrt{q^2 + 20q + 4}\}/8q$.

For a fixed t, $\Theta_{i,t,q}$ measures the cumulative effect of η_0 on $(\varepsilon_{is})_{s \geq t}$. Condition (C3) assumes that the overall cumulative effect is finite and puts some restrictions on the rate of decay of $\Theta_{i,t,q}$.
The condition (C3) is fulfilled by a wide range of stationary processes $\mathcal{E}_i$. For a detailed discussion of an assumption (C3), as well as the assumptions (C1)–(C2) and some examples of the error processes that satisfy these conditions, see Khismatullina and Vogt (2020).
Regarding the independent variables $\mathbf{X}_{it}$, we assume the following 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 $\mathbf{H}_i := (H_{i1}, H_{i2}, \ldots, H_{id})^\top : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^d$ being a measurable function such that $\mathbf{H}_i(\mathcal{U}_{it})$ is well defined. We denote $\mathcal{U}_{it} = (\ldots, u_{it-1}, u_{it})$.

(C5) Let $\mathbf{N}_i$ 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 $\mathbf{N}_i$ 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 some $q' > \max\{2\theta, 4\}$, where θ will be introduced further in Assumption (C12).

(C7) $\sum_{s=0}^{\infty} \delta_{q'}(\mathbf{H}_i, s) < \infty$ for q' from Assumption (C6).

(C8) For each i it holds that $\sum_{s=t}^{\infty} \delta_{q'}(\mathbf{H}_i, s) = O(t^{-\alpha})$ for q' from Assumption (C6) and for some $\alpha > 1/2 - 1/q'$.

As with the error processes $\mathcal{E}_i$, $\mathbf{X}_i$ is guaranteed to be stationary and causal by Assumption (C4). Assumptions (C5) and (C6) are technical conditions that prevents asymptotic multicollinearity and ensures that all the necessary moments exist, respectively. Moreover, we also employ the defintion of the physical dependence measure $\delta_q(\cdot, \cdot)$ in Assumptions (C7) - (C8), that make certain that the cumulative effect of the innovation u_0 on $(\mathbf{X}_{it})_{t\geq 0}$ is finite.
To be able to prove the main theorems in Section 3, 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 $t, s \in \{1, \dots, T\}$.

(C10) Let $\zeta_{i,t} = (u_{it}, \eta_{it})^\top$. Define $\mathcal{I}_{it} = (\dots, \zeta_{i,t-1}, \zeta_{i,t})$ and $\mathbf{U}_i(\mathcal{I}_{it}) = \mathbf{H}_i(\mathcal{U}_{it})G_i(\mathcal{J}_{it})$. With this notation at hand, we assume that $\sum_{s=0}^{\infty} \delta_2(\mathbf{U}_i, s) < \infty$.

Assumption (C9) is a slightly relaxed independence assumption: even though we do not require the covariates $\mathbf{X}_{it}$ to be completely independent with the error terms ε_{it}, our theoretical results depend upon them being uncorrelated. We in particular need this assumption in order to prove asymptotic consistency for the differencing estimator $\widehat{\boldsymbol{\beta}}_i$ of $\boldsymbol{\beta}_i$ proposed in Section 5.1. In principle, it would be possible to relax this assumption even further, but that would involve much more complicated estimation procedure of $\boldsymbol{\beta}_i$ and more arduous technical arguments. Assumption (C10) ensures short-range dependence among the variables in our model. Again, we can interpret this as the fact that the cumulative effect of a single error on all future values is bounded.
We employ these assumptions to prove the main theoretical results in our paper. For detailed proofs, we refer the reader to the Appendix.

Remark 2.1. *The conditions (C4)–(C10) can be relaxed to cover nonstationary regressors as well as stationary ones. For example, (C4) will then be replaced by*

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

The other assumptions can be adjusted accordingly. However, for the sake of clarity, we restrict our attention only to stationary covariates $\mathbf{X}_{it}$.

3 Testing procedure

In this section, we develop a multiscale testing procedure for the problem of comparison of the trend curves m_i in model (2.2). 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.

3.1 Preliminary steps

Testing the null hypothesis $H_0 : m_1 = m_2 = \ldots = m_n$ in model (2.2) is not a trivial task not only because it involves nonparametric estimation of the functions $m_i(\cdot)$, but also due to the presence an unknown fixed term α_i and a vector of unknown parameters $\boldsymbol{\beta}_i$. It is clear 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. However, in reality the variables Y_{it}° are not observed since the intercept α_i and the coefficients $\boldsymbol{\beta}_i$ are not known. Nevertheless, given appropriate estimators $\widehat{\alpha}_i$ and $\widehat{\boldsymbol{\beta}}_i$, we can consider

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

Thus, the unobserved variables Y_{it}° can be approximated by $\widehat{Y}_{it}$, and in what follows we show under some mild conditions on $\widehat{\alpha}_i$ and $\widehat{\boldsymbol{\beta}}_i$, this approximation is indeed sufficient for our analysis.

First, we focus on the estimation of the vector of unknown parameters $\boldsymbol{\beta}_i$. We construct the estimator $\widehat{\boldsymbol{\beta}}_i$ in the following way.

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 can write

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

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 (by our assumption that $m_i(\cdot)$ is continuously differentiable on $[0,1]$), we can use the fact that $\big|m_i\big(\frac{t}{T}\big) - m_i\big(\frac{t-1}{T}\big)\big| = O\big(\frac{1}{T}\big)$ and rewrite

$$
\Delta Y_{it} = \boldsymbol{\beta}_i^\top \Delta \mathbf{X}_{it} + \Delta\varepsilon_{it} + O\Big(\frac{1}{T}\Big). \tag{3.1}
$$

Now, for each i we employ the least squares estimation method to estimate $\boldsymbol{\beta}_i$ in (3.1), treating $\Delta\mathbf{X}_{it}$ as the regressors and ΔY_{it} as the response variable. That is, we propose the following differencing estimator:

$$\widehat{\boldsymbol{\beta}}_i = \Big(\sum_{t=2}^{T} \Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^\top\Big)^{-1} \sum_{t=2}^{T} \Delta\mathbf{X}_{it}\Delta Y_{it} \tag{3.2}$$

We will show in Section 5.1 that $\widehat{\boldsymbol{\beta}}_i$ is a consistent estimator of $\boldsymbol{\beta}_i$ with the property $\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i = O_P(T^{-1/2})$.

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

$$\begin{aligned}\widehat{\alpha}_i &= \frac{1}{T}\sum_{t=1}^{T}\big(Y_{it} - \widehat{\boldsymbol{\beta}}_i^\top\mathbf{X}_{it}\big) = \frac{1}{T}\sum_{t=1}^{T}\big(\boldsymbol{\beta}_i^\top\mathbf{X}_{it} - \widehat{\boldsymbol{\beta}}_i^\top\mathbf{X}_{it} + \alpha_i + m_i(t/T) + \varepsilon_{it}\big) = \\ &= \big(\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i\big)^\top\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_{it} + \alpha_i + \frac{1}{T}\sum_{i=1}^{T} m_i(t/T) + \frac{1}{T}\sum_{i=1}^{T}\varepsilon_{it}.\end{aligned} \tag{3.3}$$

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 (3.3), we get that $\widehat{\alpha}_i - \alpha_i = O_P(T^{-1/2})$. Thus, the unobserved variables $Y_{it}^\circ := Y_{it} - \boldsymbol{\beta}_i^\top\mathbf{X}_{it} - \alpha_i = m_i(t/T) + \varepsilon_{it}$ can be well approximated by $\widehat{Y}_{it}$ since $\widehat{Y}_{it} = Y_{it} - \widehat{\alpha}_i - \widehat{\boldsymbol{\beta}}_i^\top\mathbf{X}_{it} = Y_{it}^\circ + O_P(T^{-1/2})$.

We now turn to the estimator of the long-run error variance $\sigma_i^2 = \sum_{\ell=-\infty}^{\infty}\mathrm{Cov}(\varepsilon_{i0}, \varepsilon_{i\ell})$. For the moment, we assume that the long-run variance does not depend on i, that is $\sigma_i^2 = \sigma^2$ for all i. We will need this further for conducting the testing procedure properly. Nevertheless, we keep the indices throughout the paper in order to be congruous in notation. We further let $\widehat{\sigma}_i^2$ be an estimator of σ_i^2 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 $1 \le t \le T$. Hence, whereas the true long-run variance is the same for all time series, the estimators are different. Throughout the section, we assume that $\widehat{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$ where the conditions on ρ_T will be provided further in Section 4. Details on how to construct $\widehat{\sigma}_i^2$ are deferred to Section 5.2.

3.2 Construction of the test statistics

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 and for any location-bandwidth pair (u, h), 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:

$$w_{t,T}(u,h) = \frac{\Lambda_{t,T}(u,h)}{\{\sum_{t=1}^{T} \Lambda_{t,T}(u,h)^2\}^{1/2}}, \tag{3.4}$$

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 $C > 0$.

Assumption (C11) allows us to use the common kernel functions such as rectangular, Epanechnikov and Gaussian kernels.

We regard the kernel average $\widehat{\psi}_{ij,T}(u,h)$ as a measure of the distance between the two trend curves m_i and m_j on the interval $\mathcal{I}_{(u,h)} = [u-h, u+h]$.

Instead with working directly with the kernel averages $\widehat{\psi}_{ij,T}(u,h)$, we replace them by their normalized version:

$$\hat{\psi}^0_{ij,T}(u,h) = \Big|\frac{\widehat{\psi}_{ij,T}(u,h)}{(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2)^{1/2}}\Big| - \lambda(h), \tag{3.5}$$

where $\lambda(h) = \sqrt{2\log\{1/(2h)\}}$ is an additive correction term that balances the significance of many test statistics that correspond to different values of bandwidth parameters (see the discussion on this topic and comparison between multiscale testing procedures with and without this correction term in Khismatullina and Vogt (2020)).

We now aggregate the test statistics $\hat{\psi}^0_{ij,T}(u,h)$ for all i and j and a wide range of different locations u and bandwidths (or scales) h:

$$\widehat{\Psi}_{n,T} = \max_{1\le i<j\le n} \max_{(u,h)\in\mathcal{G}_T} \hat{\psi}^0_{ij,T}(u,h), \tag{3.6}$$

In (3.6), $\mathcal{G}_T$ stands for the set of location-bandwidth pairs (u,h) that was mentioned in the Introduction. We use the subscript T in $\mathcal{G}_T$ to point out that the choice of the grid depends on the sample size T. Specifically, throughout the paper, we suppose that $\mathcal{G}_T$ is some subset of $\mathcal{G}_T^{\text{full}} = \{(u,h) : u = t/T$ for some $1 \le t \le T$ and $h \in [h_{\min}, h_{\max}]\}$, where $h_{\min}$ and $h_{\max}$ denote some minimal and maximal bandwidth value, respectively. As was discussed in the Introduction, we assume that the set of intervals $\{\mathcal{I}_{(u,h)} = [u-h, u+h] : (u,h) \in \mathcal{G}_T\}$ covers the whole unit interval. Furthermore, for our theoretical results, we require the following additional 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 $h_{\max} < 1/2$.

Assumption (C12) places relatively mild restrictions on the grid $\mathcal{G}_T$: we allow the grid to grow with the sample size but only at a polynomial rate T^θ with fixed θ. This is not a severe constraint because under this limitation, we can still work with the full set of location-bandwidth points $\mathcal{G}_T = \mathcal{G}_T^{\text{full}}$ which is more than enough for most applied problems. Assumption (C13) concerns the minimal and the maximal bandwidths that we use for our analysis. Specifically, according to Assumption (C13), we can choose the minimal bandwidth $h_{\min}$ that converges to zero slower than $T^{-(1-\frac{2}{q})} \log T$ as the sample size T goes to infinity. $h_{\max}$ can be picked very large.
Note that the value $\max_{(u,h)\in\mathcal{G}_T} \hat{\psi}^0_{ij,T}(u,h)$ simultaneously takes into account all intervals $\mathcal{I}_{(u,h)} = [u-h, u+h]$ with $(u,h) \in \mathcal{G}_T$. Thus, it can be interpreted as a global distance measure between the two curves m_i and m_j, and the test statistics $\widehat{\Psi}_{n,T}$ is then defined as the maximal distance between any pair of curves m_i and m_j with $i \neq j$.
In Section 3.3, we will show how to test the null hypothesis $H_0 : m_1 = m_2 = \ldots = m_n$ using the mutliscale test statistics $\widehat{\Psi}_{n,T}$.

3.3 The testing procedure

Let Z_{it} for $1 \leq t \leq T$ and $1 \leq i \leq n$ be independent standard normal random variables which are independent of the error terms ε_{js} and the covariates $\mathbf{X}_{js}$ for all $1 \leq s \leq T$ and $1 \leq j \leq n$. 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 the notation, we will omit the subscript T in $\bar{Z}_{i,T}$ in what follows. As before, for each i and j, we introduce the normalized Gaussian statistic

$$\phi^0_{ij,T}(u,h) = \left| \frac{\phi_{ij,T}(u,h)}{(\sigma_i^2 + \sigma_j^2)^{1/2}} \right| - \lambda(h), \tag{3.7}$$

where $\phi_{ij,T}(u,h) = \sum_{t=1}^T w_{t,T}(u,h) \big\{ \sigma_i(Z_{it} - \bar{Z}_i) - \sigma_j(Z_{jt} - \bar{Z}_j) \big\}$ with $w_{t,T}(u,h)$ was defined in (3.4).
Next, similarly to (3.6) we define the global Gaussian test statistics

$$\Phi_{n,T} = \max_{1\leq i<j\leq n} \max_{(u,h)\in\mathcal{G}_T} \phi^0_{ij,T}(u,h) \tag{3.8}$$

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 $\widehat{\Psi}_{n,T} > q_{n,T}(\alpha)$.

Remark 3.1. *To prove the theoretical results in Section 4, we will use the following fact. By our assumption that the long-run variance σ_i^2 does not depend on i (i.e. $\sigma_i^2 = \sigma_j^2 = \sigma^2$), we can rewrite the Gaussian statistics (3.7) as follows:*

$$\phi_{ij,T}^0(u,h) = \frac{1}{\sqrt{2}}\Big| \sum\nolimits_{t=1}^T w_{t,T}(u,h)\left\{(Z_{it} - \bar{Z}_i) - (Z_{jt} - \bar{Z}_j)\right\}\Big| - \lambda(h),$$

which means that the distribution of the Gaussian test statistics does not depend neither on the data $\mathcal{Z}_i = \{(Y_i, \mathbf{X}_i) : 1 \leq t \leq T)\}, \mathcal{Z}_j = \{(Y_j, \mathbf{X}_j) : 1 \leq t \leq T)\}$, nor on any unknown quantities (such as σ_i^2 or σ_j^2) and thus can be regarded as known. In addition to exploiting this fact while proving the theoretical results, we will also use it for calculating (approximately) the quantiles of $\Phi_{n,T}$ by the Monte Carlo simulations in Section 3.5. However, in what follows, we will stick to the definition (3.7) of $\phi_{ij,T}^0(u,h)$, which involves the long-run variances σ_i and σ_j, for the sake of similarity to $\hat{\psi}_{ij,T}^0(u,h)$.

Remark 3.2. *By construction, the $(1-\alpha)$ Gaussian quantile $q_{n,T}(\alpha)$ depends not only on the number of times series considered n and the sample size T, but on the choice of the set of location-bandwidth pairs $\mathcal{G}_T$ as well. However, we do not include this dependence in the definition explicitly since we believe it will only lead to the unnecessary complication of the notation.*

3.4 Locating the differences

Suppose we reject the null hypothesis H_0. Unfortunately, that does not provide us with a lot of information about the behaviour of the trend functions $m_i(\cdot)$. After performing the test described in Section 3.3, we can only state the fact that some of the trend functions are not equal somewhere on $[0,1]$ (with a given statistical confidence), but we can not tell which of the functions are different and where they differ. Hence, we need an additional step in the testing procedure in order to locate those differences.
Formally, for a given pair of time series (i,j) and for any given interval $\mathcal{I}_{(u,h)} = [u-h, u+h]$ such that $(u,h) \in \mathcal{G}_T$ we consider the hypothesis

$$H_0^{[i,j]}(u,h) : m_i(w) = m_j(w) \text{ for all } w \in [u-h, u+h].$$

Then the multiscale test of the hypothesis $H_0^{[i,j]}(u,h)$ is defined as follows:

> *For a given significance level $\alpha \in (0,1)$, we reject $H_0^{[i,j]}(u,h)$ if $\hat{\psi}_{ij,T}^0(u,h) > q_{n,T}(\alpha)$.*

For each pair of time series (i,j), denote the set of intervals $\mathcal{I}_{(u,h)}$ that consists of the intervals where we reject $H_0^{[i,j]}(u,h)$ at a significance level α by $\mathcal{S}^{[i,j]}(\alpha)$. We will prove later in Section 4, that we can make the following confidence statements:

> *We can state with (asymptotic) probability $1-\alpha$ that for all i,j, $1 \leq i < j \leq n$, we have that $m_i(\cdot)$ and $m_j(\cdot)$ differ on all of the intervals $\mathcal{I}_{[u,h]} \in \mathcal{S}^{[i,j]}(\alpha)$.*

3.5 Implementation of the test in practice

In practice, we implement the test procedure described in Sections 3.3 and 3.4 in the following way.

Step 1. Fix a significance level $\alpha \in (0,1)$.

Step 2. Compute the (approximated) quantile $q_{n,T}(\alpha)$ by Monte Carlo simulations. Specifically, draw a large number N (say $N = 5000$) of samples of independent standard normal random variables $\{Z_{it}^{(\ell)} : 1 \le t \le T, 1 \le i \le n\}$ for $1 \le \ell \le N$. For each sample ℓ, compute the value $\Phi_{n,T}^{(\ell)}$ of the Gaussian test statistics $\Phi_{n,T}$ and store them. Calculate the empirical $(1-\alpha)$-quantile $\hat{q}_{n,T}(\alpha)$ from the stored values $\{\Phi_T^{(\ell)} : 1 \le \ell \le N\}$. Use $\hat{q}_{n,T}(\alpha)$ as an approximated value of the quantile $q_{n,T}(\alpha)$.

Step 3. Carry out the test for the global hypothesis H_0 by calculating $\widehat{\Psi}_{n,T}$ and checking if $\widehat{\Psi}_{n,T} > q_{n,T}(\alpha)$. Reject the null if it is true.

Step 4. For each i, j, $1 \le i < j \le n$, and each $(u,h) \in \mathcal{G}_T$, carry out the test for the local null hypothesis $H_0^{[i,j]}(u,h)$ by checking if $\hat{\psi}_{ij,T}^0(u,h) > q_{n,T}(\alpha)$. Store the test results in the variable $r_{ij,T}(u,h) = \mathbb{1}(|\hat{\psi}_{ij,T}^0(u,h)| > q_{n,T}(\alpha))$, where $\mathbb{1}(\cdot)$ is an indicator function that is equal to 1 if the condition inside the brackets is met and 0 otherwise.

Step 5. Display the results. One of the possible ways to do that is to produce a separate plot for each of the pairwise comparisons and draw only the intervals where we reject the corresponding "local" null. Formally, on each of the plots that present the results of the comparison of time series i and j, we display the intervals $\mathcal{I}_{(u,h)} = [u-h, u+h] \in \mathcal{S}^{[i,j]}(\alpha)$, i.e. the (rescaled) time intervals where we reject the null $H_0^{[i,j]}(u,h)$.

4 Theoretical properties of the test

In order to invesitgate the theoretical properties of our multiscale test, we introduce an auxiliary test statistic

$$\widehat{\Phi}_{n,T} = \max_{1\le i<j\le n} \max_{(u,h)\in\mathcal{G}_T} \widehat{\phi}_{ij,T}^0(u,h), \tag{4.1}$$

where

$$
\begin{aligned}
\widehat{\phi}^0_{ij,T}(u,h) &= \Big|\frac{\widehat{\phi}_{ij,T}(u,h)}{\{\widehat{\sigma}_i^2+\widehat{\sigma}_j^2\}^{1/2}}\Big| - \lambda(h), \\
\widehat{\phi}_{ij,T}(u,h) &= \sum_{t=1}^T w_{t,T}(u,h)\big\{(\varepsilon_{it}-\bar{\varepsilon}_i) + (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top(\mathbf{X}_{it}-\bar{\mathbf{X}}_i) \\
&\qquad - (\varepsilon_{jt}-\bar{\varepsilon}_j) - (\boldsymbol{\beta}_j - \widehat{\boldsymbol{\beta}}_j)^\top(\mathbf{X}_{jt}-\bar{\mathbf{X}}_j)\big\}.
\end{aligned}
$$

Here we denote $\bar{\varepsilon}_i = \bar{\varepsilon}_{i,T} := T^{-1}\sum_{t=1}^T \varepsilon_{it}$ and $\bar{\mathbf{X}}_i = \bar{\mathbf{X}}_{i,T} := T^{-1}\sum_{t=1}^T \mathbf{X}_{it}$. Note that under the null, we have $\widehat{\phi}^0_{ij,T}(u,h) = \widehat{\psi}^0_{ij,T}(u,h)$ and $\widehat{\Phi}_{n,T} = \widehat{\Psi}_{n,T}$. Hence, in order to determine the distribution of $\widehat{\Psi}_{n,T}$ under the null, we can simply study the behaviour of $\widehat{\Phi}_{n,T}$. And our first theoretical result characterizes the asymptotic behaviour of the statistic $\widehat{\Phi}_{n,T}$.

Theorem 4.1. *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 for all i, $m_i(\cdot)$ is a continuously differentiable function on $[0,1]$ satisfying the property $\int_0^1 m_i(u)du = 0$. Furthermore, for all i, $i \in \{1,\dots,n\}$ we have $\sigma_i^2 = \sigma^2$, $\widehat{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$ and $\widehat{\widehat{\sigma}}{}_i^2 = \sigma_i^2 + o_p(\rho_T)$ with $\rho_T = o(\sqrt{h_{\min}}/\log T)$. Then*

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

Theorem 4.1 is the main stepping stone to derive the theoretical properties of our multiscale test. The full proof of the theorem is provided in the Appendix. Here we present the main stepping stones in the proof.

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

5.1 Estimation of $\boldsymbol{\beta}_i$

As was already mentioned in Section 3.1, for each i, we construct a differencing estimator $\widehat{\boldsymbol{\beta}}_i$ of the vector of unknown parameters $\boldsymbol{\beta}_i$ using the first differences:

$$\widehat{\boldsymbol{\beta}}_i = \Big(\sum_{t=2}^T \Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^\top\Big)^{-1}\sum_{t=2}^T \Delta\mathbf{X}_{it}\Delta Y_{it} \tag{5.1}$$

where $\Delta\mathbf{X}_{it} = \mathbf{X}_{it} - \mathbf{X}_{it-1}$ and $\Delta Y_{it} = Y_{it} - Y_{it-1}$. The asymptotic consistency for this differencing estimator is given by the following theorem:

Theorem 5.1. *Under the conditions of Theorem 4.1, 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 (5.1).

Detailed proof of the Theorem 5.1 is provided in the Appendix. Here we briefly outline the main steps of the proof.

After rearranging the terms, we can write

$$\begin{aligned}\sqrt{T}(\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i) =& \Big(\frac{1}{T}\sum_{t=2}^{T} \Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^{\top}\Big)^{-1} \frac{1}{\sqrt{T}}\sum_{t=2}^{T} \Delta\mathbf{X}_{it}\Delta m_{it} \\ &+ \Big(\frac{1}{T}\sum_{t=2}^{T} \Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^{\top}\Big)^{-1} \frac{1}{\sqrt{T}}\sum_{t=2}^{T} \Delta\mathbf{X}_{it}\Delta \varepsilon_{it},\end{aligned} \tag{5.2}$$

where $\Delta m_{it} = m_i\left(\frac{t}{T}\right) - m_i\left(\frac{t-1}{T}\right)$ and $\Delta\varepsilon_{it} = \varepsilon_{it} - \varepsilon_{it-1}$.

We look at each part of (5.2) separately. First, by Assumption (C6) and applying Chebyshev's and Cauchy-Schwarz inequalities we show that

$$\frac{1}{\sqrt{T}}\sum_{t=2}^{T} \Delta\mathbf{X}_{it}\Delta m_{it} = \frac{1}{\sqrt{T}}\sum_{t=2}^{T} \Delta\mathbf{H}_i(\mathcal{U}_{it})\Delta m_{it} = O_P\Big(\frac{1}{\sqrt{T}}\Big).$$

Then, by similar arguments and applying Proposition A.10, we have that

$$\left|\Big(\frac{1}{T}\sum_{t=2}^{T} \Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^{\top}\Big)^{-1}\right| = O_P(1),$$

where $|A|$ with A being a matrix is any matrix norm.

Taking these two facts together, we have shown that the first summand in (5.2) is $O_P(1/\sqrt{T})$.

Finally, we turn our attention to the second summand in (5.2). We already know that $\left|\Big(\frac{1}{T}\sum_{t=2}^{T} \Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^{\top}\Big)^{-1}\right| = O_P(1)$. Moreover, by Proposition A.13,

$$\left|\frac{1}{\sqrt{T}}\sum_{t=2}^{T} \Delta\mathbf{X}_{it}\Delta\varepsilon_{it}\right| = O_P(1).$$

Hence, we have that

$$\Big(\frac{1}{T}\sum_{t=2}^{T} \Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^{\top}\Big)^{-1} \frac{1}{\sqrt{T}}\sum_{t=2}^{T} \Delta\mathbf{X}_{it}\Delta \varepsilon_{it} = O_P(1). \tag{5.3}$$

The statement of the theorem follows.

5.2 Estimation of σ_i^2

6 Conclusion

Consider the situation that the null hypothesis $H_0 : m_1 = \ldots = m_n$ is violated in the general panel data model (1.1). Even though some of the trend functions m_i are different in this case, there may still be groups of time series with the same time trend.

Formally, a group stucture can be defined as follows within the framework of model (1.1): There exist sets or groups of time series $G_1, \ldots, G_{K_0}$ with $\{1, \ldots, n\} = \dot{\bigcup}_{k=1}^{K_0} G_k$ such that for each $1 \leq k \leq K_0$,

$$m_i = m_j \quad \text{for all } i, j \in G_k. \tag{6.1}$$

According to (6.1), the time series of a given group G_k all have the same time trend. In many applications, it is very natural to suppose that there is such a group structure in the data. An interesting statistical problem which we aim to investigate in our project is how to estimate the unknown groups $G_1, \ldots, G_{K_0}$ and their unknown number K_0 from the data.

The problem of estimating the unknown groups $G_1, \ldots, G_{K_0}$ and their unknown number K_0 in model (1.1) has close connections to functional data clustering. There, the aim is to cluster smooth random curves that are functions of (rescaled) time and that are observed with or without noise. A number of different clustering approaches have been proposed in the context of functional data models; see for example Abraham et al. (2003), Tarpey and Kinateder (2003) and Tarpey (2007) for procedures based on k-means clustering, James and Sugar (2003) and Chiou and Li (2007) for model-based clustering approaches and Jacques and Preda (2014) for a recent survey.

The problem of finding the unknown group structure in model (1.1) is also closely related to a developing literature in econometrics which aims to identify unknown group structures in parametric panel regression models. In its simplest form, the panel regression model under consideration is given by the equation $Y_{it} = \beta_i^\top X_{it} + u_{it}$ for $1 \leq t \leq T$ and $1 \leq i \leq n$, where the coefficient vectors β_i are allowed to vary across individuals i and the error terms u_{it} may include fixed effects. Similar to the trend functions in model (1.1), the coefficients β_i are assumed to belong to a number of groups: there are K_0 groups $G_1, \ldots, G_{K_0}$ such that $\beta_i = \beta_j$ for all $i, j \in G_k$ and all $1 \leq k \leq K_0$. The problem of estimating the unknown groups and their unknown number has been studied in different versions of this modelling framework; cp. Su et al. (2016), Su and Ju (2018) and Wang et al. (2018) among others. Bonhomme and Manresa (2015) considered a related model where the group structure is not imposed on the regression coefficients but rather on some unobserved time-varying fixed effect components of the panel model.

Virtually all the proposed procedures to cluster nonparametric curves in panel and functional data models related to (1.1) depend on a number of bandwidth or smoothing parameters required to estimate the nonparametric functions m_i. In general, nonparametric curve estimators are strongly affected by the chosen bandwidth parameters. A clustering procedure which is based on such estimators can be expected to be strongly influenced by the choice of bandwidths as well. Moreover, as in the context of statistical testing, there is no theory available on how to pick the bandwidths optimally for the clustering problem. Hence, as in the context of testing, it is desirable to construct a clustering procedure which is free of bandwidth or smoothing parameters that need to be selected.

One way to obtain a clustering method which does not require to select any bandwidth parameter is to use multiscale methods. This approach has recently been taken in **?**. They develop a clustering approach in the context of the panel model $Y_{it} = m_i(X_{it}) + u_{it}$, where X_{it} are random regressors and u_{it} are general error terms that may include fixed effects. Imposing the same group structure as in (6.1) on their model, they construct estimators of the unknown groups and their unknown number as follows: In a first step, they develop bandwidth-free multiscale statistics $\hat{d}_{ij}$ which measure the distance between pairs of functions m_i and m_j. To construct them, they make use of the multiscale testing methods described in part (a) of this section. In a second step, the statistics $\hat{d}_{ij}$ are employed as dissimilarity measures in a hierarchical clustering algorithm.

A Appendix

Here, we prove the theoretical results from Sections 4 and 5. 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 $x \in \mathbb{R}, x \geq 0$, we write $\lfloor x \rfloor$ and $\lceil x \rceil$ to denote the greatest integer less than or equal and the smallest integer greater than or equal to x, respectively. 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.

A.1 Statistics used in the Appendix

In the proof of Theorem 4.1, we use a number of different test statistics, either already defined in Section 3 or the auxiliary statistics defined below. Each of these statistics plays an important role in one or more steps of the proof. In the following list, we present these test statistics, describe how they are constructed and explain in which parts of the proof they are used.

- Our main multiscale statistics defined in (3.6). It is calculated based on data.

$$\begin{aligned}
\widehat{\Psi}_{n,T} &= \max_{1 \le i < j \le n} \max_{(u,h) \in \mathcal{G}_T} \widehat{\psi}^0_{ij,T}(u,h), \\
\widehat{\psi}^0_{ij,T}(u,h) &= \Big| \frac{\widehat{\psi}_{ij,T}(u,h)}{(\widehat{\sigma}^2_i + \widehat{\sigma}^2_j)^{1/2}} \Big| - \lambda(h), \\
\widehat{\psi}_{ij,T}(u,h) &= \sum_{t=1}^{T} w_{t,T}(u,h)(\widehat{Y}_{it} - \widehat{Y}_{jt}).
\end{aligned} \tag{A.1}$$

- The Gaussian statistics that is used to calculate the critical values (defined in (3.8)).

$$\begin{aligned}
\Phi_{n,T} &= \max_{1 \le i < j \le n} \max_{(u,h) \in \mathcal{G}_T} \phi^0_{ij,T}(u,h), \\
\phi^0_{ij,T}(u,h) &= \Big| \frac{\phi_{ij,T}(u,h)}{(\sigma^2_i + \sigma^2_j)^{1/2}} \Big| - \lambda(h), \\
\phi_{ij,T}(u,h) &= \sum\nolimits_{t=1}^{T} w_{t,T}(u,h) \big\{ \sigma_i (Z_{it} - \bar{Z}_i) - \sigma_j (Z_{jt} - \bar{Z}_j) \big\}.
\end{aligned} \tag{A.2}$$

- Auxiliary statistics that can be regarded as the version of our multiscale statistic under H_0 (defined in (4.1)). Our main theoretical result (Theorem 4.1) investigates the distribution of $\widehat{\Phi}_{n,T}$.

$$\begin{aligned}
\widehat{\Phi}_{n,T} &= \max_{1 \le i < j \le n} \max_{(u,h) \in \mathcal{G}_T} \widehat{\phi}^0_{ij,T}(u,h), \\
\widehat{\phi}^0_{ij,T}(u,h) &= \Big| \frac{\widehat{\phi}_{ij,T}(u,h)}{\{\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\}^{1/2}} \Big| - \lambda(h), \qquad \text{(A.3)} \\
\widehat{\phi}_{ij,T}(u,h) &= \sum_{t=1}^T w_{t,T}(u,h) \big\{ (\varepsilon_{it} - \bar{\varepsilon}_i) + (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top (\mathbf{X}_{it} - \bar{\mathbf{X}}_i) \\
&\qquad - (\varepsilon_{jt} - \bar{\varepsilon}_j) - (\boldsymbol{\beta}_j - \widehat{\boldsymbol{\beta}}_j)^\top (\mathbf{X}_{jt} - \bar{\mathbf{X}}_j) \big\}.
\end{aligned}$$

- Intermediate statistics that is close to $\widehat{\Phi}_{n,T}$ but is based on the kernel averages $\widehat{\widehat{\phi}}_{ij,T}(u,h)$ that are different from $\widehat{\phi}_{ij,T}(u,h)$ only by the fact that they do not include the covariates $\mathbf{X}_{it}$.

$$\begin{aligned}
\widehat{\widehat{\Phi}}_{n,T} &= \max_{1 \le i < j \le n} \max_{(u,h) \in \mathcal{G}_T} \widehat{\widehat{\phi}}^0_{ij,T}(u,h), \\
\widehat{\widehat{\phi}}^0_{ij,T}(u,h) &= \left| \frac{\widehat{\widehat{\phi}}_{ij,T}(u,h)}{\{\widehat{\widehat{\sigma}}_i^2 + \widehat{\widehat{\sigma}}_j^2\}^{1/2}} \right| - \lambda(h), \qquad \text{(A.4)} \\
\widehat{\widehat{\phi}}_{ij,T}(u,h) &= \sum_{t=1}^T w_{t,T}(u,h) \big\{ (\varepsilon_{it} - \bar{\varepsilon}_i) - (\varepsilon_{jt} - \bar{\varepsilon}_j) \big\}.
\end{aligned}$$

We can view these kernel averages as constructed (under the null) based on the unobserved variables $\widehat{\widehat{Y}}_{it}$ that are defined by

$$\begin{aligned}
\widehat{\widehat{Y}}_{it} &:= Y_{it} - \boldsymbol{\beta}_i^\top \mathbf{X}_{it} - \frac{1}{T} \sum_{t=1}^T \big(Y_{it} - \boldsymbol{\beta}_i^\top \mathbf{X}_{it} \big) = \\
&= m_i \Big(\frac{t}{T} \Big) - \frac{1}{T} \sum_{t=1}^T m_i \Big(\frac{t}{T} \Big) + \varepsilon_{it} - \frac{1}{T} \sum_{t=1}^T \varepsilon_{it}.
\end{aligned}$$

The definition of $\widehat{\widehat{\phi}}^0_{ij,T}(u,h)$ also includes the auxiliary estimator $\widehat{\widehat{\sigma}}_i^2$ of the long-run error variance σ_i^2 which is computed from the augmented sample $\{\widehat{\widehat{Y}}_{it} : 1 \le t \le T\}$. We thus regard $\widehat{\widehat{\sigma}}_i^2 = \widehat{\widehat{\sigma}}_i^2(\widehat{\widehat{Y}}_{i1}, \ldots, \widehat{\widehat{Y}}_{iT})$ as a function of the variables $\widehat{\widehat{Y}}_{it}$ for $1 \le t \le T$. As with $\widehat{\sigma}_i^2$, we assume that $\widehat{\widehat{\sigma}}_i^2 = \sigma_i^2 + o_p(\rho_T)$ with $\rho_T = o(\sqrt{h_{\min}} / \log T)$.

- Auxiliary statistics that has the same distribution as $\widehat{\widehat{\Phi}}_{n,T}$ for each $T = 1, 2, \ldots$. We prove that there exists such statistics with the property that they are close to the Gaussian test statistics $\Phi_{n,T}$ in Proposition A.3 using the strong approximation theory by Berkes et al. (2014).

$$
\begin{aligned}
\widetilde{\Phi}_{n,T} &= \max_{1\le i<j\le n} \max_{(u,h)\in\mathcal{G}_T} \widetilde{\phi}^0_{ij,T}(u,h), \\
\widetilde{\phi}^0_{ij,T}(u,h) &= \left|\frac{\widetilde{\phi}_{ij,T}(u,h)}{\{\widetilde{\sigma}_i^2+\widetilde{\sigma}_j^2\}^{1/2}}\right| - \lambda(h), \\
\widetilde{\phi}_{ij,T}(u,h) &= \sum\nolimits_{t=1}^T w_{t,T}(u,h)\big\{(\widetilde{\varepsilon}_{it}-\bar{\widetilde{\varepsilon}}_i)-(\widetilde{\varepsilon}_{jt}-\bar{\widetilde{\varepsilon}}_j)\big\}
\end{aligned}
\tag{A.5}
$$

with $[\widetilde{\varepsilon}_{i1},\dots,\widetilde{\varepsilon}_{iT}] \overset{\mathcal{D}}{=} [\varepsilon_{i1},\dots,\varepsilon_{iT}]$ for each i and T.

A.2 Auxiliary results

Here, we state some auxiliary results that will be used further in the proof of Theorem 4.1.

Definition A.1. *For a given $q>0$ and $\alpha>0$, we define dependence adjusted norm as $||X_\cdot||^q_{q,\alpha} = \sup_{m\ge 0}(m+1)^\alpha \sum_{t=m}^\infty \delta_q(X,t)$.*

Theorem A.1. *Wu and Wu (2016) Assume that $||X_\cdot||^q_{q,\alpha} < \infty$, 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 $x>0$,*

$$
\mathbb{P}(|S_T|\ge x) \le C_1\frac{|a|_q^q ||X_\cdot||^q_{q,\alpha}}{x^q} + C_2\exp\left(-\frac{C_3x^2}{T||X_\cdot||^2_{2,\alpha}}\right),
$$

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

Theorem A.2. *Wu (2007) Let $(\xi_i)_{i\in\mathbb{Z}}$ be a stationary and ergodic Markov chain and $g(\cdot)$ be a measurable function. Let $g(\xi_1)\in\mathcal{L}^q, q>2, \mathbb{E}[g(\xi_0)]=0$ and l be a positive, nondecreasing slowly varying function. Assume that*

$$
\sum_{i=n}^\infty \Big|\Big|\mathbb{E}[g(\xi_i)|\xi_0] - \mathbb{E}[g(\xi_i)|\xi_{-1}]\Big|\Big|_q = O\Big([\log n]^{-\beta}\Big),
$$

where $0\le\beta<1/q$ and

$$
\sum_{k=1}^\infty \frac{k^{-\beta q}}{[l(2^k)]^q} < \infty.
$$

Then $S_n = g(\xi_1)+\ldots+g(\xi_n) = o_{a.s.}\big[\sqrt{n}l(n)\big]$.

Proposition A.1. *Wu (2007) Let $(\epsilon_n)_{n\in\mathbb{Z}}$ be i.i.d. random variables, $\xi_n = (\ldots,\epsilon_{n-1},\epsilon_n)$ and $g(\cdot)$ be a measurable function such that $g(\xi_n)$ is a proper random variable for each $n\ge 0$. For $k\ge 0$ let $\tilde{\xi}_k = (\ldots,\epsilon_{-1},\epsilon_0',\epsilon_1,\ldots,\epsilon_{k-1},\epsilon_k)$, where ϵ_0' is an i.i.d. copy of ϵ_0. Let $g(\xi_0)\in\mathcal{L}^q, q>1$ and $\mathbb{E}[g(\xi_0)]=0$. For $n\ge 1$ we have*

$$
\Big|\Big|\mathbb{E}[g(\xi_n)|\xi_0] - \mathbb{E}[g(\xi_n)|\xi_{-1}]\Big|\Big|_q \le 2\Big|\Big|g(\xi_n)-g(\tilde{\xi}_n)\Big|\Big|_q.
$$

Proposition A.2. *Under the conditions of Theorem 4.1, for all $i \in \{1, \dots, n\}$ it holds that*

$$\bar{\mathbf{X}}_i = \frac{1}{T} \sum_{t=1}^{T} \mathbf{H}_i(\mathcal{U}_{it}) = o_P(1). \tag{A.6}$$

Proof of Proposition A.2. Take any $i \in \{1, \dots, n\}$. To prove (A.6), we will use two results from Wu (2007) stated above. First, fix $j \in \{1, \dots, d\}$. Denote $\xi_t = \mathcal{U}_{it}$, $\tilde{\xi}_t = \mathcal{U}'_{it}$ and $g(\cdot) = H_{i,j}(\cdot)$. Then by Assumption (C6), $g(\xi_0) = H_{i,j}(\mathcal{U}_{i0}) \in \mathcal{L}^{q'}$ for $q' > 4$ and $\mathbb{E}[g(\xi_0)] = \mathbb{E}[H_{i,j}(\mathcal{U}_{i0})] = 0$ and we can apply Proposition A.1 (Proposition 3(ii) in Wu (2007)) that says that for all $s \geq 1$ we have:

$$\big\| \mathbb{E}[g(\xi_s)|\xi_0] - \mathbb{E}[g(\xi_s)|\xi_{-1}] \big\|_{q'} \leq 2 \big\| g(\xi_s) - g(\tilde{\xi}_s) \big\|_{q'},$$

or, equivalently,

$$\big\| \mathbb{E}[H_{i,j}(\mathcal{U}_{is})|\mathcal{U}_{i0}] - \mathbb{E}[H_{i,j}(\mathcal{U}_{is})|\mathcal{U}_{i(-1)}] \big\|_{q'} \leq 2 \big\| H_{i,j}(\mathcal{U}_{is}) - H_{i,j}(\mathcal{U}'_{is}) \big\|_{q'}.$$

Since this holds simultaneously for all $j \in \{1, \dots, d\}$, we can use the obvious bound $\big\| H_{i,j}(\mathcal{U}_{is}) - H_{i,j}(\mathcal{U}'_{is}) \big\|_{q'} \leq \big\| \mathbf{H}_i(\mathcal{U}_{is}) - \mathbf{H}_i(\mathcal{U}'_{is}) \big\|_{q'} = \delta_{q'}(\mathbf{H}_i, s)$ and Assumption (C8) to write

$$0 \leq \sum_{s=t}^{\infty} \big\| \mathbb{E}[g(\xi_s)|\xi_0] - \mathbb{E}[g(\xi_s)|\xi_{-1}] \big\|_{q'} \leq \sum_{s=t}^{\infty} \delta_{q'}(\mathbf{H}_i, s) = O(t^{-\alpha}),$$

where $\alpha > 1/2 - 1/q'$.

Now we want to apply Theorem A.2 (Corollary 2(i) in Wu (2007)). As a parameter β in the theorem we can take any value satisfying assumption $0 \leq \beta < 1/q'$ because for every $\beta \geq 0$ we have

$$\sum_{s=t}^{\infty} \big\| \mathbb{E}[g(\xi_s)|\xi_0] - \mathbb{E}[g(\xi_s)|\xi_{-1}] \big\|_{q'} \leq \sum_{s=t}^{\infty} \delta_{q'}(\mathbf{H}_i, s) = O(t^{-\alpha}) = O\big([\log t]^{-\beta}\big).$$

Furthermore, as a positive, nondecreasing slowly varying function l we can take $l(x) = \log^{2/q'-\beta}(x)$. Then,

$$\begin{aligned}
\sum_{k=1}^{\infty} \frac{k^{-\beta q'}}{[l(2^k)]^{q'}} &= \sum_{k=1}^{\infty} \frac{k^{-\beta q'}}{\big[\log^{2/q'-\beta}(2^k)\big]^{q'}} \\
&= \sum_{k=1}^{\infty} \frac{k^{-\beta q'}}{k^{2-\beta q'} (\log 2)^{2-\beta q'}} \\
&= \frac{1}{(\log 2)^{2-\beta q'}} \sum_{k=1}^{\infty} \frac{1}{k^2} \\
&< \infty.
\end{aligned}$$

Hence, $S_T = g(\xi_1) + \ldots + g(\xi_T) = o_{a.s.}[\sqrt{T} \log^{2/q'-\beta}(T)]$, or, equivalently, $\bar{X}_{i,j} = S_T/T = o_{a.s.}[\log^{2/q'-\beta}(T)/\sqrt{T}] = o_P(1)$ for each $j \in \{1, \dots, d\}$. Obviously, this means that $\bar{\mathbf{X}}_i = o_P(1)$. $\square$

A.3 Proof of Theorem 4.1

The main steps of the proof of the Theorem 4.1 are described below. We will build the proof on the auxiliary results stated in A.2.

1. First, we introduce the intermediate statistics $\widehat{\widehat{\Phi}}_{n,T}$ that can be regarded as the version of the statistics $\widehat{\Phi}_{n,T}$ where we excluded the regressors $\mathbf{X}_{it}$ from the construction of the kernel averages. We show that we can replace this multiscale statistic $\widehat{\widehat{\Phi}}_{n,T}$ by an identically distributed version $\widetilde{\Phi}_{n,T}$ which is close to the Gaussian statistics $\Phi_{n,T}$ defined in (3.8). Formally, in Proposition A.3 we prove that 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
$$\big|\widetilde{\Phi}_{n,T} - \Phi_{n,T}\big| = o_p\Big(\frac{T^{1/q}}{\sqrt{Th_{\min}}} + \rho_T\sqrt{\log T}\Big),$$
where $\Phi_{n,T}$ is the Gaussian statistic.

2. Second, in Proposition A.5 we demonstrate 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 as $T \to \infty$. Or, in other words, it holds that
$$\sup_{x\in\mathbb{R}} \mathbb{P}\big(|\Phi_{n,T} - x| \leq \delta_T\big) = o(1),$$
where $\delta_T = T^{1/q}/\sqrt{Th_{\min}} + \rho_T\sqrt{\log T}$.

3. Then, we make use of Lemma A.6 to show that
$$\sup_{x\in\mathbb{R}} \big|\mathbb{P}(\widehat{\widehat{\Phi}}_{n,T} \leq x) - \mathbb{P}(\Phi_{n,T} \leq x)\big| = o(1).$$
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 $n \geq 2$, $T \geq 1$.

4. In the fourth step, in Propositions A.7 and A.8 we formally show that the introduced intermediate statistic $\widehat{\widehat{\Phi}}_{n,T}$ is close to $\widehat{\Phi}_{n,T}$, i.e. there exists a sequence of positive numbers $\gamma_{n,T}$ that converges to 0 as $T \to \infty$ such that for all $x \in \mathbb{R}$
$$\begin{aligned}\mathbb{P}\Big(\widehat{\widehat{\Phi}}_{n,T} \leq x - \gamma_{n,T}\Big) - \mathbb{P}\Big(\big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| > \gamma_{n,T}\Big) &\leq \mathbb{P}(\widehat{\Phi}_{n,T} \leq x) \\ &\leq \mathbb{P}\Big(\widehat{\widehat{\Phi}}_{n,T} \leq x + \gamma_{n,T}\Big) + \mathbb{P}\Big(\big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| > \gamma_{n,T}\Big),\end{aligned}$$
and
$$\mathbb{P}\Big(\big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| > \gamma_{n,T}\Big) = o(1). \tag{A.7}$$
Note that (A.7) does not involve x. Hence, this result is uniform over all $x \in \mathbb{R}$.

5. And finally, by the means of Proposition A.9 we prove that

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

which immediately implies the statement of Theorem 4.1.

Step 1

The auxiliary statistics $\widehat{\Phi}_{n,T}$ defined in (4.1) is equal to our multiscale statistics $\widehat{\Psi}_{n,T}$ under the null hypothesis, but has the property that it depends on the known covariates $\mathbf{X}_{it}$, whereas the Gaussian version $\Phi_{n,T}$ is independent of them. This is the reason why we need to introduce additional intermediate test statistics that do not include the covariates and connect $\widehat{\Phi}_{n,T}$ and $\Phi_{n,T}$.
We do it in the following way. For each i and j, consider the kernel averages

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

We can view these kernel averages as constructed (under the null) based on the unobserved variables $\widehat{\widehat{Y}}_{it}$ and $\widehat{\widehat{Y}}_{jt}$ defined by

$$\begin{aligned}\widehat{\widehat{Y}}_{it} :&= Y_{it} - \boldsymbol{\beta}_i^\top \mathbf{X}_{it} - \frac{1}{T}\sum_{t=1}^{T}\big(Y_{it} - \boldsymbol{\beta}_i^\top \mathbf{X}_{it}\big) = \\ &= m_i\Big(\frac{t}{T}\Big) - \frac{1}{T}\sum_{t=1}^{T} m_i\Big(\frac{t}{T}\Big) + \varepsilon_{it} - \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}.\end{aligned}$$

The intermediate statistic is then defined as

$$\widehat{\widehat{\Phi}}_{n,T} = \max_{1\le i<j\le n}\max_{(u,h)\in\mathcal{G}_T} \widehat{\widehat{\phi}}^0{}_{ij,T}(u,h) \quad \text{with}$$

$$\widehat{\widehat{\phi}}^0{}_{ij,T}(u,h) = \left|\frac{\widehat{\widehat{\phi}}_{ij,T}(u,h)}{\big\{\widehat{\widehat{\sigma}}_i^2 + \widehat{\widehat{\sigma}}_j^2\big\}^{1/2}}\right| - \lambda(h)$$

with $\widehat{\widehat{\sigma}}_i^2$ being an estmator 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 unobserved sample $\{\widehat{\widehat{Y}}_{it} : 1 \le t \le T\}$. We thus regard $\widehat{\widehat{\sigma}}_i^2 = \widehat{\widehat{\sigma}}_i^2(\widehat{\widehat{Y}}_{i1}, \ldots, \widehat{\widehat{Y}}_{iT})$ as a function of the variables $\widehat{\widehat{Y}}_{it}$ for $1 \le t \le T$. As with the estimator $\widehat{\sigma}_i^2$, we assume that $\widehat{\widehat{\sigma}}_i^2 = \sigma_i^2 + o_p(\rho_T)$ with $\rho_T = o(\sqrt{h_{\min}}/\log T)$.
The statistics $\widehat{\widehat{\Phi}}_{n,T}$ can thus be viewed as a version of the statistic $\widehat{\Phi}_{n,T}$ without the covariates. However, we formally prove that these two statistics are close in Step 4. In this step, we are interested in another matter.
Specifically, the main purpose of this part of the proof is to show that there is a version of the multiscale statistics $\widehat{\widehat{\Phi}}_{n,T}$ which is close to the Gaussian statistics $\Phi_{n,T}$ (defined in (3.8)) which distribution is known. More specifically, we prove the following result.

Proposition A.3. *Under the conditions of Theorem 4.1, there exist statistics $\widetilde{\Phi}_{n,T}$ for $T = 1, 2, \ldots$ 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)*

$$\big|\widetilde{\Phi}_{n,T} - \Phi_{n,T}\big| = o_p\Big(\frac{T^{1/q}}{\sqrt{Th_{\min}}} + \rho_T\sqrt{\log T}\Big), \tag{A.8}$$

where $\Phi_{n,T}$ is a Gaussian statistic as defined in (3.8).

Proof of Proposition A.3. For the proof, we draw on strong approximation theory for each stationary process $\mathcal{E}_i = \{\varepsilon_{it} : 1 \leq t \leq T\}$ 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 $\mathbb{B}_i$ and a sequence $\{\widetilde{\varepsilon}_{it} : t \in \mathbb{N}\}$ such that $[\widetilde{\varepsilon}_{i1}, \ldots, \widetilde{\varepsilon}_{iT}] \stackrel{\mathcal{D}}{=} [\varepsilon_{i1}, \ldots, \varepsilon_{iT}]$ for each T and

$$\max_{1\leq t\leq T}\Big|\sum_{s=1}^{t}\widetilde{\varepsilon}_{is} - \sigma_i\mathbb{B}_i(t)\Big| = o\big(T^{1/q}\big) \quad \text{a.s.}, \tag{A.9}$$

where $\sigma_i^2 = \sum_{k\in\mathbb{Z}}\text{Cov}(\varepsilon_{i0}, \varepsilon_{ik})$ denotes the long-run error variance.
We apply this result for each stationary process $\mathcal{E}_i = \{\varepsilon_{it} : 1 \leq t \leq T\}$ so that each process $\widetilde{\mathcal{E}}_i = \{\widetilde{\varepsilon}_{it} : t \in \mathbb{N}\}$ is independent of $\widetilde{\mathcal{E}}_j = \{\widetilde{\varepsilon}_{jt} : t \in \mathbb{N}\}$ for $i \neq j$.
Furthermore, we define

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

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 $1 \leq t \leq T$. Since $[\widetilde{\varepsilon}_{i1}, \ldots, \widetilde{\varepsilon}_{iT}] \stackrel{\mathcal{D}}{=} [\varepsilon_{i1}, \ldots, \varepsilon_{iT}]$, we have $\sum_{\ell=-\infty}^{\infty}\text{Cov}(\widetilde{\varepsilon}_{i0}, \widetilde{\varepsilon}_{i\ell}) = \sum_{\ell=-\infty}^{\infty}\text{Cov}(\varepsilon_{i0}, \varepsilon_{i\ell}) = \sigma_i^2$. Hence, by construction $\widetilde{\sigma}_i^2 = \sigma_i^2 + o_P(\rho_T)$.
In addition, we let

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

with $\phi_{ij,T}(u,h)$ defined in (**??**) and $Z_{it} = \mathbb{B}_i(t) - \mathbb{B}_i(t-1)$. With this notation, we can write

$$\big|\widetilde{\Phi}_{n,T} - \Phi'_{n,T}\big| \leq \big|\widetilde{\Phi}_{n,T} - \Phi^\diamond_{n,T}\big| + \big|\Phi^\diamond_{n,T} - \Phi_{n,T}\big|. \tag{A.10}$$

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

$$\big|\widetilde{\Phi}_{n,T} - \Phi^\diamond_{n,T}\big| \leq \max_{1\leq i<j\leq n}\Big(\{\widetilde{\sigma}_i^2 + \widetilde{\sigma}_j^2\}^{-1/2}\max_{(u,h)\in\mathcal{G}_T}\big|\widetilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h)\big|\Big). \tag{A.11}$$

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

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

where

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

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

$$\big|\widetilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h)\big| = \Big| \sum_{t=1}^{T-1} A_{ij,t} \big(w_{t,T}(u,h) - w_{t+1,T}(u,h) \big) \Big| \le W_T(u,h) \max_{1 \le t \le T} |A_{ij,t}|. \tag{A.12}$$

Now consider $\max_{1 \le t \le T} |A_{ij,t}|$. Straightforward calculations yield the following bound:

$$\begin{aligned}
\max_{1 \le t \le T} |A_{ij,t}| \le& \max_{1 \le t \le T} \Big| \sum_{s=1}^{t} \widetilde{\varepsilon}_{is} - \sigma_i \sum_{s=1}^{t} Z_{is} \Big| + \max_{1 \le t \le T} \Big| t(\bar{\widetilde{\varepsilon}}_i - \sigma_i \bar{Z}_i) \Big| \\
&+ \max_{1 \le t \le T} \Big| \sum_{s=1}^{t} \widetilde{\varepsilon}_{js} - \sigma_j \sum_{s=1}^{t} Z_{js} \Big| + \max_{1 \le t \le T} \Big| t(\bar{\widetilde{\varepsilon}}_j - \sigma_j \bar{Z}_j) \Big| \\
\le& 2 \max_{1 \le t \le T} \Big| \sum_{s=1}^{t} \widetilde{\varepsilon}_{is} - \sigma_i \sum_{s=1}^{t} Z_{is} \Big| + 2 \max_{1 \le t \le T} \Big| \sum_{s=1}^{t} \widetilde{\varepsilon}_{js} - \sigma_j \sum_{s=1}^{t} Z_{js} \Big| \\
=& 2 \max_{1 \le t \le T} \Big| \sum_{s=1}^{t} \widetilde{\varepsilon}_{is} - \sigma_i \sum_{s=1}^{t} \big(\mathbb{B}_i(s) - \mathbb{B}_i(s-1) \big) \Big| \\
&+ 2 \max_{1 \le t \le T} \Big| \sum_{s=1}^{t} \widetilde{\varepsilon}_{js} - \sigma_j \sum_{s=1}^{t} \big(\mathbb{B}_j(s) - \mathbb{B}_j(s-1) \big) \Big| \\
=& 2 \max_{1 \le t \le T} \Big| \sum_{s=1}^{t} \widetilde{\varepsilon}_{is} - \sigma_i \mathbb{B}_i(t) \Big| + 2 \max_{1 \le t \le T} \Big| \sum_{s=1}^{t} \widetilde{\varepsilon}_{js} - \sigma_j \mathbb{B}_j(t) \Big|.
\end{aligned}$$

Applying the strong approximation result (A.9), we can infer that

$$\max_{1 \le t \le T} |A_{ij,t}| = o_P\big(T^{1/q}\big). \tag{A.13}$$

Standard arguments show that $\max_{(u,h)\in\mathcal{G}_T} W_T(u,h) = O(1/\sqrt{Th_{\min}})$. Plugging (A.13) in (A.12) and then in (A.11), we can thus infer that

$$\begin{aligned} \big|\widetilde{\Phi}_{n,T} - \Phi^{\diamond}_{n,T}\big| &\le \{\widetilde{\sigma}_i^2 + \widetilde{\sigma}_j^2\}^{-1/2} \max_{(u,h)\in\mathcal{G}_T} W_T(u,h) \max_{1\le i<j\le n} \max_{1\le t\le T} |A_{ij,t}| \\ &= o_P\Big(\frac{T^{1/q}}{\sqrt{Th_{\min}}}\Big). \end{aligned} \tag{A.14}$$

Now consider $|\Phi^{\diamond}_{n,T} - \Phi_{n,T}|$. Since $\phi_{ij,T}(u,h)$ is distributed as $N(0,\sigma_i^2+\sigma_j^2)$ 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 $\widetilde{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$, we can establish that

$$\big|\Phi^{\diamond}_{n,T} - \Phi_{n,T}\big| \le \max_{1\le i<j\le n} \max_{(u,h)\in\mathcal{G}_T} \Big|\frac{\phi_{ij,T}(u,h)}{\{\widetilde{\sigma}_i^2+\widetilde{\sigma}_j^2\}^{1/2}} - \frac{\phi_{ij,T}(u,h)}{\{\sigma_i^2+\sigma_j^2\}^{1/2}}\Big| = o_P(\rho_T\sqrt{\log T}). \tag{A.15}$$

Plugging (A.14) and (A.15) in (A.10) completes the proof. □

Step 2

In this section, we establish some properties of the Gaussian statistic $\Phi_{n,T}$ defined in (3.8). 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.5 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.4. *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.4 is provided in **?**.

Proposition A.5. *Under the conditions of Theorem **??**, it holds that*

$$\sup_{x\in\mathbb{R}} \mathbb{P}\big(|\Phi_{n,T} - x| \le \delta_T\big) = o(1), \tag{A.16}$$

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

Proof of Proposition A.5. We write $x = (u, h)$ along with $\mathcal{G}_T = \{x : x \in \mathcal{G}_T\} = \{x_1, \ldots, x_p\}$, where $p := |\mathcal{G}_T| \leq O(T^\theta)$ for some large but fixed $\theta > 0$ by our assumptions. Moreover, for $k = 1, \ldots, p$, we set

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

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

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

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

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

and (ii) $\sigma_l^2 := \text{Var}(U'_l) = 1$ for all $1 \leq l \leq (n-1)np$. 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 $a_{(n-1)np} = b_{(n-1)np} \leq C\sqrt{\log((n-1)np)} \leq C\sqrt{\log T}$. With this notation at hand, we can apply Proposition A.4 to obtain that

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

with $\delta_T = T^{1/q}/\sqrt{Th_{\min}} + \rho_T\sqrt{\log T}$, which is the statement of Proposition A.5. $\square$

Step 3

Lemma A.6. *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*

$$\sup_{x \in \mathbb{R}} \mathbb{P}(|V_T - x| \leq \delta_T) = o(1), \tag{A.17}$$

then

$$\sup_{x \in \mathbb{R}} \big|\mathbb{P}(V_T \leq x) - \mathbb{P}(W_T \leq x)\big| = o(1). \tag{A.18}$$

Proof of this lemma is provided in Khismatullina and Vogt (2020).
Applying Lemma A.6 to $\widetilde{\Phi}_{n,T}$ and $\Phi_{n,T}$ together with the results (A.8) and (A.16) and noting the fact that $\widetilde{\Phi}_{n,T}$ is distributed as $\widehat{\widehat{\Phi}}_{n,T}$ for any $n \geq 2$, $T \geq 1$ leads us to

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

Step 4

As was already mentioned in Step 1, the statistics $\widehat{\widehat{\Phi}}_{n,T}$ can be viewed as an approximation of the statistics $\widehat{\Phi}_{n,T}$. 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 $\widehat{\boldsymbol{\beta}}_i$, $\widehat{\widehat{\sigma}}_i^2$ and assumptions on $\mathbf{X}_{it}$. In the following two propositions we prove it formally.

Proposition A.7. *For any $x \in \mathbb{R}$ and any $\gamma > 0$, we have*

$$\begin{aligned} \mathbb{P}\Big(\widehat{\widehat{\Phi}}_{n,T} \le x - \gamma\Big) - \mathbb{P}\Big(\big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| > \gamma\Big) &\le \mathbb{P}(\widehat{\Phi}_{n,T} \le x) \\ &\le \mathbb{P}\Big(\widehat{\widehat{\Phi}}_{n,T} \le x + \gamma\Big) + \mathbb{P}\Big(\big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| > \gamma\Big). \end{aligned} \tag{A.19}$$

Proof of Proposition A.7. From the law of total probability and the monotonic property of the probability function, we have

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

Analogously,

$$\mathbb{P}(\widehat{\widehat{\Phi}}_{n,T} \le x - \gamma) \le \mathbb{P}\Big(\widehat{\Phi}_{n,T} \le x\Big) + \mathbb{P}\Big(\big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| > \gamma\Big).$$

Combining these two inequalities together, we arrive at the desired result. □

The aim of the next proposition is to determine the sequence of values of $\gamma_{n,T}$ that may depend on n and 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.8. *There exists a sequence of positive random numbers $\{\gamma_{n,T}\}_T$, that converges to 0 as $T \to \infty$, such that*

$$\mathbb{P}\Big(\big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| > \gamma_{n,T}\Big) = o(1). \tag{A.20}$$

Proof of Proposition A.8. Straightforward calculations yield that

$$\begin{aligned} \big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| \le \max_{1 \le i < j \le n} \max_{(u,h) \in \mathcal{G}_T} &\left| \frac{\widehat{\widehat{\phi}}_{ij,T}(u,h)}{\big(\widehat{\widehat{\sigma}}_i^2 + \widehat{\widehat{\sigma}}_j^2\big)^{1/2}} - \frac{\widehat{\widehat{\phi}}_{ij,T}(u,h)}{\big(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\big)^{1/2}} \right| \\ + \max_{1 \le i < j \le n} \max_{(u,h) \in \mathcal{G}_T} &\left| \frac{\widehat{\widehat{\phi}}_{ij,T}(u,h)}{\big(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\big)^{1/2}} - \frac{\widehat{\phi}_{ij,T}(u,h)}{\big(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\big)^{1/2}} \right|. \end{aligned}$$

Obviously,

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

and

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

Furthermore, the difference of the kernel averages $\widehat{\widehat{\phi}}_{ij,T}(u,h)-\widehat{\phi}_{ij,T}(u,h)$ does not include the error term (it cancels out) and can be written as

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

Hence,

$$\begin{aligned}\big|\widehat{\widehat{\Phi}}_{n,T}-\widehat{\Phi}_{n,T}\big| &\le \max_{1\le i<j\le n}\left|\big(\widehat{\widehat{\sigma}}_i^2+\widehat{\widehat{\sigma}}_j^2\big)^{-1/2}-\big(\widehat{\sigma}_i^2+\widehat{\sigma}_j^2\big)^{-1/2}\right|\max_{1\le i<j\le n}\max_{(u,h)\in\mathcal{G}_T}\left|\widehat{\widehat{\phi}}_{ij,T}(u,h)\right|\\
&\quad+2\max_{1\le i<j\le n}\big(\widehat{\sigma}_i^2+\widehat{\sigma}_j^2\big)^{-1/2}\max_{1\le i\le n}\max_{(u,h)\in\mathcal{G}_T}\Big|(\boldsymbol{\beta}_i-\widehat{\boldsymbol{\beta}}_i)^\top\sum_{t=1}^T w_{t,T}(u,h)\mathbf{X}_{it}\Big|\\
&\quad+2\max_{1\le i<j\le n}\big(\widehat{\sigma}_i^2+\widehat{\sigma}_j^2\big)^{-1/2}\max_{1\le i\le n}\big|(\boldsymbol{\beta}_i-\widehat{\boldsymbol{\beta}}_i)^\top\bar{\mathbf{X}}_i\big|\max_{(u,h)\in\mathcal{G}_T}\Big|\sum_{t=1}^T w_{t,T}(u,h)\Big|.\end{aligned}\tag{A.21}$$

We consider each of the three summands separately.
We start by looking at the first summand in (A.21). Since $\widehat{\widehat{\sigma}}_i^2=\sigma_i^2+o_P(\rho_T)$ and $\widehat{\sigma}_i^2=\sigma_i^2+o_P(\rho_T)$ by our assumptions, we have that

$$\max_{1\le i<j\le n}\left|\big(\widehat{\widehat{\sigma}}_i^2+\widehat{\widehat{\sigma}}_j^2\big)^{-1/2}-\big(\widehat{\sigma}_i^2+\widehat{\sigma}_j^2\big)^{-1/2}\right|=o_P(\rho_T).\tag{A.22}$$

Then, we investigate $\max_{(u,h)\in\mathcal{G}_T}\big|\widehat{\widehat{\phi}}_{ij,T}(u,h)\big|$. Specifically, we are interested in its distribution. We know by Proposition A.3 that there exists $\widetilde{\phi}_{ij,T}(u,h)$ that has the same distribution as $\widehat{\widehat{\phi}}_{ij,T}(u,h)$ for all $1 \le i < j \le n$ and all $(u,h) \in \mathcal{G}_T$.

$$\mathbb{P}\left(\max_{(u,h)\in\mathcal{G}_T}\left|\widehat{\widehat{\phi}}_{ij,T}(u,h)\right| \le C\right) = \mathbb{P}\left(\max_{(u,h)\in\mathcal{G}_T}\left|\widetilde{\phi}_{ij,T}(u,h)\right| \le C\right).$$

So instead of looking at the distribution of $\max_{(u,h)\in\mathcal{G}_T}\left|\widehat{\widehat{\phi}}_{ij,T}(u,h)\right|$, we now turn our attention at the distribution of $\max_{(u,h)\in\mathcal{G}_T}\left|\widetilde{\phi}_{ij,T}(u,h)\right|$ instead.
In bounding this probability, we can use the strategy from the second part of the proof of Proposition A.7. For any $c_T \in \mathbb{R}$ (taking $x = \gamma = c_T/2$) we have

$$\begin{aligned}
&\mathbb{P}\left(\max_{(u,h)\in\mathcal{G}_T}|\phi_{ij,T}(u,h)| \le c_T/2\right)\\
&\le \mathbb{P}\left(\max_{(u,h)\in\mathcal{G}_T}\left|\widetilde{\phi}_{ij,T}(u,h)\right| \le c_T\right) + \mathbb{P}\left(\left|\max_{(u,h)\in\mathcal{G}_T}\left|\widetilde{\phi}_{ij,T}(u,h)\right| - \max_{(u,h)\in\mathcal{G}_T}|\phi_{ij,T}(u,h)|\right| > \frac{c_T}{2}\right)\\
&\le \mathbb{P}\left(\max_{(u,h)\in\mathcal{G}_T}\left|\widetilde{\phi}_{ij,T}(u,h)\right| \le c_T\right) + \mathbb{P}\left(\max_{(u,h)\in\mathcal{G}_T}\left|\widetilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h)\right| > \frac{c_T}{2}\right).
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathbb{P}&\left(\max_{(u,h)\in\mathcal{G}_T}\left|\widehat{\widehat{\phi}}_{ij,T}(u,h)\right| \le c_T\right) = \mathbb{P}\left(\max_{(u,h)\in\mathcal{G}_T}\left|\widetilde{\phi}_{ij,T}(u,h)\right| \le c_T\right)\\
&\ge \mathbb{P}\left(\max_{(u,h)\in\mathcal{G}_T}|\phi_{ij,T}(u,h)| \le c_T/2\right) - \mathbb{P}\left(\max_{(u,h)\in\mathcal{G}_T}\left|\widetilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h)\right| > \frac{c_T}{2}\right).
\end{aligned} \tag{A.23}$$

By (A.15) we have

$$\max_{(u,h)\in\mathcal{G}_T}\left|\widetilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h)\right| = o_P\left(\frac{T^{1/q}}{\sqrt{Th_{\min}}}\right).$$

Furthermore, $\phi_{ij,T}(u,h)$ is distributed as $N(0, \sigma_i^2 + \sigma_j^2)$ for all $(u,h) \in \mathcal{G}_T$ and all $1 \le i < j \le n$, and $|\mathcal{G}_T| = O(T^\theta)$ for some large but fixed constant θ by Assumption (C12). By the standard results from the probability theory, we know that

$$\max_{(u,h)\in\mathcal{G}_T}|\phi_{ij,T}(u,h)| = O_P(\sqrt{\log T}).$$

Since $T^{1/q}/\sqrt{Th_{\min}} \ll \sqrt{\log T}$, we can take $c_T = o(\sqrt{\log T})$ in (A.23) to get the following:

$$\begin{aligned}
&\mathbb{P}\left(\max_{(u,h)\in\mathcal{G}_T}\left|\widehat{\widehat{\phi}}_{ij,T}(u,h)\right| \le c_T\right)\\
&\ge \mathbb{P}\left(\max_{(u,h)\in\mathcal{G}_T}|\phi_{ij,T}(u,h)| \le \frac{c_T}{2}\right) - \mathbb{P}\left(\max_{(u,h)\in\mathcal{G}_T}\left|\widetilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h)\right| > \frac{c_T}{2}\right)\\
&= 1 - o(1) - o(1)\\
&= 1 - o(1),
\end{aligned}$$

which means that

$$\max_{(u,h)\in\mathcal{G}_T}\left|\widehat{\widehat{\phi}}_{ij,T}(u,h)\right| = o_P(\sqrt{\log T}). \tag{A.24}$$

Combining (A.22) and (A.24) and taking into consideration that n is fixed, we get the following:

$$\begin{aligned}\max_{1\le i<j\le n}\left|\big(\widehat{\widehat{\sigma}}_i^2+\widehat{\widehat{\sigma}}_j^2\big)^{-1/2}-\big(\widehat{\sigma}_i^2+\widehat{\sigma}_j^2\big)^{-1/2}\right| & \max_{1\le i<j\le n}\max_{(u,h)\in\mathcal{G}_T}\big|\widehat{\widehat{\phi}}_{ij,T}(u,h)\big| \\ &= o_P(\rho_T)\cdot o_P(\sqrt{\log T}) \\ &= o_P(1)\end{aligned} \tag{A.25}$$

since by our assumption $\rho_T = O(\sqrt{h_{\min}}/\log T)$.
Now we evaluate the second summand in (A.21).
First, by our assumptions $\widehat{\sigma}_i^2 = \sigma_i^2 + o_P(\rho_T)$. Moreover, for all $i \in \{1,\dots,n\}$ we know that $\sigma_i^2 \neq 0$. Hence,

$$\max_{1\le i<j\le n}\big(\widehat{\sigma}_i^2+\widehat{\sigma}_j^2\big)^{-1/2} = O_P(1). \tag{A.26}$$

Then, by Theorem 5.1, we know that

$$\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i = O_P\Big(\frac{1}{\sqrt{T}}\Big). \tag{A.27}$$

Now consider $\sum_{t=1}^T w_{t,T}(u,h)\mathbf{X}_{it}$. 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

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

Note that

$$\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} w_{t,T}^2(u,h) = \sum_{t=1}^T w_{t,T}^2(u,h) = \sum_{t=1}^T \frac{\Lambda_{t,T}^2(u,h)}{\sum_{s=1}^T \Lambda_{s,T}^2(u,h)} = 1. \tag{A.28}$$

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 $2Th \le D_{T,u,h} \le 2Th+2$) and using (A.28), we can normalize 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(ii) in Wu and Wu (2016)), if we define the weights from the theorem as $a_t = \sqrt{D_{T,u,h}} \cdot w_{t,T}(u,h)$, we can bound the probability as follows:

$$\begin{aligned}
&\mathbb{P}\left(\Big|\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} \sqrt{D_{T,u,h}} \cdot w_{t,T}(u,h) X_{it}\Big| \geq x\right) \\
&\quad \leq C_1 \frac{\big(\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} |\sqrt{D_{T,u,h}} \cdot w_{t,T}(u,h)|^{q'}\big)||X_{i\cdot}||_{q',\alpha}^{q'}}{x^{q'}} + C_2 \exp\left(-\frac{C_3 x^2}{D_{T,u,h}||X_{i\cdot}||_{2,\alpha}^2}\right),
\end{aligned} \tag{A.29}$$

where $||X_{i\cdot}||_{q,\alpha}^q$ is the dependence adjusted norm as defined by Definition A.1. Taking any $\delta > 0$ and applying Boole's inequality and (A.29) subsequently, we get

$$\begin{aligned}
&\mathbb{P}\left(\max_{(u,h)\in\mathcal{G}_T}\Big|\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} w_{t,T}(u,h) X_{it}\Big| \geq \delta\sqrt{T}\right) \\
&\quad \leq \sum_{(u,h)\in\mathcal{G}_T} \mathbb{P}\left(\Big|\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} w_{t,T}(u,h) X_{it}\Big| \geq \delta\sqrt{T}\right) \\
&\quad = \sum_{(u,h)\in\mathcal{G}_T} \mathbb{P}\left(\Big|\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} \sqrt{D_{T,u,h}} \cdot w_{t,T}(u,h) X_{it}\Big| \geq \delta\sqrt{D_{T,u,h}T}\right) \\
&\quad \leq \sum_{(u,h)\in\mathcal{G}_T} \left[C_1 \frac{(\sqrt{D_{T,u,h}})^{q'}\big(\sum |w_{t,T}(u,h)|^{q'}\big)||X_{i\cdot}||_{q',\alpha}^{q'}}{\big(\delta\sqrt{D_{T,u,h}T}\big)^{q'}} + C_2 \exp\left(-\frac{C_3\big(\delta\sqrt{D_{T,u,h}T}\big)^2}{D_{T,u,h}||X_{i\cdot}||_{2,\alpha}^2}\right)\right] \\
&\quad = \sum_{(u,h)\in\mathcal{G}_T} \left[C_1 \frac{\big(\sum |w_{t,T}(u,h)|^{q'}\big)||X_{i\cdot}||_{q',\alpha}^{q'}}{\big(\delta\sqrt{T}\big)^{q'}} + C_2 \exp\left(-\frac{C_3\delta^2 T}{||X_{i\cdot}||_{2,\alpha}^2}\right)\right] \\
&\quad \leq C_1 \frac{T^\theta ||X_{i\cdot}||_{q',\alpha}^{q'}}{T^{q'/2}\cdot\delta^{q'}} \max_{(u,h)\in\mathcal{G}_T}\left(\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} |w_{t,T}(u,h)|^{q'}\right) + C_2 T^\theta \exp\left(-\frac{C_3\delta^2 T}{||X_{i\cdot}||_{2,\alpha}^2}\right) \\
&\quad = C\frac{T^{\theta-q'/2}}{\delta^{q'}} + CT^\theta \exp\left(-CT\delta^2\right).
\end{aligned}$$

where the symbol C denotes a universal real constant that does not depend neither on T, nor on δ, and takes a different value on each occurrence. Here in the last equality we used the following facts:

1. $||X_{i\cdot}||_{q',\alpha}^{q'} = \sup_{t\geq 0}(t+1)^\alpha \sum_{s=t}^\infty \delta_{q'}(H_i,s) < \infty$ holds true since $\sum_{s=t}^\infty \delta_{q'}(H_i,s) = O(t^{-\alpha})$ by Assumption (C8);

2. $\max_{(u,h)\in\mathcal{G}_T}\left(\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} |w_{t,T}(u,h)|^{q'}\right) \leq 1$ because for every $x\in[0,1]$ we have $0 \leq x^{q'/2} \leq x \leq 1$. Thus, since $\sum_{t=1}^T w_{t,T}^2(u,h) = 1$ by (A.28) we have $0 \leq w_{t,T}^2(u,h) \leq 1$ for all $t\in\{1,\dots,T\}$ and all $(u,h)\in\mathcal{G}_T$, we get

$$0 \leq |w_{t,T}(u,h)|^{q'} = (w_{t,T}^2(u,h))^{q'/2} \leq w_{t,T}^2(u,h) \leq 1.$$

This leads to a bound

$$\max_{(u,h)\in\mathcal{G}_T}\left(\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil}|w_{t,T}(u,h)|^{q'}\right)\leq\max_{(u,h)\in\mathcal{G}_T}\left(\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil}w_{t,T}^2(u,h)\right)=1.$$

3. $||X_{i\cdot}||_{2,\alpha}^2<\infty$ (follows from 1).

By Assumption (C6), $\theta-q'/2<0$ and the term on the RHS of the above inequality is converging to 0 as $T\to\infty$ for any fixed $\delta>0$. Hence,

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

and similarly,

$$\max_{(u,h)\in\mathcal{G}_T}\left|\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil}w_{t,T}(u,h)\mathbf{X}_{it}\right|=o_P(\sqrt{T}).\tag{A.30}$$

Combining (A.26), (A.27) and (A.30), we get the following:

$$\begin{aligned}\max_{1\leq i<j\leq n}\left(\widehat{\sigma}_i^2+\widehat{\sigma}_j^2\right)^{-1/2}\max_{1\leq i\leq n}\max_{(u,h)\in\mathcal{G}_T}&\left|(\boldsymbol{\beta}_i-\widehat{\boldsymbol{\beta}}_i)^\top\sum_{t=1}^{T}w_{t,T}(u,h)\mathbf{X}_{it}\right|\\&=O_P(1)\cdot O_P(1/\sqrt{T})\cdot o_P(\sqrt{T})\\&=o_P(1).\end{aligned}\tag{A.31}$$

Now consider the third summand in (A.21). Similarly as before,

$$\max_{1\leq i<j\leq n}\left(\widehat{\sigma}_i^2+\widehat{\sigma}_j^2\right)^{-1/2}=O_P(1)\tag{A.32}$$

and

$$\boldsymbol{\beta}_i-\widehat{\boldsymbol{\beta}}_i=O_P\Big(\frac{1}{\sqrt{T}}\Big).\tag{A.33}$$

Furthermore, by Proposition A.2,

$$\bar{\mathbf{X}}_i=o_P(1).\tag{A.34}$$

Finally, consider the local linear kernel weights $w_{t,T}(u,h)$ defined in (3.4). Again, 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 $\left|\sum_{t=1}^{T}w_{t,T}(u,h)\right|$ for all

$(u,h) \in \mathcal{G}_T$ using the Cauchy-Schwarz inequality:

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

Hence,

$$\max_{(u,h)\in\mathcal{G}_T}\Big|\sum_{t=1}^{T} w_{t,T}(u,h)\Big| = O(\sqrt{T}). \tag{A.35}$$

Combining (A.32), (A.33), (A.34) and (A.35), we get the following:

$$\begin{aligned}
\max_{1\leq i<j\leq n}\big(\widehat{\sigma}_i^2+\widehat{\sigma}_j^2\big)^{-1/2}\max_{1\leq i\leq n}\big|(\boldsymbol{\beta}_i-\widehat{\boldsymbol{\beta}}_i)^\top\bar{\mathbf{X}}_i\big|\max_{(u,h)\in\mathcal{G}_T}\Big|\sum_{t=1}^{T} w_{t,T}(u,h)\Big| \\
= O_P(1)\cdot O_P(1/\sqrt{T})\cdot o_P(1)\cdot O(\sqrt{T}) \\
= o_P(1).
\end{aligned} \tag{A.36}$$

Plugging (A.25), (A.31) and (A.36) in (A.21), we get that $\big|\widehat{\widehat{\Phi}}_{n,T}-\widehat{\Phi}_{n,T}\big| = o_P(1)$ and the statement of the theorem follows. □

Step 5

Proposition A.9. *Under the conditions of Theorem 4.1, it holds that*

$$\sup_{x\in\mathbb{R}}\big|\mathbb{P}(\widehat{\Phi}_{n,T}\leq x)-\mathbb{P}(\Phi_{n,T}\leq x)\big| = o(1). \tag{A.37}$$

Proof of Proposition A.9. First, we consider those $x \in \mathbb{R}$ such that $\mathbb{P}(\widehat{\Phi}_{n,T}\leq x)\geq\mathbb{P}(\Phi_{n,T}\leq x)$. Then by Proposition **??** for $\gamma_{n,T}>0$ from the Proposition

?? we have

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

Now consider such $x \in \mathbb{R}$ that $\mathbb{P}(\widehat{\Phi}_{n,T} \le x) < \mathbb{P}(\Phi_{n,T} \le x)$. Analogously,

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

Note that since $\gamma_{n,T} \to 0$, we can use the anticoncentration results (A.16) for the Gaussian statistic $\Phi_{n,T}$ to get $\sup_{x\in\mathbb{R}} \mathbb{P}\Big(\big|\Phi_{n,T} - x\big| \le \gamma_{n,T}\Big) = o(1)$. Moreover,

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

by Proposition A.8 and this probability does not depend on x.
Thus,

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

□

A.4 Proof of Theorem 5.1

Before proceeding to the proof of Theorem 5.1, we first prove several auxiliary results. In order to do that, 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,

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

We now can prove the following propositions.

Proposition A.10. *Under Assumptions (C4) and (C6),* $||\Delta\mathbf{H}_i(\mathcal{U}_{it})||_4 < \infty$.

Proof of Proposition A.10. By Assumption (C6) and the triangle inequality,

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

□

Proposition A.11. *Under Assumption (C9),* $\Delta\mathbf{X}_{it}$ *(elementwise) and* $\Delta\varepsilon_{it}$ *are uncorrelated for each* $t \in \{1, \ldots, T\}$.

Proof of Proposition A.11. By Assumption (C9),

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

□

Proposition A.12. *Define*

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

Under Assumptions (C2), (C3), (C6), (C7) and (C10), we have that $\sum_{s=1}^{\infty} \delta_2(\Delta\mathbf{U}_i, s) < \infty$.

Proof of Proposition A.12. By the triangle inequality and the definition of the physical dependence measure δ_2, we have that

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

Here $\mathcal{U}'_{it} = (\ldots, u_{i(-1)}, u'_{i0}, u_{i1}, \ldots, u_{it-1}, u_{it})$, $\mathcal{U}'_{i(t-1)} = (\ldots, u_{i(-1)}, u'_{i0}, u_{i1}, \ldots, u_{it-1})$, $\mathcal{J}'_{it} = (\ldots, \eta_{i(-1)}, \eta'_{i0}, \eta_{i1}, \ldots, \eta_{it-1}, \eta_{it})$, $\mathcal{J}'_{i(t-1)} = (\ldots, \eta_{i(-1)}, \eta'_{i0}, \eta_{i1}, \ldots, \eta_{it-1})$ 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}.

Therefore,

$$
\begin{aligned}
\sum_{s=1}^{\infty} \delta_2(\Delta \mathbf{U}_i, s) &\leq \sum_{s=0}^{\infty} \delta_2(\mathbf{U}_i, s) + \sum_{s=1}^{\infty} \delta_2(\mathbf{U}_i, s-1) \\
&+ \sum_{s=1}^{\infty} \big(\delta_2(\mathbf{H}_i, s-1) + \delta_2(\mathbf{H}_i, s)\big)||G_i|| + \sum_{s=1}^{\infty} \big(\delta_2(G_i, s-1) + \delta_2(G_i, s)\big)||\mathbf{H}_i||.
\end{aligned}
$$

By Assumptions (C2), (C3), (C6), (C7) and (C10), the RHS is finite. Statement of the proposition follows. □

Proposition A.13. *Under Assumptions (C1) - (C10),*

$$
\Big|\frac{1}{\sqrt{T}} \sum_{t=2}^{T} \Delta \mathbf{X}_{it} \Delta \varepsilon_{it}\Big| = O_P(1).
$$

Proof of Proposition A.13. For this proof, we will need the following notation:

$$\begin{aligned}
\mathcal{P}_{i,t}(\cdot) &:= \mathbb{E}[\cdot|\mathcal{I}_{it}] - \mathbb{E}[\cdot|\mathcal{I}_{i,t-1}], \\
\kappa_i &:= \frac{1}{T-1}\sum_{t=2}^{T} \Delta\mathbf{X}_{it}\Delta\varepsilon_{it}, \\
\kappa_{i,s}^{\mathcal{P}} &:= \frac{1}{T-1}\sum_{t=2}^{T} \mathcal{P}_{i,t-s}\big(\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}\big).
\end{aligned}$$

Since $\mathcal{P}_{i,t}(\Delta\mathbf{X}_{it}\Delta\varepsilon_{it})$, $2 \le t \le T$, are martingale differences, we have that

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

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 $\mathbb{E}\big(\Delta\mathbf{X}'_{it,s}\Delta\varepsilon'_{it,s}|\mathcal{I}_{i(t-s-1)}\big) = \mathbb{E}\big(\Delta\mathbf{X}'_{it,s}\Delta\varepsilon'_{it,s}|\mathcal{I}_{i(t-s)}\big)$. Furthermore, by linearity of the expectation and Jensen's inequality, we have

$$\begin{aligned}
||\kappa_{i,s}^{\mathcal{P}}||^2 &\le \frac{1}{(T-1)^2}\sum_{t=2}^{T} \Big|\Big|\mathbb{E}(\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}|\mathcal{I}_{i(t-s)}) - \mathbb{E}(\Delta\mathbf{X}'_{it,s}\Delta\varepsilon'_{it,s}|\mathcal{I}_{i(t-s)})\Big|\Big|^2 \\
&\le \frac{1}{(T-1)^2}\sum_{t=2}^{T} \Big|\Big|\Delta\mathbf{X}_{it}\Delta\varepsilon_{it} - \Delta\mathbf{X}'_{it,s}\Delta\varepsilon'_{it,s}\Big|\Big|^2 \\
&= \frac{1}{(T-1)^2}\sum_{t=2}^{T} \Big|\Big|\Delta\mathbf{H}_i(\mathcal{U}_{it})\Delta G_i(\mathcal{J}_{it}) - \Delta\mathbf{H}_i(\mathcal{U}'_{it,s})\Delta G_i(\mathcal{J}'_{it,s})\Big|\Big|^2 \\
&= \frac{1}{(T-1)^2}\sum_{t=2}^{T} \Big|\Big|\Delta\mathbf{U}_i(\mathcal{I}_{it}) - \Delta\mathbf{U}_i(\mathcal{I}'_{it,s})\Big|\Big|^2 \\
&\le \frac{1}{(T-1)^2}\sum_{t=2}^{T} \delta_2^2(\Delta\mathbf{U}_i, s) \\
&= \frac{1}{T-1}\delta_2^2(\Delta\mathbf{U}_i, s)
\end{aligned}$$

with $\mathcal{U}'_{it,s} = (\ldots, u_{i(t-s-1)}, u'_{i(t-s)}, u_{i(t-s+1)}, \ldots, u_{it})$, $u'_{i(t-s)}$ being an i.i.d. copy of $u_{i(t-s)}$, $\mathcal{J}'_{it,s} = (\ldots, \eta_{i(t-s-1)}, \eta'_{i(t-s)}, \eta_{i(t-s+1)}, \ldots, \eta_{it})$, $\eta'_{i(t-s)}$ being an i.i.d. copy of $\eta_{i(t-s)}$, and

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

$$\begin{aligned}
\kappa_i - \mathbb{E}\kappa_i &= \frac{1}{T-1}\sum_{t=2}^{T} \Delta\mathbf{X}_{it}\Delta\varepsilon_{it} - \mathbb{E}\kappa_i \\
&= \frac{1}{T-1}\sum_{t=2}^{T} \mathbb{E}(\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}|\mathcal{I}_{it}) - \mathbb{E}\kappa_i \\
&= \frac{1}{T-1}\sum_{t=2}^{T} \big(\mathbb{E}(\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}|\mathcal{I}_{it}) - \mathbb{E}(\mathbf{X}_{it}\Delta\varepsilon_{it})\big) \\
&= \frac{1}{T-1}\sum_{t=2}^{T}\sum_{s=0}^{\infty} \big(\mathbb{E}(\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}|\mathcal{I}_{i(t-s)}) - \mathbb{E}(\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}|\mathcal{I}_{i(t-s-1)})\big) \\
&= \frac{1}{T-1}\sum_{t=2}^{T}\sum_{s=0}^{\infty} \mathcal{P}_{i,t-s}(\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}) = \sum_{s=0}^{\infty} \kappa^{\mathcal{P}}_{i,s}.
\end{aligned}$$

Thus, by Proposition A.12,

$$||\kappa_i - \mathbb{E}\kappa_i|| \le \sum_{s=0}^{\infty} ||\kappa^{\mathcal{P}}_{i,s}|| \le \frac{1}{\sqrt{T-1}}\sum_{s=0}^{\infty}\delta_2(\Delta\mathbf{U}_i, s) = O\left(\frac{1}{\sqrt{T}}\right)$$

Since $\mathbb{E}\kappa_i = 0$ by Proposition A.11, we conclude that

$$\Big|\Big|\frac{1}{T}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}\Big|\Big| = O\left(\frac{1}{\sqrt{T}}\right).$$

Therefore, the proposition follows. □

Proof of Theorem 5.1. Before we begin, we need to introduce some additional notation that we will use throughout the proof. First, define $\Delta m_{it} = m_i\left(\frac{t}{T}\right) - m_i\left(\frac{t-1}{T}\right)$. Then, by Assumption (C4), we can rewrite the first-differenced regressors $\Delta\mathbf{X}_{it}$ as

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

with $\Delta\mathbf{H}_i(\mathcal{U}_{it}) := (\Delta H_{i1}, \Delta H_{i2}, \ldots, \Delta H_{id})^\top$.
Similarly, by Assumption (C1), we have

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

Then, the differencing estimator $\widehat{\boldsymbol{\beta}}_i$ can be written as

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_i &= \Big(\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^\top\Big)^{-1}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta Y_{it} \\
&= \Big(\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^\top\Big)^{-1}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Big(\Delta\mathbf{X}_{it}^\top\boldsymbol{\beta}_i + \Delta m_{it} + \Delta\varepsilon_{it}\Big) \\
&= \boldsymbol{\beta}_i + \Big(\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^\top\Big)^{-1}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta m_{it} + \Big(\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^\top\Big)^{-1}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}.
\end{aligned}$$

Hence,

$$\sqrt{T}(\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i) = \Big(\frac{1}{T}\sum_{t=2}^{T} \Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^\top\Big)^{-1}\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta m_{it} + \Big(\frac{1}{T}\sum_{t=2}^{T} \Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^\top\Big)^{-1}\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}. \tag{A.38}$$

We look at the parts that constitute (A.38) independently and for clarification purposes, we break the proof into three steps.
For the sake of simplicity, we focus our attention on the individual vector components and we prove the necessary bounds and inequalities for each of the components separately, combining them together in the end.
Step 1.
First, we take a closer look at the part of the first summand in (A.38), specifically, $\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta m_{it}$.
Fix $j \in 1,\ldots,d$. By Chebyshev's inequality, for any $a > 0$ we have

$$\mathbb{P}\left(\frac{1}{T}\sum_{t=2}^{T}|\Delta H_{ij}(\mathcal{U}_{it})| > a\right) \le \frac{\mathbb{E}\Big[\big(\sum_{t=2}^{T}|\Delta H_{ij}(\mathcal{U}_{it})|\big)^2\Big]}{(T-1)^2a^2} \tag{A.39}$$

and

$$\mathbb{E}\Big[\big(\sum_{t=2}^{T}|\Delta H_{ij}(\mathcal{U}_{it})|\big)^2\Big] = \sum_{t=2}^{T}\mathbb{E}\left[\Delta H_{ij}^2(\mathcal{U}_{it})\right] + \sum_{\substack{t=2,s=2,\\ t\neq s}}^{T}\mathbb{E}\big[\,|\Delta H_{ij}(\mathcal{U}_{it})\Delta H_{ij}(\mathcal{U}_{is})|\,\big]. \tag{A.40}$$

Note that by the Cauchy-Schwarz inequality for all t and s we have

$$\mathbb{E}\big[\,|H_{ij}(\mathcal{U}_{it})H_{ij}(\mathcal{U}_{is})|\,\big] \le \sqrt{\mathbb{E}\big[H_{ij}^2(\mathcal{U}_{it})\big]}\sqrt{\mathbb{E}\big[H_{ij}^2(\mathcal{U}_{is})\big]} = \mathbb{E}\left[H_{ij}^2(\mathcal{U}_{i0})\right] \tag{A.41}$$

and

$$\big|\mathbb{E}\big[H_{ij}(\mathcal{U}_{it})H_{ij}(\mathcal{U}_{is})\big]\big| \le \mathbb{E}\big[\,|H_{ij}(\mathcal{U}_{it})H_{ij}(\mathcal{U}_{is})|\,\big] \le \mathbb{E}\left[H_{ij}^2(\mathcal{U}_{i0})\right].$$

Hence,

$$\begin{aligned}\mathbb{E}\left[\Delta H_{ij}^2(\mathcal{U}_{it})\right] &= \mathbb{E}\left[H_{ij}^2(\mathcal{U}_{it})\right] - 2\mathbb{E}\left[H_{ij}(\mathcal{U}_{it})H_{ij}(\mathcal{U}_{it-1})\right] + \mathbb{E}\left[H_{ij}^2(\mathcal{U}_{it-1})\right]\\ &\le \mathbb{E}\left[H_{ij}^2(\mathcal{U}_{i0})\right] + 2\mathbb{E}\left[H_{ij}^2(\mathcal{U}_{i0})\right] + \mathbb{E}\left[H_{ij}^2(\mathcal{U}_{i0})\right]\\ &= 4\mathbb{E}\left[H_{ij}^2(\mathcal{U}_{i0})\right]\end{aligned}$$

and the first summand in (A.40) can be bounded by $4(T-1)\mathbb{E}\left[H_{ij}^2(\mathcal{U}_{i0})\right]$, where the expectation is finite due to Assumption (C6).

Now to the second summand in (A.40):

$$\begin{aligned}\mathbb{E}\big[\,|\Delta H_{ij}(\mathcal{U}_{it})\Delta H_{ij}(\mathcal{U}_{is})|\,\big] &\leq \mathbb{E}\big[\,|H_{ij}(\mathcal{U}_{it})H_{ij}(\mathcal{U}_{is})|\,\big] + \mathbb{E}\big[\,|H_{ij}(\mathcal{U}_{it-1})H_{ij}(\mathcal{U}_{is})|\,\big] \\ &\quad + \mathbb{E}\big[\,|H_{ij}(\mathcal{U}_{it})H_{ij}(\mathcal{U}_{is-1})|\,\big] + \mathbb{E}\big[\,|H_{ij}(\mathcal{U}_{it-1})H_{ij}(\mathcal{U}_{is-1})|\,\big] \\ &\leq 4\mathbb{E}\left[H_{ij}^2(\mathcal{U}_{i0})\right],\end{aligned}$$

where in the last inequality we used (A.41). This means that the second summand in (A.40) can be bounded by $4(T-1)(T-2)\mathbb{E}\left[H_{ij}^2(\mathcal{U}_{i0})\right]$.
Plugging these bounds in (A.40), we get

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{t=2}^{T}\left|\Delta H_{ij}(\mathcal{U}_{it})\right|\right)^2\right] &\leq 4(T-1)\mathbb{E}\left[H_{ij}^2(\mathcal{U}_{i0})\right] + 4(T-1)(T-2)\mathbb{E}\left[H_{ij}^2(\mathcal{U}_{i0})\right] \\ &= 4(T-1)^2\mathbb{E}\left[H_{ij}^2(\mathcal{U}_{i0})\right],\end{aligned}$$

which together with (A.39) leads to $\frac{1}{T}\sum_{t=2}^{T}\left|\Delta H_{ij}(\mathcal{U}_{it})\right| = O_P(1)$.
Next, by the assumption in Theorem 5.1, $m_i(\cdot)$ is Lipschitz continuous, that is, $|\Delta m_{it}| = \left|m_i\left(\frac{t}{T}\right) - m_i\left(\frac{t-1}{T}\right)\right| \leq C\frac{1}{T}$ for all $t \in \{1,\dots,T\}$ and some constant $C > 0$. Hence,

$$\begin{aligned}\Big|\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\Delta H_{ij}(\mathcal{U}_{it})\Delta m_{it}\Big| &\leq \frac{1}{\sqrt{T}}\sum_{t=2}^{T}\left|\Delta H_{ij}(\mathcal{U}_{it})\right|\cdot\left|\Delta m_{it}\right| \\ &\leq \frac{C}{\sqrt{T}}\cdot\frac{1}{T}\sum_{t=2}^{T}|\Delta H_{ij}(\mathcal{U}_{it})| \\ &= O_P\Big(\frac{1}{\sqrt{T}}\Big).\end{aligned}$$

Since it holds for each $j \in \{1,\dots,d\}$ (and d is fixed), it is obvious that

$$\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta m_{it} = \frac{1}{\sqrt{T}}\sum_{t=2}^{T}\Delta\mathbf{H}_i(\mathcal{U}_{it})\Delta m_{it} = O_P\Big(\frac{1}{\sqrt{T}}\Big). \tag{A.42}$$

Step 2.
Now we look at the other part of the first summand in (A.38), specifically, $\left(\frac{1}{T}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^\top\right)^{-1}$. Using similar arguments as in Step 1 and applying Proposition A.10, we can show that

$$\Big|\frac{1}{T}\sum_{t=2}^{T}\Delta H_{ij}(\mathcal{U}_{it})\Delta H_{ik}(\mathcal{U}_{it})\Big| = O_P(1),$$

for each $j,k \in \{1,\dots,d\}$, which trivially leads to

$$\Big|\frac{1}{T}\sum_{t=2}^{T}\Delta\mathbf{H}_i(\mathcal{U}_{it})\Delta\mathbf{H}_i(\mathcal{U}_{it})^\top\Big| = \Big|\frac{1}{T}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^\top\Big| = O_P(1),$$

where $|A|$ with A being a matrix is any matrix norm.
Furthermore, 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,

$$\left|\left(\frac{1}{T}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^\top\right)^{-1}\right| = O_P(1). \tag{A.43}$$

Step 3
Here we turn our attention to the second summand in (A.38). We already know that $\left|\left(\frac{1}{T}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^\top\right)^{-1}\right| = O_P(1)$. Moreover, by Proposition A.13,

$$\left|\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}\right| = O_P(1).$$

Taking these two facts together, we have that

$$\left(\frac{1}{T}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^\top\right)^{-1}\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta\varepsilon_{it} = O_P(1). \tag{A.44}$$

Finally, from (A.42) and (A.43) we get that the first summand in (A.38) is $O_P(1/\sqrt{T})$, and by (A.44) the second summand is $O_P(1)$. The statement of the theorem follows. $\square$