Multiscale Testing for Equality of Nonparametric Trend Curves

Marina Khismatullina[1]
Erasmus University Rotterdam

Michael Vogt[2]
Ulm University

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, it may be useful to know exactly which of the time trends are different. 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 show that the proposed test has asymptotic power of one against a certain class of local alternatives.

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 functional forms of the objective regression curves are unknown, hence, the parametric approach is not applicable. In this paper, we propose a novel approach that addresses this particular problem in a nonparametric context. Specifically, we present a new testing procedure for detecting differences between 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 $\mathcal{T}_i = \{(Y_{it}, \boldsymbol{X}_{it}) : 1 \le t \le T\}$ for $1 \le i \le n$, where Y_{it} are real-valued random variables and $\boldsymbol{X}_{it} = (X_{it,1}, \dots, X_{it,d})^\top$ are d-dimensional random vectors. Each time series $\mathcal{T}_i$ is modelled by the equation

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

for $1 \le t \le T$, where $\boldsymbol{\beta}_i$ is a $d \times 1$ vector of unknown parameters, $\boldsymbol{X}_{it}$ 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 $\mathcal{E}_i = \{\varepsilon_{it} : 1 \le t \le T\}$ is a zero-mean stationary error process.

[1]Corresponding author. Address: Erasmus School of Economics, Erasmus University Rotterdam, 3062 PA Rotterdam, Netherlands. Email: `khismatullina@ese.eur.nl`.
[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 a common trend. In other words, the researchers would like to know if m_i are the same for all i. When there is evidence that this is not the case, one of the major related statistical problems is to determine which of the trends are different and whether we can group the time series with the similar trends together. Moreover, 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 characterised 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

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

$H_0^{[i,j]}(u,h)$ can thus be viewed as a local null hypothesis that characterises the behaviour of two trend functions locally, whereas H_0 specified in (1.2) is the global null hypothesis that is concerned with the comparison of all of the trends on the whole unit interval.

Trend comparison is a common statistical problem that arises in various contexts. For example, in economics the researchers compare 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, the temperature time series in different geographical areas are investigated in the context of the regional and global warming trends (Karoly and Wu, 2005). Finally, in industry, mobile phone providers are interested in finding the differences between the cell phone download activity in various 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 trends 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 under the assumption that the observed time series have a common time trend. However, in many applications the restriction of including a common time trend in the model is questionable at best. For instance, when we observe a large number of time series it is reasonable to expect that at least some of the trends are different from the others. Consequently, it often makes sense to relax the assumption of a common trend, which leads to more flexible panel settings with heterogeneous trends. Such models 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) and 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 $\boldsymbol{\beta}_i = \boldsymbol{\beta}$ for all i. All of these tests have an important drawback: they involve classical nonparametric estimation of the trend functions that depends on one or several bandwidth parameters which imposes a certain limit on the applicability of such tests since in most cases it is far from clear how to choose these parameters in an appropriate way. Contrary to the aforementioned methods, our multiscale testing procedure allows us to consider a large collection of bandwidths simultaneously avoiding the problem of choosing only one bandwidth altogether.

In this paper, we introduce a multiscale method that allows us to test the hypotheses $H_0^{[i,j]}(u,h)$ in the model (1.1) simultaneously for all pairs (i,j) and all intervals $\mathcal{I}_{(u,h)}$ under consideration in a statistically rigorous way. By using appropriate critical values that depend on the scale of the problem (i.e. the number of hypotheses tested simultaneously and the relationship between them), our methods accounts for the multiple testing problem which arises from considering multiple statistical tests and allows us to make simultaneous confidence statements. Moreover, we show that the suggested procedure for obtaining critical values leads to good theoretical properties of the proposed test: it has the correct (asymptotic) level and an (asymptotic) power of one against a certain class of local alternatives. Based on our test method, we further construct an algorithm which clusters the observed time series into groups with the same trend.

Recently, Khismatullina and Vogt (2021) proposed a new inference method that allows researchers to detect differences between epidemic time trends in the context of the COVID-19 pandemic. In their paper, the authors present a statistically rigorous procedure that, similarly to ours, not only allows to compare trends across different countries, but to pinpoint the time intervals where the differences occur as well. Moreover, they also circumvent the need to pick a bandwidth parameter by using a multiscale testing approach. However, the model in Khismatullina and Vogt (2021) is only a special case of the model (1.1) which includes neither the covariates $\boldsymbol{X}_{it}$, nor the fixed effects α_i. Furthermore, the authors place major restriction on the error terms: in their model, ε_{it} are independent across t. In contrast, our model (1.1) can be regarded as a generalised version of the one in Khismatullina and Vogt (2021) that allows for a wider range of economic and financial applications.

To sum up, the main theoretical contribution of the current paper is the multiscale testing method that allows to make simultaneous confidence statements about which of the time trends are distinct and the regions where they 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 significantly simplified version of the model (1.1). In addition to restricted model, Park et al. (2009) derive theoretical results for their analysis only for the special case of observing only two time series, whereas in other cases, the algorithm is provided without detailed 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, and the technical details and extended proofs are deferred to the Appendix. Section 5 describes a simple clustering algorithm that can be applied to group the time series with the similar trends together. We complement the theoretical analysis of the paper by a simulation study in Section 6, where we investigate the finite sample properties of the test methods and the clustering algorithm from Sections 3 and 5. Section 7 presents two empirical applications to illustrate the usage of our method: testing for common trend in the GDP growth data and cross-country trend comparison of the housing prices. Section 8 concludes.

2 The model framework

2.1 Notation

Throughout the paper, we adopt the following notation. For a vector $\mathbf{v} = (v_1, \ldots, v_m) \in \mathbb{R}^m$, we write $\|\mathbf{v}\|_q = \big(\sum_{i=1}^m v_i^q\big)^{1/q}$ to denote its ℓ_q-norm and use the shorthand $\|\mathbf{v}\| = \|\mathbf{v}\|_2$ in the special case $q = 2$. For a random variable V, we define its $\mathcal{L}^q$-norm by $\|V\|_q = (\mathbb{E}|V|^q)^{1/q}$ and write $\|V\| := \|V\|_2$ in the case $q = 2$.
Let η_t $(t \in \mathbb{Z})$ be independent and identically distributed (i.i.d.) random variables, write $\mathcal{F}_t = (\ldots, \eta_{t-1}, \eta_t)$ and let $g : \mathbb{R}^\infty \to \mathbb{R}$ be a measurable function such that $g(\mathcal{F}_t) = g(\ldots, \eta_{t-1}, \eta_t)$ is a properly defined random variable. Following Wu (2005), we define the *physical dependence measure* of the process $\{g(\mathcal{F}_t)\}_{t=-\infty}^{\infty}$ by

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

where $\mathcal{F}_t' = (\ldots, \eta_{-1}, \eta_0', \eta_1, \ldots, \eta_t)$ is a coupled version of $\mathcal{F}_t$ with η_0' being an i.i.d. copy of η_0. Evidently, $\delta_q(g, t)$ measures the dependency of the random variable $g(\mathcal{F}_t)$ on the innovation term η_0.

2.2 Model

We observe a panel of n time series $\mathcal{T}_i = \{(Y_{it}, \boldsymbol{X}_{it}) : 1 \le t \le T\}$ of length T for $1 \le i \le n$. Each time series $\mathcal{T}_i$ satisfies the model equation

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

for $1 \le t \le T$, where $\boldsymbol{\beta}_i$ is a $d \times 1$ vector of unknown parameters, $\boldsymbol{X}_{it}$ is a $d \times 1$ vector of individual covariates, m_i is an unknown nonparametric trend function defined on the unit interval $[0, 1]$ with $\int_0^1 m_i(u)du = 0$ for all i, α_i is a (deterministic or random) intercept term and $\mathcal{E}_i = \{\varepsilon_{it} : 1 \le t \le T\}$ 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; see e.g. Robinson (1989), Dahlhaus (1997) and Vogt and Linton (2014) for a discussion of rescaled time in nonparametric estimation. The condition $\int_0^1 m_i(u)du = 0$ is required for identification in the presence of the intercept terms α_i. Without imposing this condition, one can freely shift the functions m_i by any (positive or negative) constant c_i while simultaneously subtracting this constant from α_i:

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

The term α_i can be regarded as a fixed effect error term which captures unobserved characteristics of the time series $\mathcal{T}_i$ that remain constant over time. We allow the error terms α_i to be dependent across i in an arbitrary way. We also allow for an arbitrary dependence structure of the individual covariates $\boldsymbol{X}_{it}$ across i. Hence, we allow the n time series $\mathcal{T}_i$ in our panel to be correlated with each other. Whereas the variables α_i and $\boldsymbol{X}_{it}$ may be correlated across time series, the error processes $\mathcal{E}_i$ are assumed to be independent across i. Technical conditions regarding the model are discussed below. 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.

2.3 Assumptions

The error processes $\mathcal{E}_i = \{\varepsilon_{it} : 1 \leq t \leq T\}$ satisfy the following conditions.

(C1) The variables ε_{it} allow for the representation $\varepsilon_{it} = g_i(\mathcal{F}_{it})$, where $\mathcal{F}_{it} = (\dots, \eta_{it-2}, \eta_{it-1}, \eta_{it})$, the variables η_{it} are i.i.d. across t, and $g_i : \mathbb{R}^\infty \to \mathbb{R}$ is a measurable function. It holds that $\mathbb{E}[\varepsilon_{it}] = 0$ and $\|\varepsilon_{it}\|_q \leq C < \infty$ for some $q > 4$ and a sufficiently large constant C.

(C2) The processes $\mathcal{E}_i = \{\varepsilon_{it} : 1 \leq t \leq T\}$ are independent across i.

Assumption (C1) implies that the error processes $\mathcal{E}_i$ are stationary and causal (in the sense that ε_{it} does not depend on future innovations η_{is} with $s > t$). The class of error processes that satisfy condition (C1) is very large. It includes linear processes, nonlinear transformations thereof, 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 $\delta_q(g_i, t)$ defined in (2.1). In particular, we assume the following:

(C3) For each i, it holds that $\sum_{s \geq t} \delta_q(g_i, s) = O(t^{-\gamma}(\log t)^{-A})$ with q from (C1), where $A > \frac{2}{3}(1/q + 1 + \gamma)$ and $\gamma = \{q^2 - 4 + (q-2)\sqrt{q^2 + 20q + 4}\}/8q$.

For fixed i and t, the expression $\sum_{s \geq t} \delta_q(g_i, s)$ measures the cumulative effect of the innovation η_0 on the variables $\varepsilon_{it}, \varepsilon_{it+1}, \dots$ in terms of the $\mathcal{L}^q$-norm. Condition (C3) puts some restrictions on the decay of $\sum_{s \geq t} \delta_q(g_i, s)$ (as a function of t) and in particular implies that $\sum_{s \geq 0} \delta_q(g_i, s)$ is finite. It is fulfilled by a wide range of stationary processes $\mathcal{E}_i$. For a detailed discussion of (C1)–(C3) and some examples of error processes that satisfy these conditions, see Khismatullina and Vogt (2020).

The covariates $\boldsymbol{X}_{it} = (X_{it,1}, \dots, X_{it,d})^\top$ are assumed to have the following properties.

(C4) The variables $X_{it,j}$ allow for the representation $X_{it,j} = h_{ij}(\mathcal{G}_{it,j})$, where $\mathcal{G}_{it,j} = (\ldots, \xi_{it-1,j}, \xi_{it,j})$, the random variables $\xi_{it,j}$ are i.i.d. across t and $h_{ij} : \mathbb{R}^\infty \to \mathbb{R}$ is a measurable function such that $X_{it,j}$ is well-defined. We use the notation $\boldsymbol{X}_{it} = \boldsymbol{h}_i(\mathcal{G}_{it})$ with $\boldsymbol{h}_i = (h_{i1}, \ldots, h_{id})^\top$ and $\mathcal{G}_{it} = (\mathcal{G}_{it,1}, \ldots, \mathcal{G}_{it,d})^\top$. It holds that $\mathbb{E}[X_{it,j}] = 0$ and $\|X_{it,j}\|_{q'} < \infty$ for all i and j, where $q' > \max\{2, \theta q\}$ with q from (C1) and θ specified in (C10) below.

(C5) The matrix $\mathbb{E}[\boldsymbol{X}_{it}\boldsymbol{X}_{it}^\top]$ is invertible for each i.

(C6) For each i and j, it holds that $\sum_{s=t}^\infty \delta_{q'}(h_{ij}, s) = O(t^{-\alpha})$ for some $\alpha > 1/2 - 1/q'$ with q' from (C4).

Assumption (C4) guarantees that the process $\{\boldsymbol{X}_{it} : 1 \le t \le T\}$ is stationary and causal for each i. Similar to the restrictions on the error processes, we employ the definition of the physical dependence measure $\delta_{q'}(h_{ij}, s)$ in Assumption (C6), thus ensuring that the cumulative effect of the innovation ξ_{i0} on the variables $\boldsymbol{X}_{i0}, \boldsymbol{X}_{i1}, \boldsymbol{X}_{i2}, \ldots$ is finite. We finally impose some assumptions on the relationship between the covariates and the errors and on the trend functions m_i.

(C7) The random variables $\Delta \boldsymbol{X}_{it} = \boldsymbol{X}_{it} - \boldsymbol{X}_{it-1}$ and $\Delta\varepsilon_{it} = \varepsilon_{it} - \varepsilon_{it-1}$ are uncorrelated, that is, $\text{Cov}(\Delta \boldsymbol{X}_{it}, \Delta\varepsilon_{it}) = \mathbb{E}[\Delta \boldsymbol{X}_{it} \Delta\varepsilon_{it}] = 0$.

(C8) The trend functions m_i are continuously differentiable on $[0, 1]$ and satisfy the property $\int_0^1 m_i(u)du = 0$ for each i.

Remark 2.1. *The conditions (C4)–(C6) can be relaxed to cover nonstationary regressors as well as stationary ones. For example, (C4) may then be replaced by*

(C4) The variables $X_{it,j}$ allow for the representation $X_{it,j} = h_{ij}(t; \mathcal{G}_{it,j})$, where $\mathcal{G}_{it,j} = (\ldots, \xi_{it-1,j}, \xi_{it,j})$, the random variables $\xi_{it,j}$ are i.i.d. across t and $h_{ij} : \mathbb{R}^\infty \to \mathbb{R}$ is a measurable function such that $X_{it,j}$ is well-defined. We use the notation $\boldsymbol{X}_{it} = \boldsymbol{h}_i(t; \mathcal{G}_{it})$ with $\boldsymbol{h}_i = (h_{i1}, \ldots, h_{id})^\top$ and $\mathcal{G}_{it} = (\mathcal{G}_{it,1}, \ldots, \mathcal{G}_{it,d})^\top$. It holds that $\mathbb{E}[X_{it,j}] = 0$ and $\|X_{it,j}\|_{q'} < \infty$ for all i, j and t, where $q' > \max\{2, \theta q\}$ with q from (C1) and θ specified in (C10) below.*

The other assumptions can be adjusted accordingly. Our main theoretical results will in principle still hold in this case, however, the complexity of the technical arguments will increase drastically. Hence, for the sake of clarity, we restrict our attention only to stationary covariates $\boldsymbol{X}_{it}$.

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 global 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 a challenging task not only because it involves nonparametric estimation of the functions $m_i(\cdot)$, but also due to the presence of 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

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

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

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

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

But before we proceed further, we show how to construct consistent estimates $\widehat{\alpha}_i$ and $\widehat{\boldsymbol{\beta}}_i$. To begin with, 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 $\Delta\mathcal{T}_i = \{(\Delta Y_{it}, \Delta \boldsymbol{X}_{it}) : 2 \le t \le T\}$ of the first differences $\Delta Y_{it} = Y_{it} - Y_{it-1}$ and $\Delta \boldsymbol{X}_{it} = \boldsymbol{X}_{it} - \boldsymbol{X}_{it-1}$. We can write

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

where $\Delta\varepsilon_{it} = \varepsilon_{it} - \varepsilon_{it-1}$. Since m_i is Lipschitz by Assumption (C8), 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

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

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

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

In Lemma A.5 in the Appendix, we show that $\widehat{\boldsymbol{\beta}}_i$ is a consistent estimator of $\boldsymbol{\beta}_i$ with the property $\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i = O_p(T^{-1/2})$ under our assumptions.
Next, given $\widehat{\boldsymbol{\beta}}_i$, consider an appropriate estimator $\widehat{\alpha}_i$ for the intercept α_i calculated by

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

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 normalisation $\int_0^1 m_i(u)du = 0$. Furthermore, $\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{X}_{it} = O_p(1)$ 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 $Y_{it}^\circ := Y_{it} - \boldsymbol{\beta}_i^\top \boldsymbol{X}_{it} - \alpha_i = m_i(t/T) + \varepsilon_{it}$ can be well approximated by $\widehat{Y}_{it}$ since $\widehat{Y}_{it} = Y_{it} - \widehat{\alpha}_i - \widehat{\boldsymbol{\beta}}_i^\top \boldsymbol{X}_{it} = Y_{it}^\circ + O_p(T^{-1/2})$.
We now turn to the estimator of the long-run error variance $\sigma_i^2 = \sum_{\ell=-\infty}^{\infty}\mathrm{Cov}(\varepsilon_{i0}, \varepsilon_{i\ell})$ which is necessary for the construction of the test statistics later on. 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 paper, 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. Our theory works with any estimator $\widehat{\sigma}_i^2$ that has this property.
We now discuss some possible choices of $\widehat{\sigma}_i^2$. Following Kim (2016), we can estimate σ_i for each i by a variant of the subseries variance estimator proposed first by Carlstein (1986) and then extended by Wu and Zhao (2007). Formally, we set

$$\begin{aligned}\widehat{\sigma}_i^2 = \frac{1}{2(M-1)s_T}\sum_{m=1}^{M}\Bigg[\sum_{t=1}^{s_T}\Big(&Y_{i(t+ms_T)} - Y_{i(t+(m-1)s_T)}\\ &- \widehat{\boldsymbol{\beta}}_i^\top\big(\boldsymbol{X}_{i(t+ms_T)} - \boldsymbol{X}_{i(t+(m-1)s_T)}\big)\Big)\Bigg]^2,\end{aligned} \tag{3.4}$$

where s_T is the length of subseries and $M = \lfloor T/s_T \rfloor$ is the largest integer not exceeding T/s_T. As per the optimality result in Carlstein (1986), we set $s_T \asymp T^{1/3}$. For a

finite sample, we choose $s_T = \lfloor T^{1/3} \rfloor$. According to Lemma A.6 in Appendix, $\widehat{\sigma}_i^2$ is an asymptotically consistent estimator of σ_i^2 with the rate of convergence $O_p(T^{-1/3})$.
The largest advantage of the subseries variance estimator described above is that it can be used without imposing any additional assumptions on the error processes $\mathcal{E}_i$. However, as noted in Khismatullina and Vogt (2020), estimating the long-run error variance in the presence of a pronounced trend and under general weak dependence conditions is a particularly difficult problem. Estimators often tend to be quite imprecise. Hence, in practice we opt for imposing certain some time series models on the error processes $\mathcal{E}_i$. For example, we can assume that for each i the error process $\mathcal{E}_i$ has the AR(∞) structure which covers a wide range of applications. In this case, it is possible to obtain more precise estimates $\widehat{\sigma}_i^2$ using the difference-based method described in Khismatullina and Vogt (2020). The detailed discussion of the behaviour of such an estimator in the presence of a trend and comparison with other long-run variance estimators can be found in Khismatullina and Vogt (2020).

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

$$\widehat{\psi}_{ij,T}(u, h) = \sum_{t=1}^{T} w_{t,T}(u, h)(\widehat{Y}_{it} - \widehat{Y}_{jt}), \tag{3.5}$$

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

$$w_{t,T}(u, h) = \frac{\Lambda_{t,T}(u, h)}{\{\sum_{t=1}^{T} \Lambda_{t,T}(u, h)^2\}^{1/2}}, \tag{3.6}$$

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 $\ell = 1, 2$ and K is a kernel function. As common in nonparametric estimation, we assume that K has the following properties:

(C9) 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, $|K(v) - K(w)| \leq C_K|v - w|$ for any $v, w \in \mathbb{R}$ and some constant $C_K > 0$.

Assumption (C9) allows us to use usual 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 $\mathcal{I}_{(u,h)} = [u-h, u+h]$. However, instead of working

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

$$\hat{\psi}^0_{ij,T}(u,h) = \left| \frac{\widehat{\psi}_{ij,T}(u,h)}{(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2)^{1/2}} \right| - \lambda(h). \tag{3.7}$$

Here, $\lambda(h) = \sqrt{2\log\{1/(2h)\}}$ 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 (scales) h:

$$\widehat{\Psi}_{n,T} = \max_{1\le i<j\le n} \max_{(u,h)\in\mathcal{G}_T} \hat{\psi}^0_{ij,T}(u,h), \tag{3.8}$$

In (3.8), $\mathcal{G}_T$ stands for the set of location-bandwidth pairs (u,h) that was mentioned in Section 1. 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 $\mathcal{G}_T^{\text{full}} = \{(u,h) : u = t/T \text{ and } h = s/T \text{ for some } 1 \le t, s \le T \text{ such that } h \in [h_{\min}, h_{\max}]\}$, where $h_{\min}$ and $h_{\max}$ denote some minimal and maximal bandwidth value, respectively. As was already discussed in Section 1, 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:

(C10) $|\mathcal{G}_T| = O(T^\theta)$ for some arbitrarily large but fixed constant $\theta > 0$, where $|\mathcal{G}_T|$ denotes the cardinality of $\mathcal{G}_T$.

(C11) $h_{\min} \gg T^{-(1-\frac{2}{q})} \log T$, that is, $h_{\min}/\{T^{-(1-\frac{2}{q})}\log T\} \to \infty$ with $q > 4$ defined in (C2) and $h_{\max} = o(1)$.

Assumption (C10) 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 (C11) is concerned with the minimal and the maximal bandwidths that we use for our analysis. Specifically, according to Assumption (C11), 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. The maximal bandwidth $h_{\max}$ needs only to slowly converge to zero as the sample size grows, hence, in finite samples it can be picked quite 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 show how to test the null hypothesis $H_0 : m_1 = m_2 = \ldots = m_n$ using the multiscale 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 $\boldsymbol{X}_{js}$ 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. Similarly as with $\hat{\psi}^0_{ij,T}(u,h)$, for each i and j, we introduce the normalised and corrected Gaussian kernel averages

$$\phi^0_{ij,T}(u,h) = \left| \frac{\phi_{ij,T}(u,h)}{(\sigma_i^2 + \sigma_j^2)^{1/2}} \right| - \lambda(h), \tag{3.9}$$

where

$$\phi_{ij,T}(u,h) = \sum_{t=1}^{T} w_{t,T}(u,h) \left\{ \sigma_i (Z_{it} - \bar{Z}_i) - \sigma_j (Z_{jt} - \bar{Z}_j) \right\} \tag{3.10}$$

with $w_{t,T}(u,h)$ defined in (3.6).
Next, in the same way as in (3.8), we define the global Gaussian test statistic

$$\Phi_{n,T} = \max_{1 \leq i < j \leq n} \max_{(u,h) \in \mathcal{G}_T} \phi^0_{ij,T}(u,h) \tag{3.11}$$

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 normalised kernel averages (3.9) as*

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

which means that the distribution of the Gaussian test statistics does not depend neither on the data $\mathcal{T}_i = \{(Y_{it}, \boldsymbol{X}_{it}) : 1 \leq t \leq T)\}, \mathcal{T}_j = \{(Y_{jt}, \boldsymbol{X}_{jt}) : 1 \leq t \leq T)\}$, 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 (approximately) calculating the quantiles of $\Phi_{n,T}$ by the Monte Carlo simulations in Sections 6 and 7. However, throughout the paper, we will stick to the definition (3.9) for the sake of similarity to $\hat{\psi}^0_{ij,T}(u,h)$.

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 explicitly include this dependence 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. This fact does not provide us with a lot of information about the behaviour of the trend functions m_i: we can only state with a given statistical confidence that some of the trend functions are not equal somewhere on $[0, 1]$, 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

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

We view $H_0^{[i,j]}(u, h)$ as the local null hypothesis because it compares two trend functions m_i and m_j locally, on the interval $\mathcal{I}_{(u,h)} = [u-h, u+h]$. In contrast, we refer to H_0 introduced in (1.2) as the global null hypothesis.

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

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

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 at least* $1-\alpha$ *that for all* i, j*,* $1 \leq i < j \leq n$*, we have that* m_i *and* m_j *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 $\{Z^{(\ell)}_{it} : 1 \leq t \leq T, 1 \leq i \leq n\}$ for

$1 \leq \ell \leq N$. 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 $\{\Phi_{n,T}^{(\ell)} : 1 \leq \ell \leq N\}$. 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 $\widehat{\Psi}_{n,T} > \hat{q}_{n,T}(\alpha)$. Reject the null if this is the case.

Step 4. For each i, j, $1 \leq i < j \leq n$, 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 $\hat{\psi}^0_{ij,T}(u,h) > \hat{q}_{n,T}(\alpha)$. For each pair of time series (i,j), find the set of intervals $\mathcal{S}^{[i,j]}(\alpha)$ that consists of the intervals $\mathcal{I}_{(u,h)}$ where we reject $H_0^{[i,j]}(u,h)$.

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 $H_0^{[i,j]}(u,h)$.

In some cases, the number of intervals in the set $\mathcal{S}^{[i,j]}(\alpha)$ may be quite large, making the visual representation of the graphical summary of the results more complicated. To overcome this drawback, we work with the subset of minimal intervals $\mathcal{S}^{[i,j]}_{\min}(\alpha) \subseteq \mathcal{S}^{[i,j]}(\alpha)$ which is constructed as follows: As in Dümbgen (2002), we call an interval $\mathcal{I}_{(u,h)} \in \mathcal{S}^{[i,j]}(\alpha)$ minimal if there is no other interval $\mathcal{I}_{(u',h')} \in \mathcal{S}^{[i,j]}(\alpha)$ such that $\mathcal{I}_{(u',h')} \subset \mathcal{I}_{(u,h)}$. We denote the set of all minimal intervals by $\mathcal{S}^{[i,j]}_{\min}(\alpha)$. All of our theoretical results can be rewritten using the subset of minimal intervals $\mathcal{S}^{[i,j]}_{\min}(\alpha)$ instead of $\mathcal{S}^{[i,j]}(\alpha)$.

4 Theoretical properties of the test

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

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

where

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

and

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

We here use the notation $\bar{\varepsilon}_i = \bar{\varepsilon}_{i,T} := T^{-1}\sum_{t=1}^{T}\varepsilon_{it}$ and $\bar{\boldsymbol{X}}_i = \bar{\boldsymbol{X}}_{i,T} := T^{-1}\sum_{t=1}^{T}\boldsymbol{X}_{it}$. By construction, it holds that $\widehat{\phi}_{ij,T}(u,h) = \widehat{\psi}_{ij,T}(u,h)$ under $H_0^{[i,j]}(u,h)$. This implies that $\widehat{\Phi}_{n,T}$ is identical to $\widehat{\Psi}_{n,T}$ under the global null H_0. Hence, in order to determine the distribution of our main test statistic $\widehat{\Psi}_{n,T}$ under H_0, we can study the auxiliary statistic $\widehat{\Phi}_{n,T}$. The following theorem shows that the distribution of $\widehat{\Phi}_{n,T}$ is close to the distribution of the Gaussian statistic $\Phi_{n,T}$ introduced in (3.11).

Theorem 4.1. *Let (C1)–(C11) be fulfilled and assume that $\sigma_i^2 = \sigma^2$, $\widehat{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$ with $\rho_T = o(\sqrt{h_{\min}}/\log T)$ for all i. Then*

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

Theorem 4.1 is the principal instrument for deriving theoretical properties of our multiscale test. Its proof is provided in the Appendix.

Remark 4.2. *The proof of Theorem 4.1 builds on two important theoretical results: strong approximation theory developed in Berkes et al. (2014) and anti-concentration results proved in Chernozhukov et al. (2015). These results were already combined together for the purpose of developing the multiscale test for dependent data in Khismatullina and Vogt (2020). We can say that our proof can be regarded as a further development of the proof strategy in Khismatullina and Vogt (2020) where they proposed a similar testing procedure for investigating properties of the trend function in one time series. We extend their theoretical result not only by working with multiple time series, but also by including the covariate terms in the model (1.1). Hence, our proof strategy builds on the similar stones but is much more technically involved.*

Now we examine the theoretical properties of the testing procedure proposed in Sections 3.3 and 3.4 with the help of Theorem 4.1. The following proposition states that our test has correct (asymptotic) size.

Proposition 4.3. *Suppose that the conditions of Theorem 4.1 are satisfied. Then under H_0, we have*

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

The next proposition characterises the behaviour of our multiscale test under a certain class of local alternatives. To formulate it, we consider a sequence of pairs of functions $m_i := m_{i,T}$ and $m_j := m_{j,T}$ that depend on the sample size and that are locally sufficiently far from each other.

Proposition 4.4. *Let the conditions of Theorem 4.1 be satisfied. Moreover, assume that for some pair of indices i and j, the functions $m_i = m_{i,T}$ and $m_j = m_{j,T}$ have the following property: There exists $(u,h) \in \mathcal{G}_T$ with $[u-h,u+h] \subseteq [0,1]$ such that $m_{i,T}(w) - m_{j,T}(w) \ge c_T\sqrt{\log T/(Th)}$ for all $w \in [u-h,u+h]$ or $m_{j,T}(w) - m_{i,T}(w) \ge$*

$c_T\sqrt{\log T/(Th)}$ *for all* $w \in [u-h, u+h]$, *where* $\{c_T\}$ *is any sequence of positive numbers with* $c_T \to \infty$. *Then*

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

Finally, we turn our attention to the local null hypotheses $H_0^{[i,j]}(u,h)$. Since we are testing many hypotheses at the same time, we would like to bound the probability of making even one false discovery. For this purpose, we employ the notion of the family-wise error rate (FWER) which is equal to the probability of making one or more type I errors. Formally, the FWER is defined as

$$\begin{aligned}\text{FWER}(\alpha) = \mathbb{P}\Big(\exists\, i,j \in \{1,\ldots,n\}, (u,h) \in \mathcal{G}_T : \mathcal{I}_{(u,h)} \in \mathcal{S}^{[i,j]}(\alpha) \\ \text{and } H_0^{[i,j]}(u,h) \text{ is true}\Big).\end{aligned}$$

We say that the FWER is controlled at level α if $\text{FWER}(\alpha) \le \alpha$. The following result assures that for our testing procedure, this is indeed the case.

Proposition 4.5. *Suppose that the conditions of Theorem 4.1 are satisfied. Then*

$$\text{FWER}(\alpha) \le \alpha.$$

The following corollary is an immediate consequence of Proposition 4.5 and gives the theoretical justification necessary for making simultaneous confidence statements about the locations of the differences between the trends.

Corollary 4.6. *Under the conditions of Theorem 4.1, for any given* $\alpha \in (0,1)$, *we have*

$$\begin{aligned}\mathbb{P}\Big(\forall\, i,j \in \{1,\ldots,n\}, (u,h) \in \mathcal{G}_T \text{ s.t. } H_0^{[i,j]}(u,h) \text{ is true} : \hat{\psi}^0_{ij,T}(u,h) \le q_{n,T}(\alpha)\Big) \\ \ge 1-\alpha+o(1).\end{aligned}$$

With the help of Corollary 4.6, we are able to make simultaneous confidence statements about which of the trends are different and where:

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

5 Clustering

Consider a situation in which the null hypothesis $H_0 : m_1 = m_2 = \ldots = m_n$ is violated. Even though some of the trend functions are different in this case, part of them may still be the same. Put differently, there may be groups of time series which have the same time trend. Formally speaking, we define a group structure as follows: There exist sets or groups of time series $G_1, \ldots, G_N$ with $N \le n$ and $\{1,\ldots,n\} = \dot{\bigcup}_{\ell=1}^{N} G_\ell$ such that for each $1 \le \ell \le N$,

$$m_i = f_\ell \quad \text{for all } i \in G_\ell,$$

where f_ℓ are group-specific trend functions. Hence, the time series which belong to the group G_ℓ all have the same time trend f_ℓ. Throughout the section, we suppose that the group-specific trend functions f_ℓ have the following properties: For each ℓ, $f_\ell = f_{\ell,T}$ is a Lipschitz continuous function with $\int_0^1 f_{\ell,T}(w)dw = 0$. In particular, it holds that $|f_{\ell,T}(v) - f_{\ell,T}(w)| \le L|v-w|$ for all $v, w \in [0,1]$ and some constant $L < \infty$ that does not depend on T. Moreover, for any $\ell \neq \ell'$, the trends $f_{\ell,T}$ and $f_{\ell',T}$ are assumed to differ in the following sense: There exists $(u,h) \in \mathcal{G}_T$ with $\mathcal{I}_{(u,h)} = [u-h, u+h] \subseteq [0,1]$ such that $f_{\ell,T}(w) - f_{\ell',T}(w) \ge c_T\sqrt{\log T/(Th)}$ for all $w \in \mathcal{I}_{(u,h)}$ or $f_{\ell',T}(w) - f_{\ell,T}(w) \ge c_T\sqrt{\log T/(Th)}$ for all $w \in \mathcal{I}_{(u,h)}$, where $0 < c_T \to \infty$.

In many applications, it is natural to suppose that there is a group structure in the data. In this case, a particular interest lies in estimating the unknown groups from the data at hand. In what follows, we combine our multiscale methods with a clustering algorithm to achieve this. More specifically, we use the aggregated multiscale statistics $\max_{(u,h)\in\mathcal{G}_T} \hat{\psi}^0_{ij,T}(u,h)$ calculated for each i and j, $1 \le i < j \le n$, as distance measures which are fed into a hierarchical clustering algorithm. To describe the algorithm, we first need to introduce the notion of a dissimilarity measure: Let $S \subseteq \{1,\dots,n\}$ and $S' \subseteq \{1,\dots,n\}$ be two sets of time series from our sample. We define a dissimilarity measure between S and S' by setting

$$\widehat{\Delta}(S,S') = \max_{\substack{i\in S,\\ j\in S'}} \Big(\max_{(u,h)\in\mathcal{G}_T} \hat{\psi}^0_{ij,T}(u,h) \Big). \tag{5.1}$$

This is commonly called a complete linkage measure of dissimilarity. Alternatively, we may work with an average or a single linkage measure. We now combine the dissimilarity measure $\widehat{\Delta}$ with a hierarchical agglomerative clustering (HAC) algorithm which proceeds as follows:

Step 0 (Initialisation): Let $\widehat{G}^{[0]}_i = \{i\}$ denote the i-th singleton cluster for $1 \le i \le n$ and define $\{\widehat{G}^{[0]}_1, \dots, \widehat{G}^{[0]}_n\}$ to be the initial partition of time series into clusters.

Step r (Iteration): Let $\widehat{G}^{[r-1]}_1, \dots, \widehat{G}^{[r-1]}_{n-(r-1)}$ be the $n-(r-1)$ clusters from the previous step. Determine the pair of clusters $\widehat{G}^{[r-1]}_\ell$ and $\widehat{G}^{[r-1]}_{\ell'}$ for which

$$\widehat{\Delta}(\widehat{G}^{[r-1]}_\ell, \widehat{G}^{[r-1]}_{\ell'}) = \min_{1\le k<k'\le n-(r-1)} \widehat{\Delta}(\widehat{G}^{[r-1]}_k, \widehat{G}^{[r-1]}_{k'})$$

and merge them into a new cluster.

Iterating this procedure for $r = 1,\dots,n-1$ yields a tree of nested partitions $\{\widehat{G}^{[r]}_1, \dots \dots, \widehat{G}^{[r]}_{n-r}\}$, which can be graphically represented by a dendrogram. Roughly speaking, the HAC algorithm merges the n singleton clusters $\widehat{G}^{[0]}_i = \{i\}$ step by step until we end up with the cluster $\{1,\dots,n\}$. In each step of the algorithm, the closest two clusters are merged, where the distance between clusters is measured in terms of the dissimilarity $\widehat{\Delta}$. We refer the reader to Section 14.3.12 in Hastie et al. (2009) for an overview of hierarchical clustering methods.

When the number of groups N is known, we estimate the group structure $\{G_1, \ldots, G_N\}$ by the N-partition $\{\widehat{G}_1^{[n-N]}, \ldots, \widehat{G}_N^{[n-N]}\}$ produced by the HAC algorithm. When N is unknown, we estimate it by the $\widehat{N}$-partition $\{\widehat{G}_1^{[n-\widehat{N}]}, \ldots, \widehat{G}_{\widehat{N}}^{[n-\widehat{N}]}\}$, where $\widehat{N}$ is an estimator of N. The latter is defined as

$$\widehat{N} = \min\Big\{ r = 1, 2, \ldots \,\Big|\, \max_{1 \le \ell \le r} \widehat{\Delta}\big(\widehat{G}_\ell^{[n-r]}\big) \le q_{n,T}(\alpha) \Big\},$$

where we write $\widehat{\Delta}(S) = \widehat{\Delta}(S, S)$ for short and $q_{n,T}(\alpha)$ is the $(1-\alpha)$-quantile of $\Phi_{n,T}$ defined in Section 3.3.

The following proposition summarises the theoretical properties of the estimators $\widehat{N}$ and $\{\widehat{G}_1, \ldots, \widehat{G}_{\widehat{N}}\}$, where we use the shorthand $\widehat{G}_\ell = \widehat{G}_\ell^{[n-\widehat{N}]}$ for $1 \le \ell \le \widehat{N}$.

Proposition 5.1. *Let the conditions of Theorem 4.1 be satisfied. Then*

$$\mathbb{P}\Big(\big\{ \widehat{G}_1, \ldots, \widehat{G}_{\widehat{N}} \big\} = \big\{ G_1, \ldots, G_N \big\} \Big) \ge (1-\alpha) + o(1)$$

and

$$\mathbb{P}\big(\widehat{N} = N \big) \ge (1-\alpha) + o(1).$$

This result allows us to make statistical confidence statements about the estimated clusters $\{\widehat{G}_1, \ldots, \widehat{G}_{\widehat{N}}\}$ and their number $\widehat{N}$. In particular, we can claim with asymptotic confidence $\ge 1-\alpha$ that the estimated group structure is identical to the true group structure. Note that it is possible to let the significance level α depend on the sample size T in Proposition 5.1. In particular, we can allow $\alpha = \alpha_T$ to converge slowly to zero as $T \to \infty$, in which case we obtain that $\mathbb{P}(\{\widehat{G}_1, \ldots, \widehat{G}_{\widehat{N}}\} = \{G_1, \ldots, G_N\}) \to 1$ and $\mathbb{P}(\widehat{N} = N) \to 1$. The proof of Proposition 5.1 can be found in the Appendix.

Our multiscale methods do not only allow us to compute estimators of the unknown groups $G_1, \ldots, G_N$. They also provide information on the locations where two group-specific trend functions f_ℓ and $f_{\ell'}$ differ from each other. To turn this claim into a mathematically precise statement, we need to introduce some notation. First of all, note that the indexing of the estimators $\widehat{G}_1, \ldots, \widehat{G}_{\widehat{N}}$ is completely arbitrary. We could, for example, change the indexing according to the rule $\ell \mapsto \widehat{N} - \ell + 1$. In what follows, we suppose that the estimated groups are indexed such that $P(\widehat{G}_\ell = G_\ell \text{ for all } \ell) \ge (1-\alpha) + o(1)$. Proposition 5.1 implies that this is possible without loss of generality. Keeping this convention in mind, we define the sets

$$\mathcal{A}_{n,T}^{[\ell,\ell']}(\alpha) = \Big\{ (u,h) \in \mathcal{G}_T : \widehat{\psi}_{ij,T}^0(u,h) > q_{n,T}(\alpha) \text{ for some } i \in \widehat{G}_\ell, j \in \widehat{G}_{\ell'} \Big\}$$

and

$$\mathcal{S}_{n,T}^{[\ell,\ell']}(\alpha) = \big\{ \mathcal{I}_{(u,h)} = [u-h, u+h] : (u,h) \in \mathcal{A}_{n,T}^{[\ell,\ell']}(\alpha) \big\}$$

for $1 \leq \ell < \ell' \leq \widehat{N}$. An interval $\mathcal{I}_{(u,h)}$ is contained in $\mathcal{S}^{[\ell,\ell']}_{n,T}(\alpha)$ if our multiscale test indicates a significant difference between the trends m_i and m_j on the interval $\mathcal{I}_{(u,h)}$ for some $i \in \widehat{G}_\ell$ and $j \in \widehat{G}_{\ell'}$. Put differently, $\mathcal{I}_{(u,h)} \in \mathcal{S}^{[\ell,\ell']}_{n,T}(\alpha)$ if the test suggests a significant difference between the trends of the ℓ-th and the ℓ'-th group on the interval $\mathcal{I}_{(u,h)}$. We further let

$$E^{[\ell,\ell']}_{n,T}(\alpha) = \Big\{\forall \mathcal{I}_{(u,h)} \in \mathcal{S}^{[\ell,\ell']}_{n,T}(\alpha) : f_\ell(v) \neq f_{\ell'}(v) \text{ for some } v \in \mathcal{I}_{(u,h)} = [u-h, u+h]\Big\}$$

be the event that the group-specific time trends f_ℓ and $f_{\ell'}$ differ on all intervals $\mathcal{I}_{(u,h)} \in \mathcal{S}^{[\ell,\ell']}_{n,T}(\alpha)$. With this notation at hand, we can make the following formal statement whose proof is given in the Appendix. (add proof)

Proposition 5.2. *Under the conditions of Theorem 4.1, the event*

$$E_{n,T}(\alpha) = \Big\{\bigcap_{1\leq \ell<\ell'\leq \widehat{N}} E^{[\ell,\ell']}_{n,T}(\alpha)\Big\} \cap \Big\{\widehat{N} = N \text{ and } \widehat{G}_\ell = G_\ell \text{ for all } \ell\Big\}$$

asymptotically occurs with probability $\geq 1-\alpha$, *that is,*

$$\mathbb{P}\big(E_{n,T}(\alpha)\big) \geq (1-\alpha) + o(1).$$

The statement of Proposition 5.2 remains to hold true when the sets of intervals $\mathcal{S}^{[\ell,\ell']}_{n,T}(\alpha)$ are replaced by the corresponding sets of minimal intervals. According to Proposition 5.2, the sets $\mathcal{S}^{[\ell,\ell']}_{n,T}(\alpha)$ allow us to locate, with a pre-specified confidence, time regions where the group-specific trend functions f_ℓ and $f_{\ell'}$ differ from each other. In particular, we can claim with asymptotic confidence $\geq 1-\alpha$ that the trend functions f_ℓ and $f_{\ell'}$ differ on all intervals $\mathcal{I}_{(u,h)} \in \mathcal{S}^{[\ell,\ell']}_{n,T}(\alpha)$.

6 Simulations

In this Section, we present the results for the simulation study for the testing method described in Section 3. The simulation design is as follows. We generate data from the model $Y_{it} = m_i(\frac{t}{T}) + \varepsilon_{it}$, where the number of time series i is set to $n = 15$ and we consider different time series lengths T. For each i, the errors ε_{it} are drawn from the AR(1) model $\varepsilon_{it} = a\varepsilon_{i,t-1} + \eta_{it}$, where $a = 0.5$ and the innovations η_{it} are i.i.d. normally distributed with mean 0 and variance 0.25. To generate data under the null $H_0 : m_1 = \ldots = m_n$, we let $m_i = 0$ for all i without loss of generality. To produce data under the alternative, we define $m_1(u) = \beta\,(u - 0.5)$ with $\beta = 1, 2, 3$ and set $m_i = 0$ for all $i \neq 1$. Hence, all trend functions are the same except for m_1 which is an increasing linear function. We use a linear function as the normalisation constraint $\int_0^1 m_1(u)du = 0$ is directly satisfied in this case. The test is implemented as follows. We use an Epanechnikov kernel and we define the grid as $\mathcal{G}_T = U_T \times H_T$ with

$$U_T = \Big\{u \in [0,1] : u = \tfrac{t}{T} \text{ for some } t \in \mathbb{N}\Big\}$$
$$H_T = \Big\{h \in \big[\tfrac{\log T}{T}, \tfrac{1}{4}\big] : h = \tfrac{5t-3}{T} \text{ for some } t \in \mathbb{N}\Big\}.$$

Hence, the number of data points in each interval $\mathcal{I}_{u,h} = [u-h, u+h]$ is equal to $5, 15, 25, \ldots$. Furthermore, since the errors follow an AR(1) process, we estimate the long-run variances σ_i^2 using the difference-based estimator proposed in Khismatullina and Vogt (2020) setting the tuning parameters q and r to 10 and 25, respectively. In order to compute the critical values of the test, we simulate 1000 values of the statistic $\Phi_{n,T}$ defined in Section 3.3 in (3.11) and compute their empirical $(1-\alpha)$ quantile $q_{n,T}(\alpha)$. Note that in general the statistic $\Phi_{n,T}$ depends on the long-run error variances σ_i^2. However, as stated in the Remark 3.1, under the assumption that the long-run error variance is known to be the same across i (that is, $\sigma_i^2 = \sigma^2$ for all i), we can simplify the formula for the kernel averages $\phi^0_{ij,T}(u,h)$ so that the distribution of $\Phi_{n,T}$ does not depend on any unknown quantities, including the value of the long-run variance σ_i^2. We employ this fact to (approximately) calculate the gaussian quantiles $q_{n,T}(\alpha)$ which are further used for testing.

The simulation results are reported in Tables **??** and **??**. The entries of the tables are computed as follows. Inspecting Table **??**, the actual size of the test can be seen to approximate the nominal target α quite well even for small values of T. Moreover, as can be seen from Table **??**, the test also has reasonable power against the alternatives considered. For the smallest slope $\beta = 1$ and the smallest sample size $T = 250$, the power is only moderate, reflecting the fact that the alternative is not very far away from the null. However, as we increase the slope β and the sample size T, the power quickly increases. For the largest slope $\beta = 3$ and $T =$??, we already reach a power of 1.00.

We finally investigate the finite sample performance of the clustering algorithm from Section 5. To do so, we partition the $n = 15$ time series into $N = 3$ groups, each containing 5 time series. Specifically, we set $G_1 = \{1, \ldots, 5\}$, $G_2 = \{6, \ldots, 10\}$ and $G_3 = \{11, \ldots, 15\}$. Moreover, we define the group-specific trend functions g_1, g_2 and g_3 by $g_1(u) = 0$, $g_2(u) = 1 \cdot (u - 0.5)$ and $g_3(u) = (-1) \cdot (u - 0.5)$. In order to compute our estimators of the groups G_1, G_2, G_3 and their number $N = 3$, we use the same implementation as before followed by the clustering procedure from Section 5. The estimation results are reported in Table 1. The entries in Table 1a are computed as the number of simulations for which $\widehat{N} = N$ divided by the total number of simulations. They thus specify the empirical probabilities with which the estimate $\widehat{N}$ is equal to the true number of groups $N = 3$. Analogously, the entries of Table 1b give the empirical probabilities with which the estimated group structure $\{\widehat{G}_1, \ldots, \widehat{G}_{\widehat{N}}\}$ equals the true one $\{G_1, G_2, G_3\}$.

The simulation results nicely illustrate the theoretical properties of our clustering algorithm. According to Proposition 5.1, the probability that $\widehat{N} = N$ and $\{\widehat{G}_1, \ldots, \widehat{G}_{\widehat{N}}\} = \{G_1, G_2, G_3\}$ should be at least $(1-\alpha)$ asymptotically. For the sample size $T = 1000$, the empirical probabilities reported in Table 1 can indeed be seen to exceed the value $(1-\alpha)$ as predicted by Proposition 5.1. For the smaller sample sizes $T = 250$ and $T = 500$, in contrast, some of the empirical probabilities are much smaller than $(1-\alpha)$.

Table 1: Clustering results for different sample sizes T and nominal sizes α.

(a) Empirical probabilities that $\widehat{N} = N$

	nominal size α		
T	0.01	0.05	0.1
250	0.112	0.267	0.398
500	0.647	0.851	0.895
1000	0.995	0.974	0.952

(b) Empirical probabilities that $\{\widehat{G}_1, \dots, \widehat{G}_{\widehat{N}}\} = \{G_1, G_2, G_3\}$

	nominal size α		
T	0.01	0.05	0.1
250	0.021	0.051	0.089
500	0.524	0.688	0.723
1000	0.994	0.973	0.951

This reflects the asymptotic nature of Proposition 5.1 and is not very surprising. It simply mirrors the fact that for small sample sizes, the effective noise level in the simulated data is quite high. Even though some of the empirical probabilities for $T = 250$ and $T = 500$ are clearly below the target $(1-\alpha)$, they are still quite substantial. Hence, even for these small sample sizes, our estimates $\widehat{N}$ and $\{\widehat{G}_1, \dots, \widehat{G}_{\widehat{N}}\}$ are equal to the true values in a large number of simulations.

7 Applications

7.1 Analysis of the GDP growth

To illustrate our test method from Section 3, we repeat an application example from Zhang et al. (2012) where the authors test the hypothesis of a common trend in the GDP growth data for 16 OECD countries. Since we do not have access to the original dataset from Zhang et al. (2012) and to the exact data specifications the authors used in their paper, we perform our analysis on the data available from the common sources: Refinitiv Datastream, OECD.Stat database, Federal Reserve Economics Data (FRED) and Barro-Lee Educational Attainment dataset (Barro and Lee, 2013). In our illustration example, we consider the specification of the data that is as close as possible to the one in Zhang et al. (2012) with one important distinction. In the original study, the authors examine 16 OECD countries (not specifying which ones), whereas we consider only 11 countries[1] due to the availability of the data. The OECD.Stat database is the main source for the data on major economic indicators, and it contains quarterly data of good quality covering the time period used in the original study only for 11 countries used in our analysis. Similar data for other OECD countries in OECD.Stat database and in other notable sources contain many missing values. In the appendix, we repeat the analysis for aforementioned 11 countries and additional 5 countries using linear interpolation to impute the missing values in the data. Details how this interpolation is

[1]Australia, Austria, Canada, Finland, France, Germany, Japan, Norway, Switzerland, UK, and USA.

done are deferred to the Appendix.
We consider the same time period as in Zhang et al. (2012) from the fourth quarter of 1975 up to and including the third quarter of 2010. We collect the data from multiple sources. In the following list, we provide the specifications for the variables that we use in our analysis.

1. **GDP:** We use data on the Gross Domestic Product – Expenditure Approach (GDP) from the OECD.Stat database (https://stats.oecd.org/Index.aspx). The data are freely available and were accessed on 7 December 2021. To be as close as possible to the specification of the data in Zhang et al. (2012), we use the seasonally adjusted quarterly data on the GDP expressed in millions of 2015 US dollars.[2] The data span from 1960 to 2021 which fully covers the analysed time period.

2. **Capital:** We use data on Gross Fixed Capital Formation (K) from the OECD.Stat database (https://stats.oecd.org/Index.aspx). The data are freely available and were accessed on 7 December 2021. The values are at a quarterly frequency, seasonally adjusted, and expressed in the millions of the 2015 US dollars. In contrast with Zhang et al. (2012), where they use the data on Capital Stock at Constant National Prices, we choose to work with Gross Fixed Capital Formation due to the availability of the data. It is worth noting that since accurate data on capital stock is notoriously difficult to collect, the use of gross fixed capital formation as a measure of capital while explaining economic growth is standard in the literature (see, e.g., Sharma and Dhakal (1994), Lee et al. (2002), Lee (2005)).

3. **Labour:** We collect the data on the Number of Employed People (L) from various sources. For most of the countries (Austria, Australia, Canada, Germany, Japan, UK and the USA) we download the OECD data on Employed Population: Aged 15 and Over retrieved from FRED (https://fred.stlouisfed.org/, accessed on 7 December 2021). The data for France and Switzerland were downloaded from Refinitiv Datastream on 7 December 2021. For all of the aforementioned countries the observations are at a quarterly frequency and seasonally adjusted. The data for Finland and Norway were also obtained via Refinitiv Datastream on 7 December 2021, however, the only quarterly time series that are long enough for our purposes are not seasonally adjusted. Hence, for these two countries we perform the seasonal adjustment ourselves. We do it using the default method of the function `seas` from an `R` package `seasonal` (Sax and Eddelbuettel, 2018) which is an interface to X-13-ARIMA-SEATS, the seasonal adjustment software

[2]Since the publication of the original paper in 2013, the OECD reference year has changed from 2005 to 2015. We have decided to analyse the latest version of the data in order to be able to make more accurate and up-to-date conclusions.

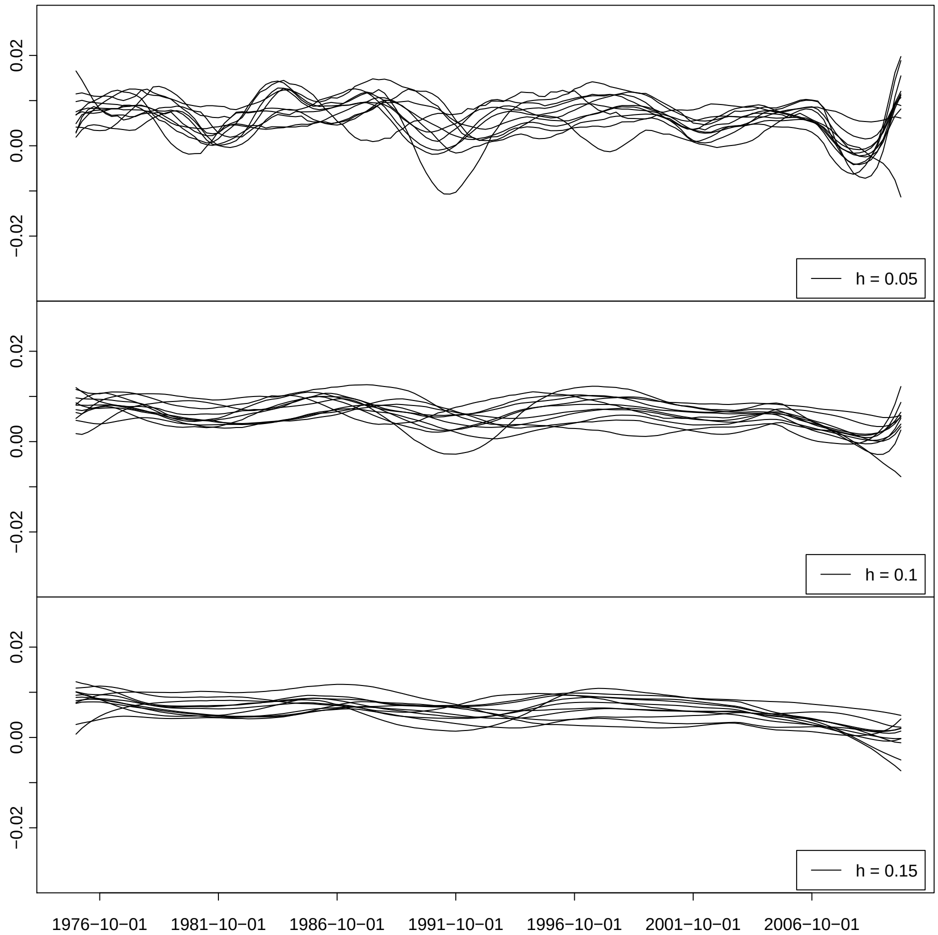

Figure 1: Local linear kernel estimates of the $n = 11$ original time trends from the application of Section 7.1. Each panel shows the estimates for a different bandwidth h.

used by the US Census Bureau. We repeat the analysis using not seasonally adjusted data for all of the 11 countries as a robustness check and we report the results in the Appendix.

For all of the countries, the observations are given in thousands of persons.

4. **Human capital:** We use Educational Attainment for Population Aged 25 and Over (H) collected from http://www.barrolee.com (accessed on 7 December 2021) as a measure of human capital. Since the only available data is five-year census data, we follow Zhang et al. (2012) and use linear interpolation between the observations and constant extrapolation on the boundaries (second and third quarters of 2010) to obtain the quarterly time series.

We thus observe a panel of $n = 11$ time series $\mathcal{T}_i = \{(Y_{it}, \boldsymbol{X}_{it}) : 1 \leq t \leq T\}$ of length $T = 140$ for each country $i \in \{1, \ldots, 11\}$, where $Y_{it} = \Delta \ln GDP_{it} := \ln GDP_{it} - \ln GDP_{i(t-1)}$, $\boldsymbol{X}_{it} = (\Delta \ln L_{it}, \Delta \ln K_{it}, \Delta \ln H_{it})^\top$ with $\Delta \ln L_{it} := \ln L_{it} - \ln L_{i(t-1)}$, $\Delta \ln K_{it} := \ln K_{it} - \ln K_{i(t-1)}$ and $\Delta \ln H_{it} := \ln H_{it} - \ln H_{i(t-1)}$. Without loss of gen-

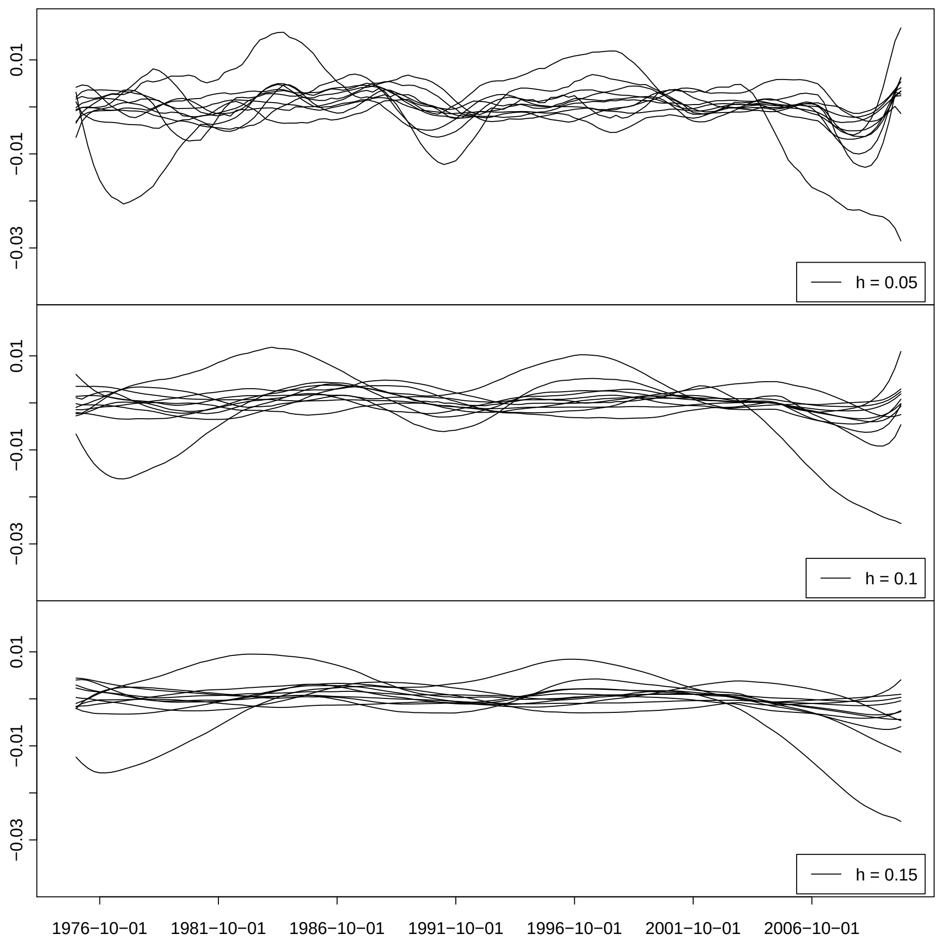

Figure 2: Local linear kernel estimates of the $n = 11$ augmented time trends from the application of Section 7.1. Each panel shows the estimates for a different bandwidth h.

erality, we let $\Delta \ln GDP_{i1} = \Delta \ln L_{i1} = \Delta \ln K_{i1} = \Delta \ln H_{i1} = 0$. The time series $\mathcal{T}_i$ is assumed to follow the model

$$Y_{it} = \boldsymbol{\beta}_i^\top \boldsymbol{X}_{it} + m_i\Big(\frac{t}{T}\Big) + \alpha_i + \varepsilon_{it} \tag{7.1}$$

for $1 \leq t \leq T$, where $\boldsymbol{\beta}_i = (\beta_{i,1}, \beta_{i,2}, \beta_{i,3})^\top$ is a vector of unknown parameters, m_i is a country-specific unknown nonparametric time trend, and α_i is a fixed-effect term. Similarly to Zhang et al. (2012), we rewrite the model (7.1) as

$$\Delta \ln GDP_{it} = \beta_{i,1} \Delta \ln L_{it} + \beta_{i,2} \Delta \ln K_{it} + \beta_{i,3} \Delta \ln H_{it} + m_i\Big(\frac{t}{T}\Big) + \alpha_i + \varepsilon_{it}, \tag{7.2}$$

for $i \in \{1, \dots, 11\}$ and $t \in \{1, \dots, 140\}$.

Our test procedure depends on the estimator of the long-run variance σ_i^2 for the error process $\mathcal{E}_i = \{\varepsilon_{it} : 1 \leq t \leq T\}$. As was mentioned in Section 3.1, without imposing any assumptions on the error terms, the estimator tends to be imprecise. To improve the behaviour of the estimator in a finite sample, we decide for imposing a certain time series structure on the error process in our analysis. Specifically, we assume that for each i the error process $\mathcal{E}_i$ follows an AR(p_i) model $\varepsilon_{it} = \sum_{j=1}^{p_i} a_{i,j}\varepsilon_{i(t-j)} + \eta_{it}$, where the order of the process p_i is country-specific and not known and η_{it} are i.i.d. innovations with mean zero. We choose p_i as the minimiser of the Bayesian Information Criterion (BIC).[3] This yields $p_i = 3$ for Australia, Canada and the UK, and $p_i = 1$ for all other countries. This assumption allows us to use a difference-based long-run variance estimator proposed in Khismatullina and Vogt (2020).

We aim to test whether the time trend m_i is the same for all 11 countries. In other words, we want to test the null hypothesis $H_0 : m_1 = \ldots = m_n$ with $n = 11$ in model (7.2). To do so, we implement the multiscale test from Section 3 in the following way.

1. We choose K to be an Epanechnikov kernel.

2. We let $\mathcal{G}_T = U_T \times H_T$ with

$$U_T = \Big\{u \in [0,1] : u = \tfrac{8t+1}{2T} \text{ for some } t \in \mathbb{N}\Big\}$$
$$H_T = \Big\{h \in \big[\tfrac{\log T}{T}, \tfrac{1}{4}\big] : h = \tfrac{4t}{T} \text{ for some } t \in \mathbb{N}\Big\}.$$

   We thus take into account all locations u on an equidistant grid U_T with step length $4/T$ and all bandwidths $h = 4/T, 8/T, 12/T, \ldots$ with $\log T/T \le h \le 1/4$. Note that $\mathcal{G}_T$ is not a subset of $\mathcal{G}_T^{\text{full}}$ since we do not assume that the considered locations take the form of $u = t/T$ for some $1 \le t \le T$. In our case, the choice of the grid U_T is motivated by the structure of the data: for each interval $\mathcal{I}_{(u,h)} = [u-h, u+h]$, the effective sample size is $8, 16, 24, \ldots$ quarters, i.e. $2, 4, 6, \ldots$ years. It can be shown that with this choice of the grid the theoretical properties of our method remain the same. Furthermore, the lower bound $\log T/T$ is explained by Assumption (C11) which requires that $\log T/T \ll h_{\min}$ (given that all moments of ε_{it} exist).

3. We estimate the unknown parameters $\boldsymbol{\beta}_i = (\beta_{i,1}, \beta_{i,2}, \beta_{i,3})^\top$ for each country i separately using the first-differencing approach described in Section 3.1.

4. We compute the estimator of the fixed-effect term α_i by employing (3.3). We then work with the augmented time series $\widehat{Y}_{it} = Y_{it} - \widehat{\boldsymbol{\beta}}_i^\top \boldsymbol{X}_{it} - \widehat{\alpha}_i$ instead of the original data on the GDP growth rate Y_{it}.

5. To obtain the estimator $\hat{\sigma}_i^2$ of the long-run error variance σ_i^2, for each i we apply the procedure from Khismatullina and Vogt (2020) to the augmented values $\{\widehat{Y}_{it} : 1 \le t \le T\}$. The details are summarised in Section **??** of the Appendix. We use $\hat{\sigma}_i^2$ for calculating the value of our main test statistic $\widehat{\Psi}_{n,T}$.

[3]We also calculate the values of other information criteria such as FPE, AIC and HQ which, in most of the cases, resulted in the same values of p_i.

6. To obtain the (approximate) critical values $q_{n,T}(\alpha)$ of the multiscale test, we simulate 5000 values of the statistic $\Phi_{n,T}$ defined in Section 3.3 and compute their empirical $(1-\alpha)$ quantile $\hat{q}_{n,T}(\alpha)$.

Figure 1 depicts smoothed version of the original time series on the GDP growth rate $\{Y_{it} = \Delta \ln GDP_{it} : 1 \leq t \leq T\}$ for each country i of the $n = 11$ countries under consideration. Figure 2 presents local linear estimates of the trend functions m_i for these countries after factoring out the effects of the covariates and the fixed-effect terms (i.e. calculated from the augmented time series $\widehat{Y}_{it}$). In both figures, each panel corresponds to a different value of the bandwidth h.

As can be seen in Figure 1, in the original data on the GDP growth there are some notable differences between the countries for bandwidths $h = 0.05$ and $h = 0.1$. For example, while most of the countries experience the increase in the growth rate in the last two years, the data for one of the countries (Norway) suggests a decrease in the same period of time. Moreover, using the smallest bandwidth ($h = 0.05$) allows us to notice considerable differences in the behaviour of the time series in the middle of our time region. In contrast, the value of the bandwidth $h = 0.15$ is too big to detect any heterogeneity in the behaviour of the trends. As we see, the choice of the bandwidth is crucial in making conclusions in this example.

Now we look at the local linear estimates of the trend functions m_i after excluding the effects of the covariates (Figure 2). We can see that the differences between the trends are much more pronounced and some heterogeneity between the countries is notable even for large values of h.

The results of applying our test are in line with our conclusions from the visual inspection: our test rejects the null hypothesis H_0 at the most common levels $\alpha = 0.01, 0.05, 0.1$. This result is consistent with the findings in Zhang et al. (2012) where the authors report a rejection of the null hypothesis of a common trend at the level $\alpha = 0.1$. However, in contrast to Zhang et al. (2012), our method provides a more detailed comparison of the trends in the GDP growth rate in these 11 countries. Specifically, we can make simultaneous confidence statements about which of the countries have different trends and where they differ. With the help of our multiscale method, we simultaneously test all local null hypotheses $H_0^{[i,j]}(u,h)$ that $m_i = m_j$ on the interval $\mathcal{I}_{(u,h)} = [u-h, u+h]$ for each $i, j, 1 \leq i < j \leq n$, and each $(u,h) \in \mathcal{G}_T$. The results for $\alpha = 0.05$[4] are presented in Figures 3–13, with each figure corresponding to a specific pair of countries (i,j) from our sample.[5] Each figure is divided into three panels (a)–(c). Panel (a) shows the augmented time series $\{\widehat{Y}_{it} : 1 \leq t \leq T\}$ and $\{\widehat{Y}_{jt} : 1 \leq t \leq T\}$ for the two countries i and j that are being compared. Panel (b) presents smoothed versions of the time series from (a), that is, it shows nonparametric kernel estimates

[4]The results for $\alpha = 0.1$ are similar and thus ommitted.

[5]We present the results only for those pairs (i,j) where the set $\mathcal{S}^{[i,j]}(\alpha)$ is non-empty.

(specifically, Nadaraya-Watson estimates) of the two trend functions m_i and m_j, where the bandwidth is set to 14 quarters (which corresponds to $h = 0.1$) and a rectangular kernel is used. Finally, panel (c) presents the results produced by our test for a significance level $\alpha = 0.05$: it depicts in grey the set $\mathcal{S}^{[i,j]}(\alpha)$ of all the intervals for which the test rejects the local null $H_0^{[i,j]}(u,h)$. The set of minimal intervals $\mathcal{S}^{[i,j]}_{min}(\alpha) \subseteq \mathcal{S}^{[i,j]}(\alpha)$ is highlighted in black. Note that according to (4.6), we can make the following simultaneous confidence statement about the intervals plotted in panels (c) of Figures 3–13: we can claim, with confidence of about 95%, that there is a difference between

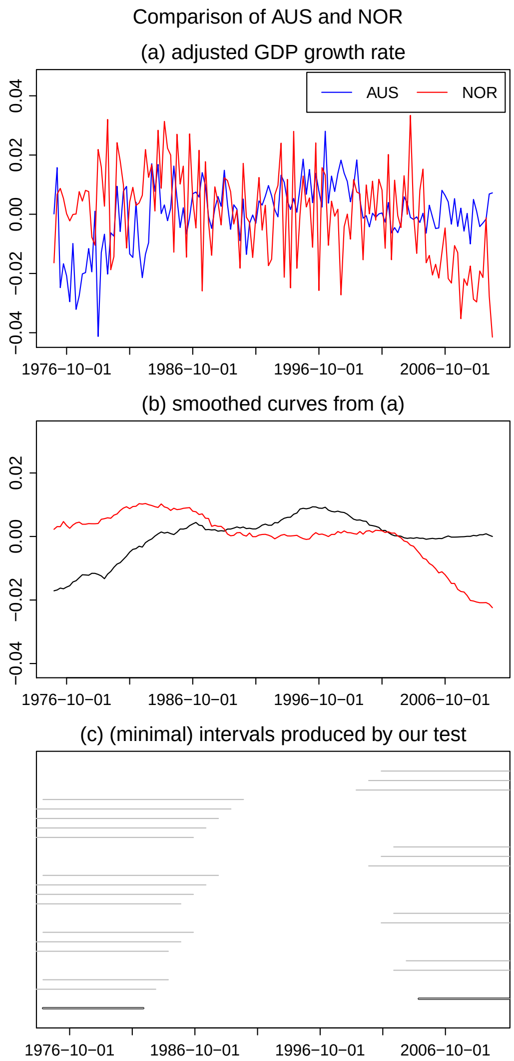

Figure 3: Test results for the comparison of Australia and Norway.

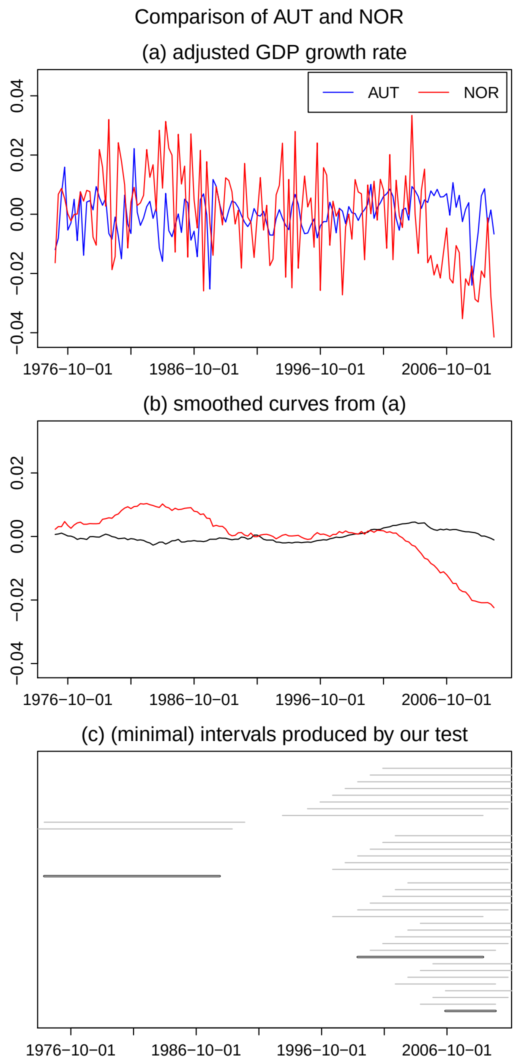

Figure 4: Test results for the comparison of Austria and Norway.

Note: In each figure, panel (a) shows the two augmented time series, panel (b) presents smoothed versions of the augmented time series, and panel (c) depicts the set of intervals $\mathcal{S}^{[i,j]}(\alpha)$ in grey and the subset of minimal intervals $\mathcal{S}^{[i,j]}_{min}(\alpha)$ in black.

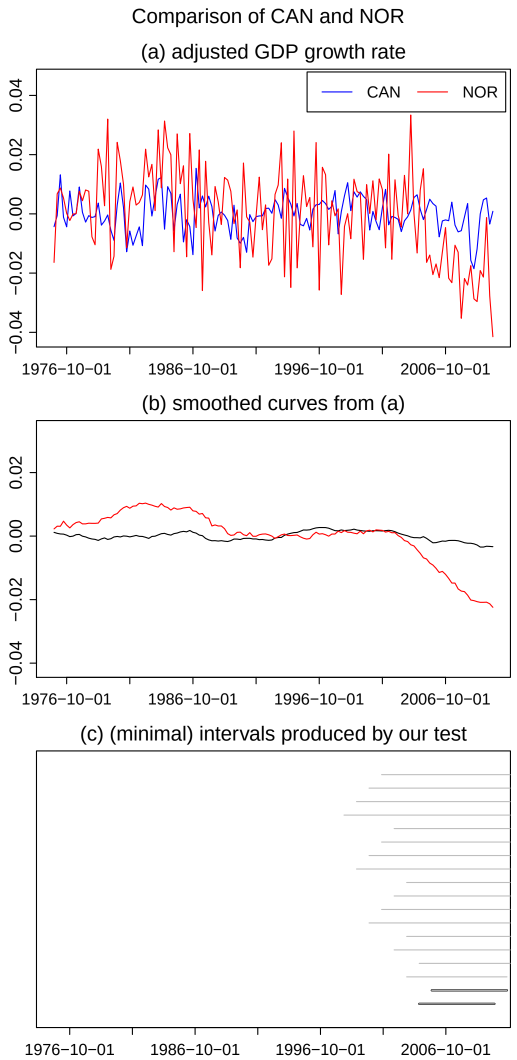

Figure 5: Test results for the comparison of Canada and Norway.

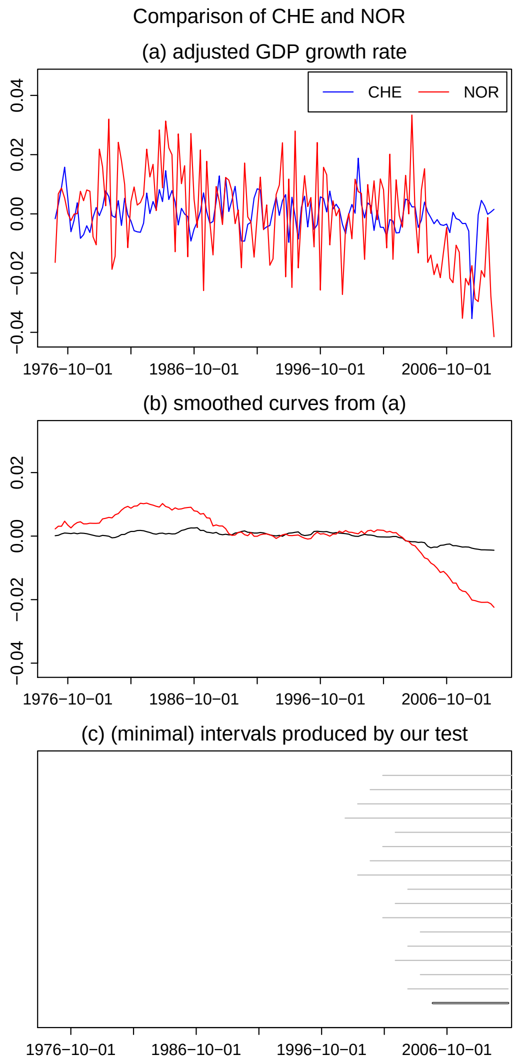

Figure 6: Test results for the comparison of Switzerland and Norway.

Note: In each figure, panel (a) shows the two augmented time series, panel (b) presents smoothed versions of the augmented time series, and panel (c) depicts the set of intervals $\mathcal{S}^{[i,j]}(\alpha)$ in grey and the subset of minimal intervals $\mathcal{S}^{[i,j]}_{min}(\alpha)$ in black.

the functions m_i and m_j on each of these intervals.

Now we briefly comment on the results. Out of 55 pairwise comparisons, our test detects the differences in the trends 11 times. In 9 cases, one of the countries involved in the comparison is Norway; these comparisons are depicted in Figures 3 – 11. Note that the differences in the trends seem to be more pronounced towards the end of the considered period: the set of intervals $\mathcal{S}^{[i,j]}(\alpha)$ for those pairs where one of the countries involved in comparison is Norway always covers last 10 years of the analysed time period (from the first quarter in 2000 up to the third quarter in 2010). This coincides with the

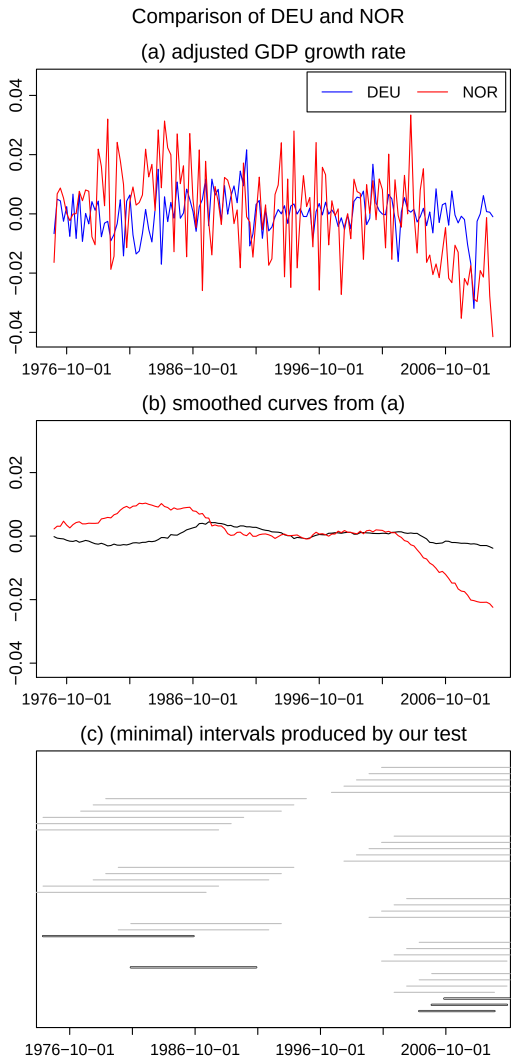

Figure 7: Test results for the comparison of Germany and Norway.

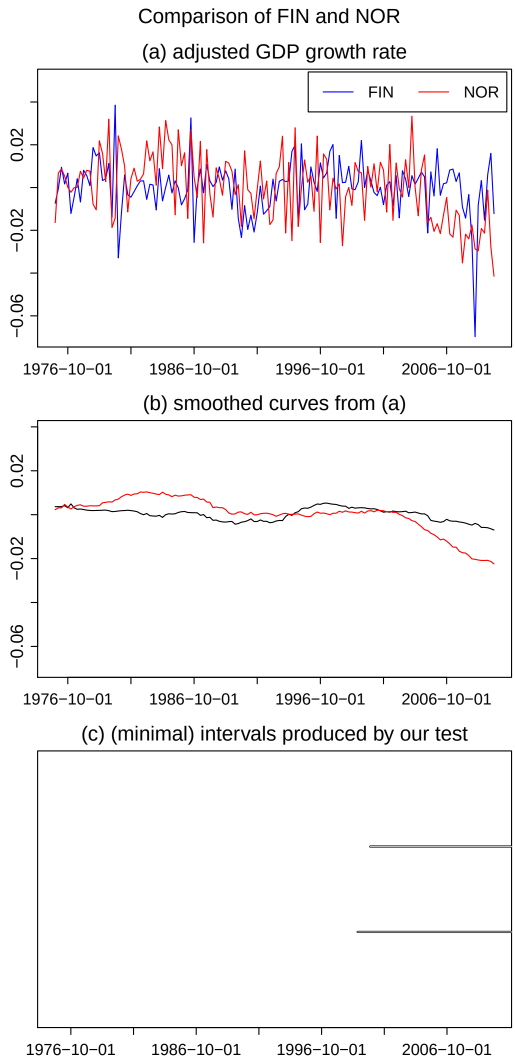

Figure 8: Test results for the comparison of Finland and Norway.

Note: In each figure, panel (a) shows the two augmented time series, panel (b) presents smoothed versions of the augmented time series, and panel (c) depicts the set of intervals $\mathcal{S}^{[i,j]}(\alpha)$ in grey and the subset of minimal intervals $\mathcal{S}^{[i,j]}_{min}(\alpha)$ in black.

visual observation mentioned above that in the last 10 years only Norway exhibited downward movement of the trend in the GDP growth, and this tendency becomes even more pronounced after factoring out the effect of the individual covariates. All of the other countries appear to have a slight increase in the trend function in the same time period. Our test supports this observation.

Figures 12 and 13 present the results of the pairwise comparison between Australia and France and between the USA and France respectively. In both cases, our test detects the differences between the underlying trends in the GDP growth only in the

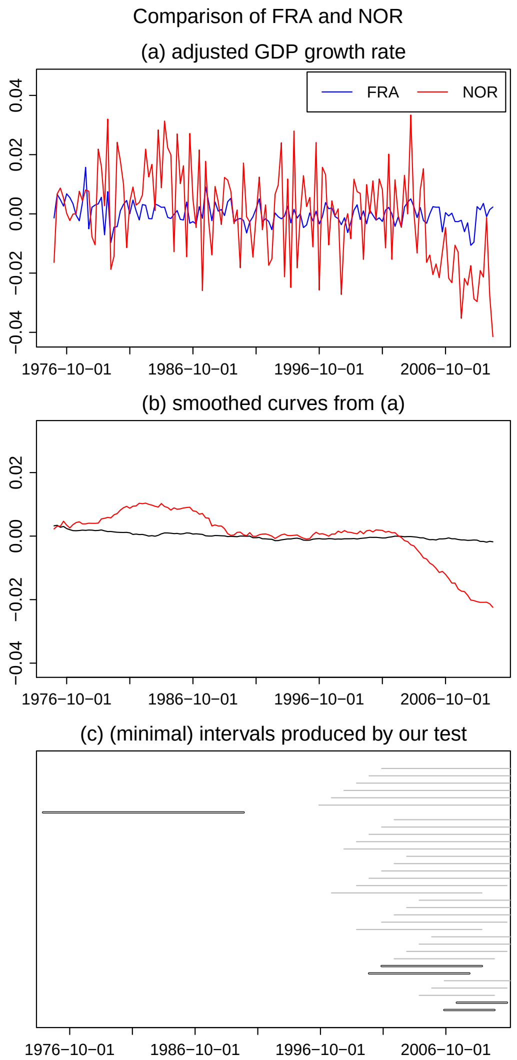

Figure 9: Test results for the comparison of France and Norway.

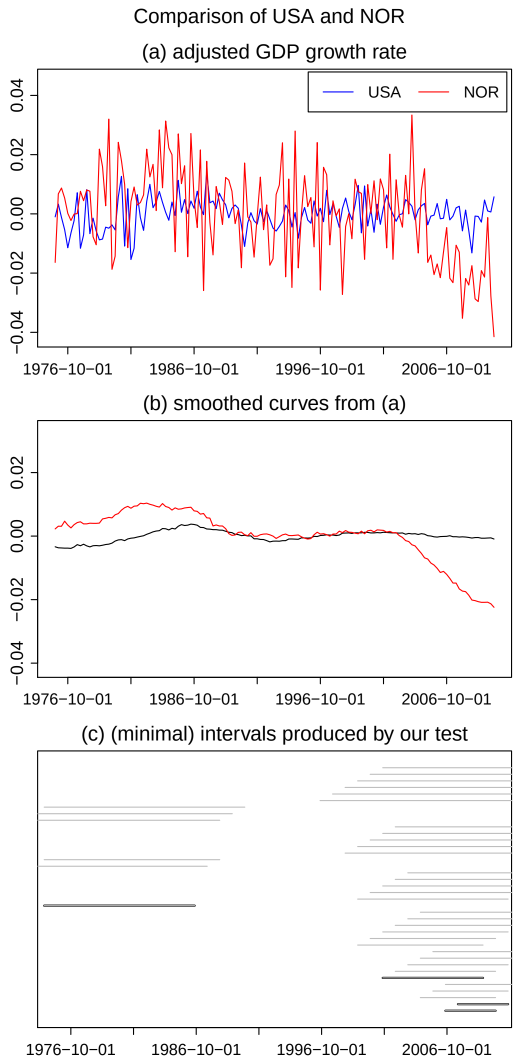

Figure 10: Test results for the comparison of the USA and Norway.

Note: In each figure, panel (a) shows the two augmented time series, panel (b) presents smoothed versions of the augmented time series, and panel (c) depicts the set of intervals $\mathcal{S}^{[i,j]}(\alpha)$ in grey and the subset of minimal intervals $\mathcal{S}^{[i,j]}_{min}(\alpha)$ in black.

beginning of the considered time period. In the case of Australia and France, the systematic difference between the trends is clearly visible even in the raw data (panel (a) in Figure 13), whereas in the case of USA and France, the difference is not so obvious. Our test allows us to detect differences in both situations, and we can claim with confidence no less than 95%, that there are significant differences between the trends for the USA and France (for Australia and France) up to the fourth quarter in 1991 (the fourth quarter in 1987), but there is no evidence of any differences between the trends from 1992 (1988) onwards.

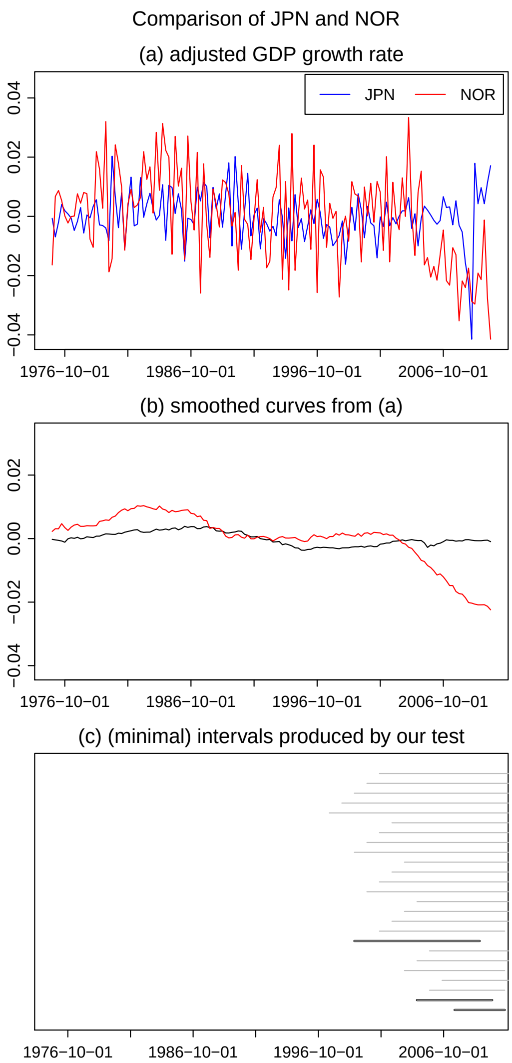

Figure 11: Test results for the comparison of Japan and Norway.

Note: Panel (a) shows the two augmented time series, panel (b) presents smoothed versions of the augmented time series, and panel (c) depicts the set of intervals $\mathcal{S}^{[i,j]}(\alpha)$ in grey and the subset of minimal intervals $\mathcal{S}^{[i,j]}_{min}(\alpha)$ in black.

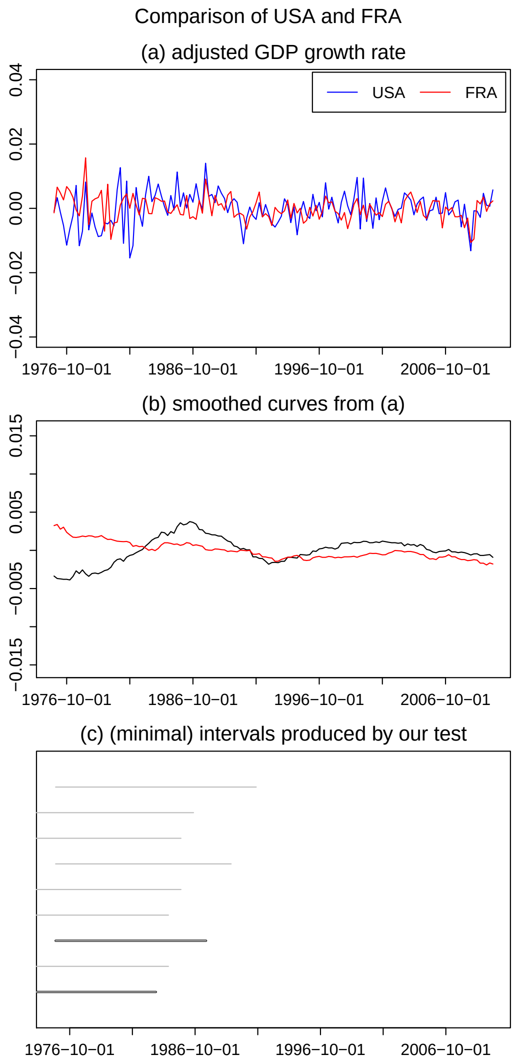

Figure 12: Test results for the comparison of the USA and France.

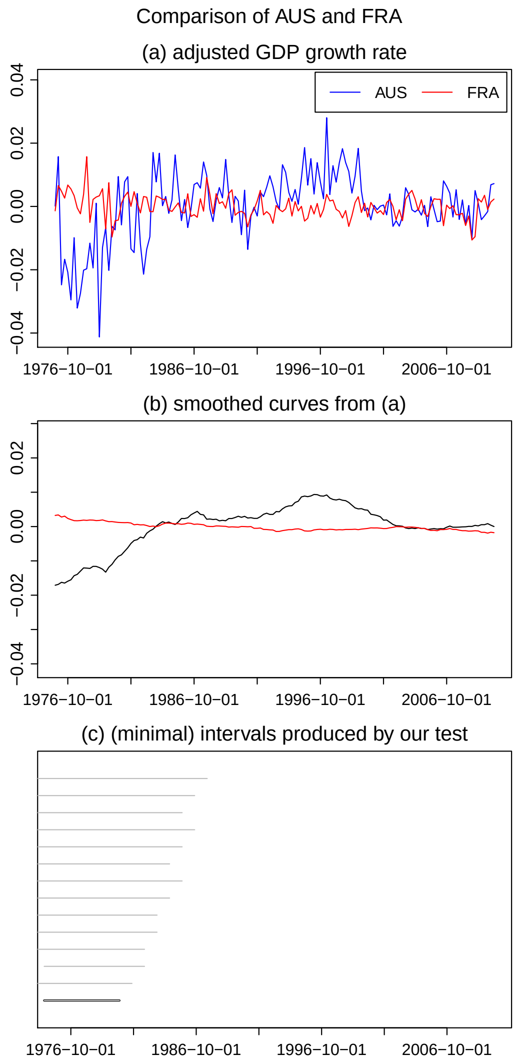

Figure 13: Test results for the comparison of Australia and France.

Note: In each figure, panel (a) shows the two augmented time series, panel (b) presents smoothed versions of the augmented time series, and panel (c) depicts the set of intervals $\mathcal{S}^{[i,j]}(\alpha)$ in grey and the subset of minimal intervals $\mathcal{S}^{[i,j]}_{min}(\alpha)$ in black.

7.2 Analysis of the house prices

We next analyse the historical dataset on nominal annual house prices from Knoll et al. (2017) that contains the data for 14 advanced economies that covers 143 years from 1870 to 2012. To ensure minimal manipulation with the data for all time series, we consider observations over the period 1890 – 2012 for 8 countries[6]. Specifically, the data on house prices for all of the countries in our analysis except one (Belgium) contains no missing values, and for Belgium there are only five missing observations[7] which we impute using linear interpolation from the nearest years. Each of the time series for other 6 countries present in the dataset from Knoll et al. (2017) contains more than 10 missing values, and hence these countries are excluded from our analysis.

We deflate the nominal house prices with the corresponding consumer price index (CPI) to obtain real house prices (HP). Variables that potentially can influence the average house prices are numerous, and there seems to be no general consensus about what the main causes of fluctuations are. Possible determinants include, but are not limited to, demographic factors such as population growth (Holly et al. 2010, Wang and Zhang 2014, Churchill et al. 2021); fundamental economic factors such as real GDP (Huang et al. 2013, Churchill et al. 2021), interest rate and inflation (Abelson et al. 2005, Otto 2007, Huang et al. 2013, Jordà et al. 2015); urbanisation (Chen et al. 2011, Wang et al. 2017); government subsidies and regulations (Malpezzi 1999); stock markets (Gallin 2006); etc. In our analysis, we focus on the following determinants of the average house prices: real GDP (GDP), population size (POP), long-term interest rate (I) and inflation ($INFL$) which is measured as change in CPI. Most other factors (such as government regulations, construction costs, and real wages) vary rather slowly over time and can be captured by time trend, fixed effects and slope heterogeneity.

Data for CPI, real GDP, population size and long-term interest rate are taken from the Jordà-Schularick-Taylor Macrohistory Database,[8] which is freely available at `http://www.macrohistory.net/data/` (accessed on 13 January 2022).

We thus observe a panel of $n = 8$ time series $\mathcal{T}_i = \{(Y_{it}, \boldsymbol{X}_{it}) : 1 \leq t \leq T\}$ of length $T = 123$ for each country $i \in \{1, \dots, 8\}$, where $Y_{it} = \ln HP_{it}$ and $\boldsymbol{X}_{it} = (\ln GDP_{it}, \ln POP_{it}, I_{it}, INFL_{it})^\top$. The time series $\mathcal{T}_i$ is assumed to follow the model

$$Y_{it} = \boldsymbol{\beta}_i^\top \boldsymbol{X}_{it} + m_i\Big(\frac{t}{T}\Big) + \alpha_i + \varepsilon_{it} \tag{7.3}$$

for $1 \leq t \leq T$, where $\boldsymbol{\beta}_i = (\beta_{i,1}, \beta_{i,2}, \beta_{i,3}, \beta_{i,4})^\top$ is a vector of unknown parameters, m_i is a country-specific unknown nonparametric time trend, and α_i is a fixed-effect term.

[6] Australia, Belgium, Denmark, France, the Netherlands, Norway, Sweden and the USA.

[7] The missing values in the time series on the house prices for Belgium span five years during the World War I.

[8] See Jordà et al. (2017) for the detailed description of the variable construction.

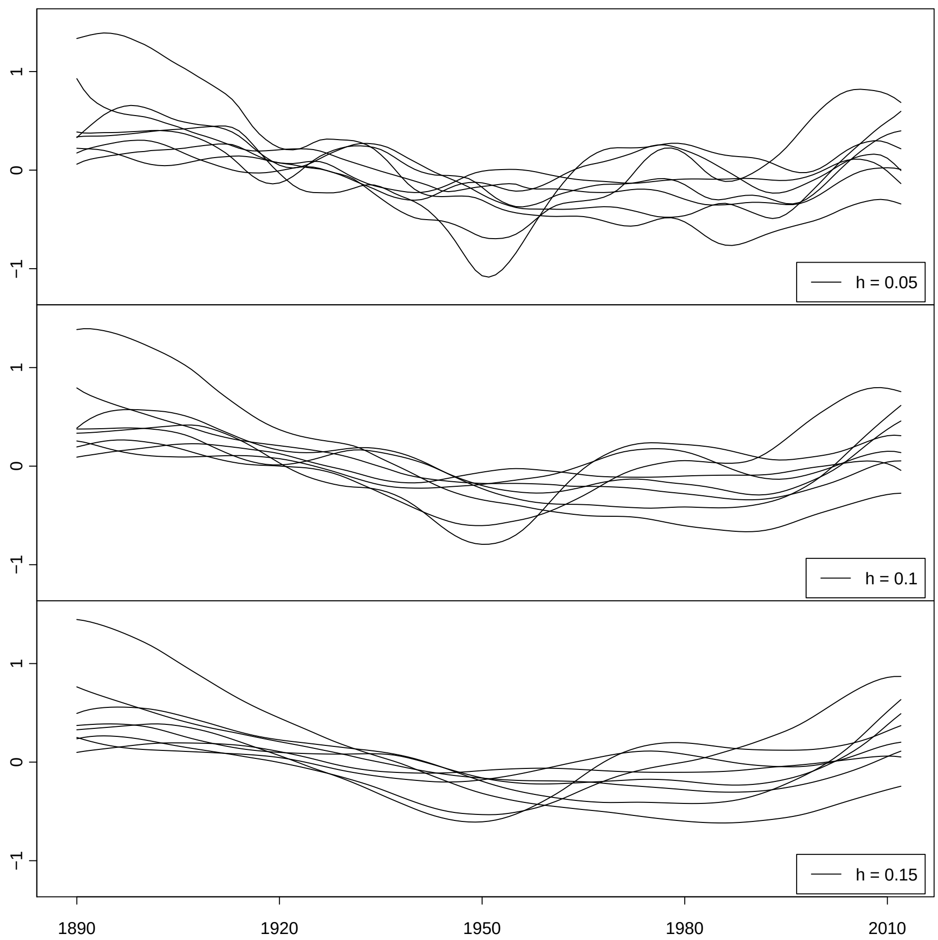

Figure 14: Local linear kernel estimates of the $n = 8$ augmented time trends from the application of Section 7.2. Each panel shows the estimates for a different bandwidth h.

As in Section 7.1, we rewrite the model (7.3) as

$$\ln HP_{it} = \beta_{i,1} \ln GDP_{it} + \beta_{i,2} \ln POP_{it} + \beta_{i,3} I_{it} + \beta_{i,4} INFL_{it} + m_i\Big(\frac{t}{T}\Big) + \alpha_i + \varepsilon_{it}. \quad (7.4)$$

The motivation for including nonparametric trend function m_i in the model 7.4 is supported by the literature. For example, Ugarte et al. (2009) model the temporal trends in average Spanish house prices using splines. Winter et al. (2022) include a long-term trend stochastic component when describing the dynamic behaviour of the time series of real house prices in eight advanced economies. Including nonparametric trend function while modelling the evolution of the house prices is also the main conclusion in Zhang et al. (2016), where the authors show that the time series for the logarithmic US house prices is trend-stationary, i.e. it can be transformed into a stationary series by subtracting the deterministic trend.

As in Section 7.1, we assume that for each i the error process $\mathcal{E}_i = \{\varepsilon_{it} : 1 \le t \le T\}$ follows an AR(p_i) model $\varepsilon_{it} = \sum_{j=1}^{p_i} a_{i,j} \varepsilon_{i(t-j)} + \eta_{it}$ with η_{it} being the innovations that are i.i.d. across t and have zero mean. As before, we choose p_i as a minimiser of BIC.

For 7 out of 8 countries the order p_i determined by BIC[9] is equal to 1 and, for the sake of simplicity, we assume that $p_i = 1$ for all $i = 1, \ldots, 8$.

Our goal is to test whether the time trend m_i is the same in all 8 countries. In other words, we would like to test the null hypothesis $H_0 : m_1 = \ldots = m_n$ with $n = 8$ in the model (7.4). To do so, we implement the multiscale test from Section 3 in the

[9] Applying other information criteria such as FPE, AIC and HQ yields exactly the same results in these cases.

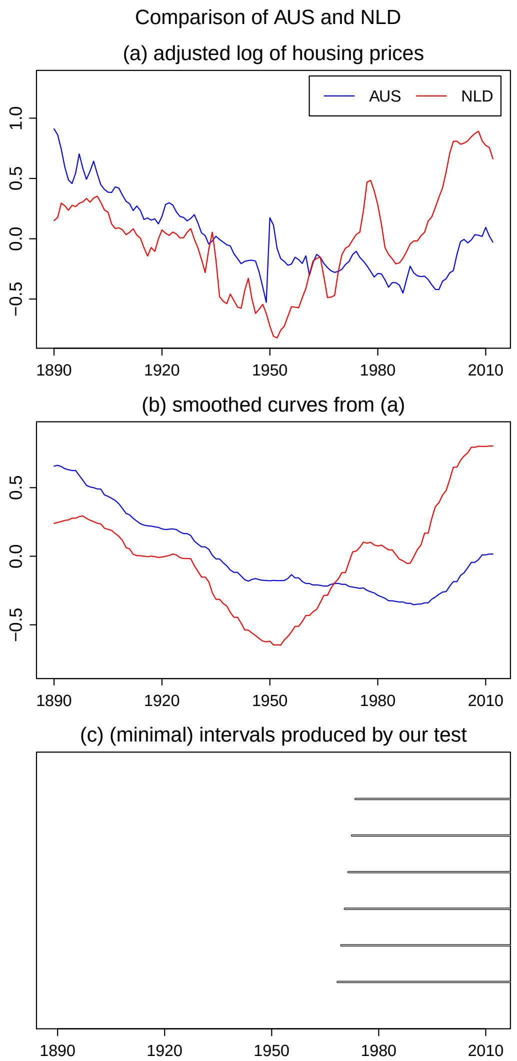

Figure 15: Test results for the comparison of the housing prices in Australia and the Netherlands.

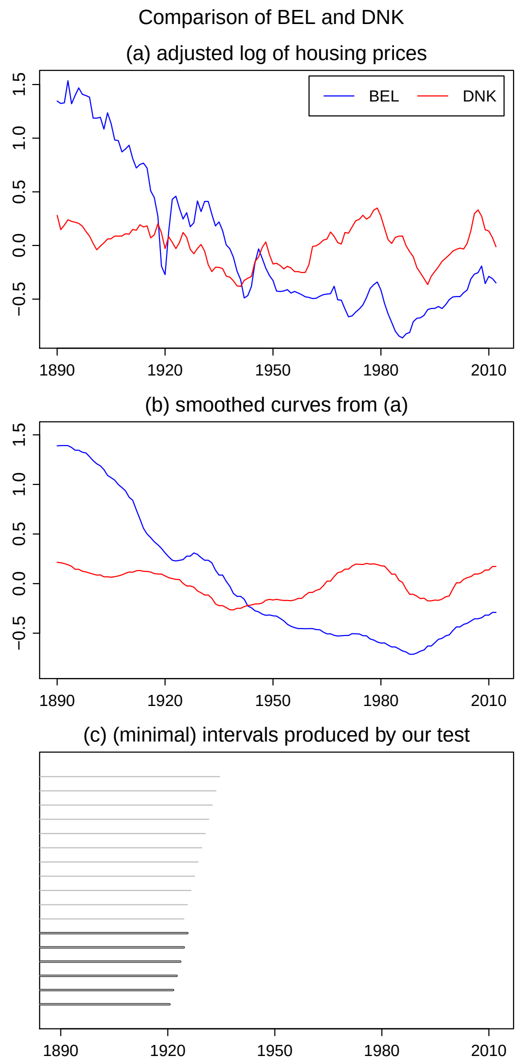

Figure 16: Test results for the comparison of the housing prices in Belgium and Denmark.

Note: In each figure, panel (a) shows the two augmented time series of the house prices, panel (b) presents smoothed versions of the augmented time series, and panel (c) depicts the set of intervals $\mathcal{S}^{[i,j]}(\alpha)$ in grey and the subset of minimal intervals $\mathcal{S}^{[i,j]}_{min}(\alpha)$ in black.

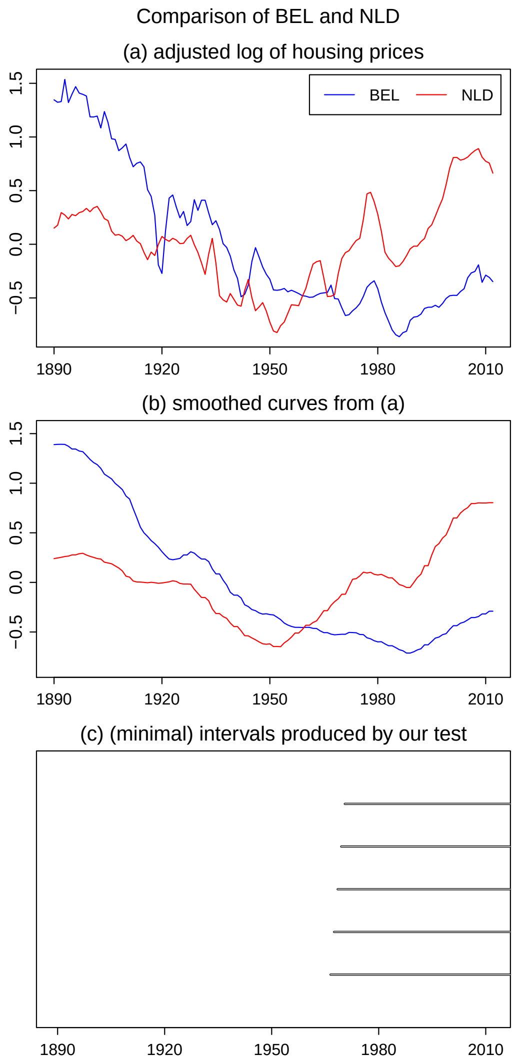

Figure 17: Test results for the comparison of the housing prices in Belgium and the Netherlands.

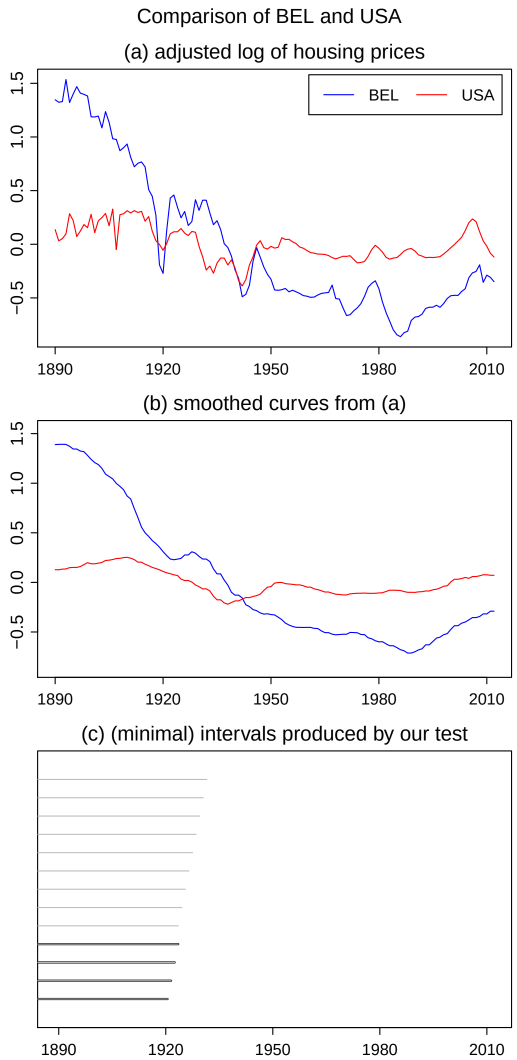

Figure 18: Test results for the comparison of the housing prices in Belgium and the USA.

Note: In each figure, panel (a) shows the two augmented time series of the house prices, panel (b) presents smoothed versions of the augmented time series, and panel (c) depicts the set of intervals $\mathcal{S}^{[i,j]}(\alpha)$ in grey and the subset of minimal intervals $\mathcal{S}^{[i,j]}_{min}(\alpha)$ in black.

same way as in the previous application example with one minor modification: we let $\mathcal{G}_T = U_T \times H_T$ with

$$U_T = \big\{ u \in [0,1] : u = \tfrac{t}{T} \text{ for some } t \in \mathbb{N} \big\}$$
$$H_T = \big\{ h \in \big[\tfrac{\log T}{T}, \tfrac{1}{4} \big] : h = \tfrac{5t-3}{T} \text{ for some } t \in \mathbb{N} \big\}.$$

We thus take into account all locations u on an equidistant grid U_T with step length $1/T$ and all bandwidths $h = 2/T, 7/T, 12/T, \dots$ with $\log T/T \leq h \leq 1/4$. The effective sample sizes are therefore $5, 15, 25, \dots$ years. As before, the lower bound $\log T/T$ is motivated by Assumption (C11).

Apart from this modification, the multiscale test is implemented exactly in the same way as in the simulations of Section 6 and in the previous application example in Section 7.1.

We are now ready to apply the test procedure to the data. Figure 14 depicts local linear estimates of the trend functions m_i for the $n = 8$ different countries. Each panel corresponds to a different bandwidth h. As can be seen, for all of bandwidths h, the fits do not look very similar to each other. On the contrary, there seems to be one country (France) that shows a much more pronounced decrease in the time trend m_i in the first half of the analysed time period than all of the other countries. For the smallest bandwidth value $h = 0.05$, we also see that the behaviour of the trends around year 1950 is not homogeneous. Visual inspection thus suggests that there are strong differences between the time trends m_i. Our test confirms this impression. It rejects the null hypothesis at levels $\alpha = 0.05$ and $\alpha = 0.1$. Hence, the test provides evidence for a violation of the null.

The results of applying our test for level $\alpha = 0.05$ are presented in Figures 15 – 18. As in Section 7.1, each figure corresponds to one pairwise comparison (i, j) and is divided into three panels (a) – (c). The structure of the panels is the same as in Figures 3 – 13: Panel (a) shows the augmented time series $\{\widehat{Y}_{it} : 1 \le t \le T\}$ and $\{\widehat{Y}_{jt} : 1 \le t \le T\}$ for the two countries i and j. Panel (b) presents smoothed versions of the time series from (a) with the bandwidth window covering 15 years. Finally, panel (c) presents the results produced by our test for a significance level $\alpha = 0.05$. As before, the set of rejected intervals $\mathcal{S}^{[i,j]}(\alpha)$ and the set of minimal intervals $\mathcal{S}^{[i,j]}_{min}(\alpha) \subseteq \mathcal{S}^{[i,j]}(\alpha)$ are depicted in grey and in black respectively.

Our findings coincide with the observations in Knoll et al. (2017), where the authors find considerable cross-country heterogeneity in the trends in house prices. Moreover, according to (4.6), we can make the following simultaneous confidence statement about the intervals plotted in panels (c) of Figures 15 – 18: we can claim, with confidence of about 95%, that there is a difference between the functions m_i and m_j on each of these intervals.

8 Conclusion

In this paper, we develop a new multiscale testing procedure for multiple time series for testing hypotheses about nonparametric time trends in the presence of covariates. This procedure addresses two important statistical problems about comparison of the time trends. First and foremost, with the help of the proposed method, we are able to test if all the time trends in the observed time series are the same or not. We prove the main theoretical results of the paper that the test has (asymptotically) the correct size and has an (asymptotic) power of one against a specific class of local alternatives. Second, our multiscale procedure allows us to tell which of the time trends are different and where the differences are located. For the purpose of pinpointing the differences, we consider many local null hypotheses at the same time, each corresponding to only two time trends and a specific time interval. Our method allows us to test all of these hypotheses simultaneously controlling the family-wise error rate, i.e. the probability of wrongly rejecting at least one true null hypothesis (making at least one type I error), at a desired level α. This result allows us to make simultaneous confidence statements as follows:

> *We can state with (asymptotic) probability at least* $1 - \alpha$ *that for every pair of time series and every interval where our test rejects the local null, the trends of these time series differ at least somewhere on this particular interval.*

For the proof of the theoretical results, the main tools that are used are strong approximation theory developed in Berkes et al. (2014) and the anti-concentration bounds for Gaussian random vectors verified in Chernozhukov et al. (2015). The proof strategy that we employ in our paper has already been used in Khismatullina and Vogt (2020), however, in that paper the authors proposed a multiscale method for testing qualitative hypotheses only about one time series. Our method can be regarded as a generalised version of the test developed in Khismatullina and Vogt (2020) where we not only consider comparison between various time series, but also add the covariates to the model and propose an estimation procedure for the unknown parameters.
Regarding future research, this project suggests some interesting issues and topics for consideration. First, as was already mentioned, it should be possible to extend our theoretical results to the case where the number of time series slowly grows with the sample size. Second, the theory in this paper is developed under the assumption that the first differences of the covariates and of error terms are uncorrelated. This restriction limits possible applications of our method. Further insight can be gained by broadening the current work in these and possibly other directions.

References

Abelson, P., Joyeux, R., Milunovich, G. and Chung, D. (2005). Explaining house prices in australia: 1970–2003. *Economic record*, **81** S96–S103.

Atak, A., Linton, O. and Xiao, Z. (2011). A semiparametric panel model for unbalanced data with application to climate change in the United Kingdom. *Journal of Econometrics*, **164** 92–115.

Barro, R. J. and Lee, J. W. (2013). A new data set of educational attainment in the world, 1950–2010. *Journal of Development Economics*, **104** 184–198. URL `https://www.sciencedirect.com/science/article/pii/S0304387812000855`.

Berkes, I., Liu, W. and Wu, W. B. (2014). Komlós-Major-Tusnády approximation under dependence. *Annals of Probability*, **42** 794–817.

Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *The annals of statistics*, **14** 1171–1179.

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

Chen, J., Guo, F. and Wu, Y. (2011). One decade of urban housing reform in china: Urban housing price dynamics and the role of migration and urbanization, 1995–2005. *Habitat International*, **35** 1–8.

Chen, L. and Wu, W. B. (2018). Testing for trends in high-dimensional time series. *Forthcoming in Journal of the American Statistical Association*.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probability Theory and Related Fields*, **162** 47–70.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2017). Central limit theorems and bootstrap in high dimensions. *Annals of Probability*, **45** 2309–2352.

Christiansen, H. and Pigott, C. (1997). Long-term interest rates in globalised markets.

Churchill, S. A., Baako, K. T., Mintah, K. and Zhang, Q. (2021). Transport infrastructure and house prices in the long run. *Transport Policy*, **112** 1–12.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics*, **25** 1–37.

Degras, D., Xu, Z., Zhang, T. and Wu, W. B. (2012). Testing for parallelism among trends in multiple time series. *IEEE Transactions on Signal Processing*, **60** 1087–1097.

Delgado, M. A. (1993). Testing the equality of nonparametric regression curves. *Statistics & Probability Letters*, **17** 199–204.

Dümbgen, L. (2002). Application of local rank tests to nonparametric regression. *Journal of Nonparametric Statistics*, **14** 511–537.

GALLIN, J. (2006). The long-run relationship between house prices and income: evidence from local housing markets. *Real Estate Economics*, **34** 417–438.

GRIER, K. B. and TULLOCK, G. (1989). An empirical analysis of cross-national economic growth, 1951–1980. *Journal of Monetary Economics*, **24** 259–276.

HALL, P. and HART, J. D. (1990). Bootstrap test for difference between means in nonparametric regression. *Journal of the American Statistical Association*, **85** 1039–1049.

HÄRDLE, W. and MARRON, J. S. (1990). Semiparametric comparison of regression curves. *Annals of Statistics*, **18** 63–89.

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning*. New York, Springer.

HIDALGO, J. and LEE, J. (2014). A CUSUM test for common trends in large heterogeneous panels. In *Essays in Honor of Peter C. B. Phillips*. Emerald Group Publishing Limited, 303–345.

HOLLY, S., PESARAN, M. H. and YAMAGATA, T. (2010). A spatio-temporal model of house prices in the usa. *Journal of Econometrics*, **158** 160–173.

HUANG, M., LI, R. and WANG, S. (2013). Nonparametric mixture of regression models. *Journal of the American Statistical Association*, **108** 929–941.

JORDÀ, Ò., SCHULARICK, M. and TAYLOR, A. M. (2015). Betting the house. *Journal of International Economics*, **96** S2–S18.

JORDÀ, Ò., SCHULARICK, M. and TAYLOR, A. M. (2017). Macrofinancial history and the new business cycle facts. *NBER macroeconomics annual*, **31** 213–263.

KAROLY, D. J. and WU, Q. (2005). Detection of regional surface temperature trends. *Journal of Climate*, **18** 4337–4343.

KHISMATULLINA, M. and VOGT, M. (2020). Multiscale inference and long-run variance estimation in non-parametric regression with time series errors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **82** 5–37.

KHISMATULLINA, M. and VOGT, M. (2021). Nonparametric comparison of epidemic time trends: the case of covid-19. *Journal of econometrics*.

KIM, K. H. (2016). Inference of the trend in a partially linear model with locally stationary regressors. *Econometric Reviews*, **35** 1194–1220.

KNOLL, K., SCHULARICK, M. and STEGER, T. (2017). No price like home: Global house prices, 1870-2012. *American Economic Review*, **107** 331–53.

LEE, C.-C. (2005). Energy consumption and gdp in developing countries: a cointegrated panel analysis. *Energy economics*, **27** 415–427.

LEE, C.-H., HUANG, B.-N. ET AL. (2002). The relationship between exports and economic growth in east asian countries: A multivariate threshold autoregressive approach. *Journal* *of Economic Development*, **27** 45–68.

Li, D., Chen, J. and Gao, J. (2010). Nonparametric time-varying coefficient panel data models with fixed effects. *The Econometrics Journal*, **14** 387–408.

Malpezzi, S. (1999). A simple error correction model of house prices. *Journal of housing economics*, **8** 27–62.

Nazarov, F. (2003). On the maximal perimeter of a convex set in $\mathbb{R}^n$ with respect to a Gaussian measure. In *Geometric Aspects of Functional Analysis*. Springer, 169–187.

Otto, G. (2007). The growth of house prices in australian capital cities: What do economic fundamentals explain? *Australian Economic Review*, **40** 225–238.

Park, C., Vaughan, A., Hannig, J. and Kang, K.-H. (2009). SiZer analysis for the comparison of time series. *Journal of Statistical Planning and Inference*, **139** 3974–3988.

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

Robinson, P. M. (2012). Nonparametric trending regression with cross-sectional dependence. *Journal of Econometrics*, **169** 4–14.

Sax, C. and Eddelbuettel, D. (2018). Seasonal adjustment by X-13ARIMA-SEATS in R. *Journal of Statistical Software*, **87** 1–17.

Sharma, S. C. and Dhakal, D. (1994). Causal analyses between exports and economic growth in developing countries. *Applied Economics*, **26** 1145–1157.

Sun, Y. (2011). Robust trend inference with series variance estimator and testing-optimal smoothing parameter. *Journal of Econometrics*, **164** 345–366.

The Organisation for Economic Co-operation and Development (2021a). Main Economic Indicators - complete database. retrieved from FRED, Federal Reserve Bank of St. Louis; `https://fred.stlouisfed.org/`. Accessed on 07-12-2021.

The Organisation for Economic Co-operation and Development (2021b). OECD.stat. `http://stats.oecd.org/Index.aspx`. Accessed on 07-12-2021.

Ugarte, M. D., Goicoa, T., Militino, A. F. and Durbán, M. (2009). Spline smoothing in small area trend estimation and forecasting. *Computational statistics & data analysis*, **53(10)** 3616–3629.

Vogelsang, T. J. and Franses, P. H. (2005). Testing for common deterministic trend slopes. *Journal of Econometrics*, **126** 1–24.

Vogt, M. and Linton, O. (2014). Nonparametric estimation of a periodic sequence in the presence of a smooth trend. *Biometrika*, **101** 121–140.

Vogt, M. and Linton, O. (2017). Classification of non-parametric regression functions in longitudinal data models. *Journal of the Royal Statistical Society: Series B*, **79** 5–27.

Vogt, M. and Linton, O. (2020). Multiscale clustering of nonparametric regression curves.

Journal of Econometrics, **216** 305–325.

Wang, X.-R., Hui, E. C.-M. and Sun, J.-X. (2017). Population migration, urbanization and housing prices: Evidence from the cities in china. *Habitat International*, **66** 49–56.

Wang, Z. and Zhang, Q. (2014). Fundamental factors in the housing markets of china. *Journal of housing economics*, **25** 53–61.

Winter, J. d., Koopman, S. J. and Hindrayanto, I. (2022). Joint decomposition of business and financial cycles: evidence from eight advanced economies. *Oxford Bulletin of Economics and Statistics*, **84** 57–79.

Wu, W. B. (2005). Nonlinear system theory: another look at dependence. *Proc. Natn. Acad. Sci. USA*, **102** 14150–14154.

Wu, W.-B. and Wu, Y. N. (2016). Performance bounds for parameter estimates of high-dimensional linear models with correlated errors. *Electronic Journal of Statistics*, **10** 352–379.

Wu, W. B. and Zhao, Z. (2007). Inference of trends in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **69** 391–410.

Xu, K.-L. (2012). Robustifying multivariate trend tests to nonstationary volatility. *Journal of Econometrics*, **169** 147–154.

Zhang, J., de Jong, R. and Haurin, D. (2016). Are us real house prices stationary? new evidence from univariate and panel data. *Studies in Nonlinear Dynamics & Econometrics*, **20** 1–18.

Zhang, Y., Su, L. and Phillips, P. C. (2012). Testing for common trends in semi-parametric panel data models with fixed effects. *The Econometrics Journal*, **15** 56–100.

A Appendix

In what follows, we provide detailed proofs of the theoretical results from Section 4. 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 $a \vee b = \max\{a, b\}$. For $x \in \mathbb{R}_{\geq 0}$, we let $\lfloor x \rfloor$ denote the integer value of x and $\lceil x \rceil$ the smallest integer greater than or equal to x. For any set A, the symbol $|A|$ denotes the cardinality of A. The expression $X \overset{\mathcal{D}}{=} Y$ means that the two random variables X and Y have the same distribution. Finally, we sometimes use the notation $a_T \ll b_T$ to express that $a_T = o(b_T)$.

Auxiliary results

Let $\{Z_t\}_{t=-\infty}^{\infty}$ be a stationary time series process with $Z_t \in \mathcal{L}^q$ for some $q > 2$ and $\mathbb{E}[Z_t] = 0$. Assume that Z_t can be represented as $Z_t = g(\ldots, \eta_{t-1}, \eta_t)$, where η_t are i.i.d. variables and $g : \mathbb{R}^\infty \to \mathbb{R}$ is a measurable function. We first state a Nagaev-type inequality from Wu and Wu (2016).

Definition A.1. *Let $q > 0$ and $\alpha > 0$. The dependence adjusted norm of the process $Z. = \{Z_t\}_{t=-\infty}^{\infty}$ is given by $\|Z.\|_{q,\alpha} = \sup_{t \geq 0}(t+1)^\alpha \sum_{s=t}^{\infty} \delta_q(g, s)$.*

Proposition A.2 (Wu and Wu (2016), Theorem 2)**.** *Assume that $\|Z.\|_{q,\alpha} < \infty$ with $q > 2$ and $\alpha > 1/2 - 1/q$. Let $S_T = a_1 Z_1 + \ldots + a_T Z_T$, where $a_1, \ldots, a_T$ are real numbers with $\sum_{t=1}^{T} a_t^2 = T$. Then for any $w > 0$,*

$$\mathbb{P}(|S_T| \geq w) \leq C_1 \frac{|a|_q^q \|Z.\|_{q,\alpha}^q}{w^q} + C_2 \exp\left(-\frac{C_3 w^2}{T\|Z.\|_{2,\alpha}^2}\right),$$

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

The following lemma is a simple consequence of the above inequality.

Lemma A.3. *Let $\sum_{s=t}^{\infty} \delta_q(g, s) = O(t^{-\alpha})$ for some $q > 2$ and $\alpha > 1/2 - 1/q$. Then*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_t = O_p(1).$$

Proof of Lemma A.3. Let $\eta > 0$ be a fixed number. We apply Proposition A.2 to the sum $S_T = \sum_{t=1}^{T} a_t Z_t$ with $a_t = 1$ for all t. The assumption $\sum_{s=t}^{\infty} \delta_q(g, s) = O(t^{-\alpha})$ implies that $\|Z.\|_{2,\alpha} \leq \|Z.\|_{q,\alpha} \leq C_Z < \infty$. Hence, for w chosen sufficiently large, we obtain that

$$\begin{aligned}
\mathbb{P}\Big(\Big|\sum_{t=1}^{T} Z_t\Big| \geq \sqrt{T} w\Big) &\leq C_1 \frac{T C_Z^q}{T^{q/2} w^q} + C_2 \exp\left(-\frac{C_3 T w^2}{T C_Z^2}\right) \\
&= \frac{\{C_1 C_Z^q\} T^{1-q/2}}{w^q} + C_2 \exp\left(-\frac{C_3 w^2}{C_Z^2}\right) \leq \eta
\end{aligned}$$

for all T. This means that $\sum_{t=1}^{T} Z_t / \sqrt{T} = O_p(1)$. $\square$

Let $\Delta\varepsilon_{it} = \varepsilon_{it} - \varepsilon_{it-1}$ and $\Delta\boldsymbol{X}_{it} = \boldsymbol{X}_{it} - \boldsymbol{X}_{it-1}$. By Assumptions (C1) and (C4), $\Delta\varepsilon_{it} = \Delta g_i(\mathcal{F}_{it})$ and $\Delta\boldsymbol{X}_{it} = \Delta\boldsymbol{h}_i(\mathcal{G}_{it})$. We further define

$$\begin{aligned}
\boldsymbol{a}_i(\mathcal{H}_{it}) &:= \Delta\boldsymbol{h}_i(\mathcal{G}_{it})\Delta g_i(\mathcal{F}_{it}) \quad = \Delta\boldsymbol{X}_{it}\Delta\varepsilon_{it} \\
\boldsymbol{b}_i(\mathcal{G}_{it}) &:= \Delta\boldsymbol{h}_i(\mathcal{G}_{it})\Delta\boldsymbol{h}_i(\mathcal{G}_{it})^\top = \Delta\boldsymbol{X}_{it}\Delta\boldsymbol{X}_{it}^\top,
\end{aligned}$$

where $\boldsymbol{a}_i = (a_{ij})_{j=1}^d$, $\boldsymbol{b}_i = (b_{ikl})_{k,l=1}^d$ and $\mathcal{H}_{it} = (\mathcal{H}_{it,1}, \dots, \mathcal{H}_{it,d})^\top$ with $\mathcal{H}_{it,j} = (\dots, \nu_{it-1,j}, \nu_{it,j})$ and $\nu_{it,j} = (\eta_{it,j}, \xi_{it,j})$. The next result gives bounds on the physical dependence measures of the processes $\{\boldsymbol{a}_i(\mathcal{H}_{it})\}_{t=-\infty}^{\infty}$ and $\{\boldsymbol{b}_i(\mathcal{G}_{it})\}_{t=-\infty}^{\infty}$.

Lemma A.4. *Let Assumptions (C1), (C3), (C4) and (C6) be satisfied. Then for each i, j, k and l, it holds that*

$$\begin{aligned}
&\sum_{s=t}^{\infty} \delta_p(a_{ij}, s) = O(t^{-\alpha}) \qquad \textit{for } p = \min\{q, q'\}/2 \textit{ and some } \alpha > 1/2 - 1/p \\
&\sum_{s=t}^{\infty} \delta_p(b_{ikl}, s) = O(t^{-\alpha}) \qquad \textit{for } p = q'/2 \textit{ and some } \alpha > 1/2 - 1/p.
\end{aligned}$$

Proof of Lemma A.4. We only prove the first statement. The second one follows by analogous arguments. By the definition of the physical dependence measure and the Cauchy-Schwarz inequality, we have with $p = \min\{q, q'\}/2$ that

$$\begin{aligned}
\delta_p(a_{ij}, s) &= \|a_{ij}(\mathcal{H}_{it,j}) - a_{ij}(\mathcal{H}'_{it,j})\| \\
&= \|\Delta h_{ij}(\mathcal{G}_{it})\Delta g_i(\mathcal{F}_{it}) - \Delta h_{ij}(\mathcal{G}'_{it})\Delta g_i(\mathcal{F}'_{it})\| \\
&\le \|h_{ij}(\mathcal{G}_{it})g_i(\mathcal{F}_{it}) - h_{ij}(\mathcal{G}'_{it})g_i(\mathcal{F}'_{it})\| \\
&\quad + \|h_{ij}(\mathcal{G}_{it-1})g_i(\mathcal{F}_{it-1}) - h_{ij}(\mathcal{G}'_{it-1})g_i(\mathcal{F}'_{it-1})\| \\
&\quad + \|h_{ij}(\mathcal{G}_{it-1})g_i(\mathcal{F}_{it}) - h_{ij}(\mathcal{G}'_{it-1})g_i(\mathcal{F}'_{it})\| \\
&\quad + \|h_{ij}(\mathcal{G}_{it})g_i(\mathcal{F}_{it-1}) - h_{ij}(\mathcal{G}'_{it})g_i(\mathcal{F}'_{it-1})\| \\
&= \|\{h_{ij}(\mathcal{G}_{it}) - h_{ij}(\mathcal{G}'_{it})\}g_i(\mathcal{F}_{it}) + h_{ij}(\mathcal{G}'_{it})\{g_i(\mathcal{F}_{it}) - g_i(\mathcal{F}'_{it})\}\| \\
&\quad + \|\{h_{ij}(\mathcal{G}_{it-1}) - h_{ij}(\mathcal{G}'_{it-1})\}g_i(\mathcal{F}_{it-1}) + h_{ij}(\mathcal{G}'_{it-1})\{g_i(\mathcal{F}_{it-1}) - g_i(\mathcal{F}'_{it-1})\}\| \\
&\quad + \|\{h_{ij}(\mathcal{G}_{it-1}) - h_{ij}(\mathcal{G}'_{it-1})\}g_i(\mathcal{F}_{it}) + h_{ij}(\mathcal{G}'_{it-1})\{g_i(\mathcal{F}_{it}) - g_i(\mathcal{F}'_{it})\}\| \\
&\quad + \|\{h_{ij}(\mathcal{G}_{it}) - h_{ij}(\mathcal{G}'_{it})\}g_i(\mathcal{F}_{it-1}) + h_{ij}(\mathcal{G}'_{it})\{g_i(\mathcal{F}_{it-1}) - g_i(\mathcal{F}'_{it-1})\}\| \\
&\le \delta_{2p}(h_{ij}, t)\|g_i(\mathcal{F}_t)\|_{2p} + \delta_{2p}(g_i, t)\|h_{ij}(\mathcal{G}'_{it})\|_{2p} \\
&\quad + \delta_{2p}(h_{ij}, t-1)\|g_i(\mathcal{F}_{t-1})\|_{2p} + \delta_{2p}(g_i, t-1)\|h_{ij}(\mathcal{G}'_{it-1})\|_{2p} \\
&\quad + \delta_{2p}(h_{ij}, t-1)\|g_i(\mathcal{F}_t)\|_{2p} + \delta_{2p}(g_i, t)\|h_{ij}(\mathcal{G}'_{it-1})\|_{2p} \\
&\quad + \delta_{2p}(h_{ij}, t)\|g_i(\mathcal{F}_{t-1})\|_{2p} + \delta_{2p}(g_i, t-1)\|h_{ij}(\mathcal{G}'_{it})\|_{2p},
\end{aligned}$$

where $\mathcal{H}'_{it,j} = (\dots, \nu_{i(-1),j}, \nu'_{i0,j}, \nu_{i1,j}, \dots, \nu_{it-1,j}, \nu_{it,j})$, $\mathcal{G}'_{it,j} = (\dots, \xi_{i(-1),j}, \xi'_{i0,j}, \xi_{i1,j}, \dots, \xi_{it-1,j}, \xi_{it,j})$ and $\mathcal{F}'_{it} = (\dots, \eta_{i(-1)}, \eta'_{i0}, \eta_{i1}, \dots, \eta_{it-1}, \eta_{it})$ are coupled processes with $\nu'_{i0,j}$, $\xi'_{i0,j}$ and $\eta'_{i0,j}$ being i.i.d. copies of $\nu_{i0,j}$, $\xi_{i0,j}$ and η_{i0}. From this and Assumptions (C1), (C3), (C4) and (C6), it immediately follows that $\sum_{s=t}^{\infty} \delta_p(a_{ij}, s) = O(t^{-\alpha})$. $\square$

We now show that the estimator $\widehat{\boldsymbol{\beta}}_i$ is $\sqrt{T}$-consistent for each i under our conditions.

Lemma A.5. *Let Assumptions (C1), (C3) and (C4)–(C7) be satisfied. Then for each i, it holds that*

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

Proof of Lemma A.5. The estimator $\widehat{\boldsymbol{\beta}}_i$ can be written as

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

where $\Delta \boldsymbol{X}_{it} = \boldsymbol{X}_{it} - \boldsymbol{X}_{it-1}$, $\Delta \varepsilon_{it} = \varepsilon_{it} - \varepsilon_{it-1}$ and $\Delta m_{it} = m_i(\frac{t}{T}) - m_i(\frac{t-1}{T})$. Hence,

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

In what follows, we show that

$$\frac{1}{\sqrt{T}} \sum_{t=2}^{T} \Delta \boldsymbol{X}_{it} \Delta \varepsilon_{it} = O_p(1) \qquad (A.2)$$

$$\Big(\frac{1}{T}\sum_{t=2}^{T} \Delta \boldsymbol{X}_{it} \Delta \boldsymbol{X}_{it}^\top\Big)^{-1} = O_p(1) \qquad (A.3)$$

$$\frac{1}{\sqrt{T}} \sum_{t=2}^{T} \Delta \boldsymbol{X}_{it} \Delta m_{it} = O_p\Big(\frac{1}{\sqrt{T}}\Big). \qquad (A.4)$$

Lemma A.5 follows from applying these three statements together with standard arguments to formula (A.1).

Since $\mathbb{E}[\Delta \boldsymbol{X}_{it} \Delta \varepsilon_{it}] = 0$ by (C7) and $\sum_{s=t}^{\infty} \delta_p(a_{ij}, s) = O(t^{-\alpha})$ for some $p > 2$ and all j by Lemma A.4, the claim (A.2) follows upon applying Lemma A.3. Another application of Lemma A.3 yields that

$$\frac{1}{T}\sum_{t=2}^{T} \Big\{\Delta \boldsymbol{X}_{it} \Delta \boldsymbol{X}_{it}^\top - \mathbb{E}[\Delta \boldsymbol{X}_{it} \Delta \boldsymbol{X}_{it}^\top]\Big\} = O_p\Big(\frac{1}{\sqrt{T}}\Big).$$

As $\mathbb{E}[\Delta \boldsymbol{X}_{it} \Delta \boldsymbol{X}_{it}^\top]$ is invertible, we can invoke Slutsky's lemma to obtain (A.3). By assumption, m_i is Lipschitz continuous, which implies that $|\Delta m_{it}| = |m_i(\frac{t}{T}) - m_i(\frac{t-1}{T})| \leq$

C/T for all $t \in \{1, \ldots, T\}$ and some constant $C > 0$. Hence,

$$\begin{aligned} \Big| \frac{1}{\sqrt{T}} \sum_{t=2}^{T} \Delta X_{it,j} \Delta m_{it} \Big| &\leq \frac{1}{\sqrt{T}} \sum_{t=2}^{T} |\Delta X_{it,j}| \cdot |\Delta m_{it}| \\ &\leq \frac{C}{\sqrt{T}} \cdot \frac{1}{T} \sum_{t=2}^{T} |\Delta X_{it,j}| = O_p\Big(\frac{1}{\sqrt{T}}\Big), \end{aligned}$$

where we have used that $T^{-1} \sum_{t=2}^{T} |\Delta \boldsymbol{X}_{it,j}| = O_p(1)$ by Markov's inequality. This yields (A.4). $\square$

Lemma A.6. *Let $s_T \asymp T^{1/3}$. Under Assumptions (C1)–(C7), for each $i \in \{1, \ldots, n\}$ we have*

$$\widehat{\sigma}_i^2 = \sigma_i^2 + O_p(T^{-1/3}).$$

where $\widehat{\sigma}_i^2$ is the subseries variance estimate of σ_i^2 introduced by (3.4).

Proof of Lemma A.6. (Proof not double-checked yet.) For notational convenience, we let $Y_{it}^* = Y_{it} - \boldsymbol{\beta}_i^\top \boldsymbol{X}_{it}$. Note that

$$\begin{aligned} &Y^*_{i(t+ms_T)} - Y^*_{i(t+(m-1)s_T)} \\ &= \alpha_i + m_i \left(\frac{t + ms_T}{T} \right) + \varepsilon_{i(t+ms_T)} \\ &\quad - \alpha_i - m_i \left(\frac{t + (m-1)s_T}{T} \right) + \varepsilon_{i(t+(m-1)s_T)} \\ &= m_i \left(\frac{t + ms_T}{T} \right) + \varepsilon_{i(t+ms_T)} - m_i \left(\frac{t + (m-1)s_T}{T} \right) + \varepsilon_{i(t+(m-1)s_T)} \\ &= Y^\circ_{i(t+ms_T)} - Y^\circ_{i(t+(m-1)s_T)}, \end{aligned}$$

where Y°_{it} is the dependent variable in a well-studied standard nonparametric regression discussed in Section 3.1. Now, using simple arithmetic calculations, we can rewrite $\widehat{\sigma}_i^2$ as

$$\begin{aligned} \widehat{\sigma}_i^2 = &\frac{1}{2(M-1)s_T} \sum_{m=1}^{M} \left[\sum_{t=1}^{s_T} \left(Y^\circ_{i(t+ms_T)} - Y^\circ_{i(t+(m-1)s_T)} \right) \right]^2 \\ &+ \frac{1}{2(M-1)s_T} \sum_{m=1}^{M} \left[\sum_{t=1}^{s_T} (\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i)^\top \left(\boldsymbol{X}_{i(t+ms_T)} - \boldsymbol{X}_{i(t+(m-1)s_T)} \right) \right]^2 \\ &- \frac{1}{(M-1)s_T} \sum_{m=1}^{M} \left[\sum_{t=1}^{s_T} \left(Y^\circ_{i(t+ms_T)} - Y^\circ_{i(t+(m-1)s_T)} \right) \right. \\ &\qquad\qquad \left. \times \sum_{t=1}^{s_T} (\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i)^\top \left(\boldsymbol{X}_{i(t+ms_T)} - \boldsymbol{X}_{i(t+(m-1)s_T)} \right) \right]. \end{aligned} \tag{A.5}$$

By Carlstein (1986) and Wu and Zhao (2007), we have

$$\frac{1}{2(M-1)s_T} \sum_{m=1}^{M} \left[\sum_{t=1}^{s_T} \left(Y^\circ_{i(t+ms_T)} - Y^\circ_{i(t+(m-1)s_T)} \right) \right]^2 = \sigma_i^2 + O_p(T^{-1/3}). \tag{A.6}$$

Furthermore, by our assumption that $s_T \asymp T^{1/3}$, Assumption (C5) and Lemma A.5, we have

$$\frac{1}{2(M-1)s_T}\sum_{m=1}^{M}\left[\sum_{t=1}^{s_T}(\widehat{\boldsymbol{\beta}}_i-\boldsymbol{\beta}_i)^\top\Big(\boldsymbol{X}_{i(t+ms_T)}-\boldsymbol{X}_{i(t+(m-1)s_T)}\Big)\right]^2 = O_p(T^{-2/3}). \quad \text{(A.7)}$$

Finally, applying (A.6) and (A.7) together with the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}\frac{1}{(M-1)s_T}\sum_{m=1}^{M}\Bigg[\sum_{t=1}^{s_T}\big(Y^\circ_{i(t+ms_T)}-Y^\circ_{i(t+(m-1)s_T)}\big)\\ \times\sum_{t=1}^{s_T}(\widehat{\boldsymbol{\beta}}_i-\boldsymbol{\beta}_i)^\top\big(\boldsymbol{X}_{i(t+ms_T)}-\boldsymbol{X}_{i(t+(m-1)s_T)}\big)\Bigg] = O_p(T^{-1/3}). \quad \text{(A.8)}\end{aligned}$$

Applying (A.6)–(A.8) to (A.5), the lemma trivially follows. □

Proof of Theorem 4.1

We first summarize the main proof strategy, which splits up into five steps, and then fill in the details. In particular, we defer the proofs of some intermediate results to the end of the section.

Step 1

To start with, we consider a simplified setting where the parameter vectors $\boldsymbol{\beta}_i$ are known. In this case, we can replace the estimators $\widehat{\boldsymbol{\beta}}_i$ in the definition of the statistic $\widehat{\Phi}_{n,T}$ by the true vectors $\boldsymbol{\beta}_i$ themselves. This leads to the simpler statistic

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

where

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

and $\widehat{\widehat{\sigma}}_i^2$ is computed in exactly the same way as $\widehat{\sigma}_i^2$ except that all occurrences of $\widehat{\boldsymbol{\beta}}_i$ are replaced by $\boldsymbol{\beta}_i$. By assumption, $\widehat{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$ with $\rho_T = o(\sqrt{h_{\min}}/\log T)$. (Double-check which rate is needed for $\widehat{\sigma}_i^2$.) For most estimators of σ_i^2 including those discussed in Section 4, this assumption immediately implies that $\widehat{\widehat{\sigma}}_i^2 = \sigma_i^2 + o_p(\rho_T)$ as well. In the sequel, we thus take for granted that the estimator $\widehat{\widehat{\sigma}}_i^2$ has this property.
We now have a closer look at the statistic $\widehat{\widehat{\Phi}}_{n,T}$. We in particular show that there exists an identically distributed version $\widetilde{\Phi}_{n,T}$ of $\widehat{\widehat{\Phi}}_{n,T}$ which is close to the Gaussian statistic $\Phi_{n,T}$ from (3.11). More formally, we prove the following result.

Proposition A.7. *There exist statistics* $\{\widetilde{\Phi}_{n,T} : T = 1, 2, \ldots\}$ *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)*

$$\big|\widetilde{\Phi}_{n,T} - \Phi_{n,T}\big| = o_p(\delta_T),$$

where $\delta_T = T^{1/q}/\sqrt{Th_{\min}} + \rho_T\sqrt{\log T}$ *and* $\Phi_{n,T}$ *is a Gaussian statistic as defined in (3.11).*

The proof makes heavy use of strong approximation theory for dependent processes. As it is quite technical, it is postponed to the end of this section.

Step 2

In this step, we establish some properties of the Gaussian statistic $\Phi_{n,T}$. Specifically, we prove the following result.

Proposition A.8. *It holds that*

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

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

Roughly speaking, this proposition says that the random variable $\Phi_{n,T}$ does not concentrate too strongly in small regions of the form $[x - \delta_T, x + \delta_T]$ with δ_T converging to 0. The main technical tool for deriving it are anti-concentration bounds for Gaussian random vectors. The details are provided below.

Step 3

We now use Steps 1 and 2 to prove that

$$\sup_{x\in\mathbb{R}} \big|\mathbb{P}(\widehat{\widehat{\Phi}}_{n,T} \le x) - \mathbb{P}(\Phi_{n,T} \le x)\big| = o(1). \tag{A.9}$$

Proof of (A.9). It holds that

$$\begin{aligned}
&\sup_{x\in\mathbb{R}} \Big|\mathbb{P}(\widehat{\widehat{\Phi}}_{n,T} \le x) - \mathbb{P}(\Phi_{n,T} \le x)\Big| \\
&= \sup_{x\in\mathbb{R}} \Big|\mathbb{P}(\widetilde{\Phi}_{n,T} \le x) - \mathbb{P}(\Phi_{n,T} \le x)\Big| \\
&= \sup_{x\in\mathbb{R}} \Big|\mathbb{E}\Big[1(\widetilde{\Phi}_{n,T} \le x) - 1(\Phi_{n,T} \le x)\Big]\Big| \\
&\le \sup_{x\in\mathbb{R}} \Big|\mathbb{E}\Big[\big\{1(\widetilde{\Phi}_{n,T} \le x) - 1(\Phi_{n,T} \le x)\big\}1\big(|\widetilde{\Phi}_{n,T} - \Phi_{n,T}| \le \delta_T\big)\Big]\Big| \\
&\quad + \mathbb{E}\Big[1\big(|\widetilde{\Phi}_{n,T} - \Phi_{n,T}| > \delta_T\big)\Big].
\end{aligned}$$

Moreover, since

$$\mathbb{E}\Big[1\big(|\widetilde{\Phi}_{n,T} - \Phi_{n,T}| > \delta_T\big)\Big] = \mathbb{P}\big(|\widetilde{\Phi}_{n,T} - \Phi_{n,T}| > \delta_T\big) = o(1)$$

by Step 1 and

$$\begin{aligned}
&\sup_{x\in\mathbb{R}} \Big|\mathbb{E}\Big[\big\{1(\widetilde{\Phi}_{n,T} \le x) - 1(\Phi_{n,T} \le x)\big\}1\big(|\widetilde{\Phi}_{n,T} - \Phi_{n,T}| \le \delta_T\big)\Big]\Big| \\
&\le \sup_{x\in\mathbb{R}} \mathbb{E}\Big[1\big(|\Phi_{n,T} - x| \le \delta_T, |\widetilde{\Phi}_{n,T} - \Phi_{n,T}| \le \delta_T\big)\Big] \\
&\le \sup_{x\in\mathbb{R}} \mathbb{P}\big(|\Phi_{n,T} - x| \le \delta_T\big) = o(1)
\end{aligned}$$

by Step 2, we arrive at (A.9). □

Step 4

In this step, we show that the auxiliary statistic $\widehat{\widehat{\Phi}}_{n,T}$ is close to $\widehat{\Phi}_{n,T}$ in the following sense.

Proposition A.9. *It holds that*

$$\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T} = o_p(\delta_T)$$

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

The proof can be found at the end of this section.

Step 5

We finally show that

$$\sup_{x\in\mathbb{R}} \big|\mathbb{P}(\widehat{\Phi}_{n,T} \le x) - \mathbb{P}(\Phi_{n,T} \le x)\big| = o(1). \tag{A.10}$$

Proof of (A.10). To start with, we verify that for any $x \in \mathbb{R}$ and any $\delta > 0$,

$$\begin{aligned}
&\mathbb{P}\Big(\widehat{\widehat{\Phi}}_{n,T} \le x - \delta\Big) - \mathbb{P}\Big(\big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| > \delta\Big) \\
&\qquad \le \mathbb{P}\big(\widehat{\Phi}_{n,T} \le x\big) \le \mathbb{P}\Big(\widehat{\widehat{\Phi}}_{n,T} \le x + \delta\Big) + \mathbb{P}\Big(\big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| > \delta\Big).
\end{aligned} \tag{A.11}$$

It holds that

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

and analogously

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

Combining these two inequalities, we arrive at (A.11).

Now let $x \in \mathbb{R}$ be any point such that $\mathbb{P}(\widehat{\Phi}_{n,T} \le x) \ge \mathbb{P}(\Phi_{n,T} \le x)$. With the help of (A.11), we get that

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

Analogously, for any point $x \in \mathbb{R}$ with $\mathbb{P}(\widehat{\Phi}_{n,T} \le x) < \mathbb{P}(\Phi_{n,T} \le x)$, it holds that

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

Consequently,

$$\begin{aligned}
\sup_{x\in\mathbb{R}} \Big|\mathbb{P}\big(\widehat{\Phi}_{n,T} \le x\big) - \mathbb{P}\big(\Phi_{n,T} \le x\big)\Big| &\le \sup_{x\in\mathbb{R}} \Big|\mathbb{P}\big(\widehat{\widehat{\Phi}}_{n,T} \le x\big) - \mathbb{P}\big(\Phi_{n,T} \le x\big)\Big| \\
&\quad + \sup_{x\in\mathbb{R}} \mathbb{P}\big(|\Phi_{n,T} - x| \le \delta_T\big) \\
&\quad + \mathbb{P}\Big(\big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| > \delta_T\Big).
\end{aligned}$$

Since the three terms on the right-hand side are all $o(1)$ by Steps 2–4, we arrive at (A.10). □

Details on Steps 1–5

Proof of Proposition A.7. Consider the stationary process $\mathcal{E}_i = \{\varepsilon_{it} : 1 \le t \le T\}$ for some fixed $i \in \{1, \ldots, n\}$. 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 exist 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

$$\max_{1\le t\le T} \Big| \sum_{s=1}^{t} \widetilde{\varepsilon}_{is} - \sigma_i \mathbb{B}_i(t) \Big| = o\big(T^{1/q}\big) \quad \text{a.s.,} \tag{A.12}$$

where $\sigma_i^2 = \sum_{k \in \mathbb{Z}} \mathrm{Cov}(\varepsilon_{i0}, \varepsilon_{ik})$ denotes the long-run error variance. We apply this result separately for each $i \in \{1, \ldots, n\}$. Since the error processes $\mathcal{E}_i = \{\varepsilon_{it} : 1 \le t \le T\}$ are independent across i, we can construct the processes $\widetilde{\mathcal{E}}_i = \{\widetilde{\varepsilon}_{it} : t \in \mathbb{N}\}$ in such a way that they are independent across i as well.

We now define the statistic $\widetilde{\Phi}_{n,T}$ in the same way as $\widehat{\widehat{\Phi}}_{n,T}$ except that the error processes $\mathcal{E}_i$ are replaced by $\widetilde{\mathcal{E}}_i$. Specifically, we set

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

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 the estimator $\widetilde{\sigma}_i^2$ is constructed from the sample $\widetilde{\mathcal{E}}_i$ in the same way as $\widehat{\widehat{\sigma}}_i^2$ is constructed from $\mathcal{E}_i$. Since $[\widetilde{\varepsilon}_{i1}, \ldots, \widetilde{\varepsilon}_{iT}] \overset{\mathcal{D}}{=} [\varepsilon_{i1}, \ldots, \varepsilon_{iT}]$ and $\widehat{\widehat{\sigma}}_i^2 = \sigma_i^2 + o_p(\rho_T)$, we have that $\widetilde{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$ as well. In addition to $\widetilde{\Phi}_{n,T}$, we introduce the Gaussian statistic

$$\Phi_{n,T} = \max_{1 \le i < j \le n} \max_{(u,h) \in \mathcal{G}_T} \left\{ \left| \frac{\phi_{ij,T}(u,h)}{\left(\sigma_i^2 + \sigma_j^2\right)^{1/2}} \right| - \lambda(h) \right\}$$

and the auxiliary statistic

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

where $\phi_{ij,T}(u,h) = \sum_{t=1}^{T} w_{t,T}(u,h) \{\sigma_i (Z_{it} - \bar{Z}_i) - \sigma_j (Z_{jt} - \bar{Z}_j)\}$ and the Gaussian variables Z_{it} are chosen as $Z_{it} = \mathbb{B}_i(t) - \mathbb{B}_i(t-1)$. With this notation, we obtain the obvious bound

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

In what follows, we prove that

$$\big|\widetilde{\Phi}_{n,T} - \Phi_{n,T}^{\diamond}\big| = o_p\Big(\frac{T^{1/q}}{\sqrt{T h_{\min}}}\Big) \tag{A.13}$$

$$\big|\Phi_{n,T}^{\diamond} - \Phi_{n,T}\big| = o_p(\rho_T \sqrt{\log T}), \tag{A.14}$$

which completes the proof.

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

$$\begin{aligned} \big|\widetilde{\Phi}_{n,T} - \Phi_{n,T}^{\diamond}\big| &\le \max_{1 \le i < j \le n} \left(\widetilde{\sigma}_i^2 + \widetilde{\sigma}_j^2\right)^{-1/2} \max_{1 \le i < j \le n} \max_{(u,h) \in \mathcal{G}_T} \big|\widetilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h)\big| \Big\} \\ &= O_p(1) \cdot \max_{1 \le i < j \le n} \max_{(u,h) \in \mathcal{G}_T} \big|\widetilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h)\big|, \end{aligned} \tag{A.15}$$

where the last line follows from the fact that $\widetilde{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$. Using summation by parts (that is, $\sum_{t=1}^T a_t b_t = \sum_{t=1}^{T-1} A_t(b_t - b_{t+1}) + A_T b_T$ with $A_t = \sum_{s=1}^t a_s$), we further obtain that

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

where

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

and $A_{ij,T} = 0$ for all pairs (i,j) by construction. From this, it follows that

$$\big|\widetilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h)\big| \le W_T(u,h) \max_{1 \le t \le T} |A_{ij,t}| \tag{A.16}$$

with $W_T(u,h) = \sum_{t=1}^{T-1} |w_{t+1,T}(u,h) - w_{t,T}(u,h)|$. Straightforward calculations yield that

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

Applying the strong approximation result (A.12), we can infer that

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

Moreover, standard arguments show that $\max_{(u,h)\in\mathcal{G}_T} W_T(u,h) = O(1/\sqrt{Th_{\min}})$. Plugging these two results into (A.16), we obtain that

$$\max_{1\le i<j\le n} \max_{(u,h)\in\mathcal{G}_T} \big|\widetilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h)\big| = o_p\Big(\frac{T^{1/q}}{\sqrt{Th_{\min}}}\Big),$$

which in view of (A.15) yields that $|\widetilde{\Phi}_{n,T} - \Phi^{\diamond}_{n,T}| = o_p(T^{1/q}/\sqrt{Th_{\min}})$. This completes the proof of (A.13).

Next consider $|\Phi^{\diamond}_{n,T} - \Phi_{n,T}|$. It holds that

$$\begin{aligned}
&\big|\Phi^{\diamond}_{n,T} - \Phi_{n,T}\big| \\
&\le \max_{1\le i<j\le n} \max_{(u,h)\in\mathcal{G}_T} \left| \frac{\phi_{ij,T}(u,h)}{\{\widetilde{\sigma}_i^2 + \widetilde{\sigma}_j^2\}^{1/2}} - \frac{\phi_{ij,T}(u,h)}{\{\sigma_i^2 + \sigma_j^2\}^{1/2}} \right| \\
&\le \max_{1\le i<j\le n} \left\{ \left| \big(\widetilde{\sigma}_i^2 + \widetilde{\sigma}_j^2\big)^{-1/2} - \big(\sigma_i^2 + \sigma_j^2\big)^{-1/2} \right| \right\} \max_{1\le i<j\le n} \max_{(u,h)\in\mathcal{G}_T} |\phi_{ij,T}(u,h)| \\
&= o_p(\rho_T) \max_{1\le i<j\le n} \max_{(u,h)\in\mathcal{G}_T} |\phi_{ij,T}(u,h)|, \qquad \text{(A.17)}
\end{aligned}$$

where the last line is due to the fact that $\widetilde{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$ and $\widehat{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$. We can write $\phi_{ij,T}(u,h) = \phi^{(I)}_{ij,T}(u,h) - \phi^{(II)}_{ij,T}(u,h)$, where

$$\phi^{(I)}_{ij,T}(u,h) = \sum_{t=1}^{T} w_{t,T}(u,h)\,(\sigma_i Z_{it} - \sigma_j Z_{jt}) \sim N(0, \sigma_i^2 + \sigma_j^2)$$

$$\phi^{(II)}_{ij,T}(u,h) = \sum_{t=1}^{T} w_{t,T}(u,h)\,(\sigma_i \bar{Z}_i - \sigma_j \bar{Z}_j) \sim N\big(0, (\sigma_i^2 + \sigma_j^2) c_T(u,h)\big)$$

with $c_T(u,h) = \{\sum_{t=1}^T w_{t,T}(u,h)\}^2/T \le C < \infty$ for all $(u,h) \in \mathcal{G}_T$ and $1 \le i < j \le n$. This shows that $\phi_{ij,T}(u,h)$ are centred Gaussian random variables with bounded variance for all $(u,h) \in \mathcal{G}_T$ and $1 \le i < j \le n$. Hence, standard results on the maximum of Gaussian random variables yield that

$$\max_{1\le i<j\le n} \max_{(u,h)\in\mathcal{G}_T} \big|\phi_{ij,T}(u,h)\big| = O_p(\sqrt{\log T}), \qquad \text{(A.18)}$$

where we have used that n is fixed and $|\mathcal{G}_T| = O(T^\theta)$ for some large but fixed constant θ by Assumption (C10). Plugging this into (A.17) yields $|\Phi^{\diamond}_{n,T} - \Phi_{n,T}| = o_p(\rho_T \sqrt{\log T})$, which completes the proof of (A.14). $\square$

Proof of Proposition A.8. The proof is an application of anti-concentration bounds for Gaussian random vectors. We in particular make use of the following anti-concentration inequality from Nazarov (2003), which can also be found as Lemma A.1 in Chernozhukov et al. (2017).

Lemma A.10. *Let $\boldsymbol{Z} = (Z_1, \dots, Z_p)^\top$ be a centred Gaussian random vector in $\mathbb{R}^p$ such that $\mathbb{E}[Z_j^2] \ge b$ for all $1 \le j \le p$ and some constant $b > 0$. Then for every $\boldsymbol{z} \in \mathbb{R}^p$ and $a > 0$,*

$$\mathbb{P}(\boldsymbol{Z} \le \boldsymbol{z} + a) - \mathbb{P}(\boldsymbol{Z} \le \boldsymbol{z}) \le Ca\sqrt{\log p},$$

where the constant C only depends on b. (Add notation vector + scalar?)

To apply this result, we introduce the following notation: We write $x = (u, h)$ and $\mathcal{G}_T = \{x : x \in \mathcal{G}_T\} = \{x_1, \ldots, x_p\}$, where $p := |\mathcal{G}_T| \le O(T^\theta)$ for some large but fixed $\theta > 0$ by our assumptions. For $k = 1, \ldots, p$ and $1 \le i < j \le n$, we further let

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

along with $\lambda_{ij,2k-1} = \lambda(x_{k2})$ and $\lambda_{ij,2k} = \lambda(x_{k2})$, where $x_k = (x_{k1}, x_{k2})$. Under our assumptions, it holds that $\mathbb{E}[Z_{ij,l}] = 0$ and $\mathbb{E}[Z_{ij,l}^2] \ge b > 0$ for all i, j and l. We next construct the random vector $\boldsymbol{Z} = (Z_{ij,l} : 1 \le i < j \le n, 1 \le l \le 2p)$ by stacking the variables $Z_{ij,l}$ in a certain order (which can be chosen freely) and construct the vector $\boldsymbol{\lambda} = (\lambda_{ij,l} : 1 \le i < j \le n, 1 \le l \le 2p)$ in an analogous way. Since the variables $Z_{ij,l}$ are normally distributed, $\boldsymbol{Z}$ is a Gaussian random vector of length $(n-1)np$.
With this notation at hand, we can express the probability $\mathbb{P}(\Phi_{n,T} \le q)$ as follows for each $q \in \mathbb{R}$:

$$\begin{aligned}
\mathbb{P}(\Phi_{n,T} \le q) &= \mathbb{P}\Big(\max_{1 \le i < j \le n} \max_{1 \le l \le 2p} \big\{ Z_{ij,l} - \lambda_{ij,l} \big\} \le q \Big) \\
&= \mathbb{P}\big(Z_{ij,l} \le \lambda_{ij,l} + q \text{ for all } (i,j,l) \big) \\
&= \mathbb{P}\big(\boldsymbol{Z} \le \boldsymbol{\lambda} + q \big).
\end{aligned}$$

Consequently,

$$\begin{aligned}
\mathbb{P}\big(|\Phi_{n,T} - x| \le \delta_T\big) &= \mathbb{P}\big(x - \delta_T \le \Phi_{n,T} \le x + \delta_T\big) \\
&= \mathbb{P}\big(\Phi_{n,T} \le x + \delta_T\big) - \mathbb{P}\big(\Phi_{n,T} \le x\big) \\
&\quad + \mathbb{P}\big(\Phi_{n,T} \le x\big) - \mathbb{P}\big(\Phi_{n,T} \le x - \delta_T\big) \\
&= \mathbb{P}\big(\boldsymbol{Z} \le \boldsymbol{\lambda} + x + \delta_T\big) - \mathbb{P}\big(\boldsymbol{Z} \le \boldsymbol{\lambda} + x\big) \\
&\quad + \mathbb{P}\big(\boldsymbol{Z} \le \boldsymbol{\lambda} + x\big) - \mathbb{P}\big(\boldsymbol{Z} \le \boldsymbol{\lambda} + x - \delta_T\big) \\
&\le 2C\delta_T \sqrt{\log((n-1)np)},
\end{aligned}$$

where the last line is by Lemma A.10. This immediately implies Proposition A.8. $\square$

Proof of Proposition A.9. Straightforward calculations yield that

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

Since $\widehat{\widehat{\sigma}}_i^2 = \sigma_i^2 + o_p(\rho_T)$ and $\widehat{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$, we further get that

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

and

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

where the difference of the kernel averages $\widehat{\widehat{\phi}}_{ij,T}(u,h) - \widehat{\phi}_{ij,T}(u,h)$ does not include the error terms (they cancel out) and can be written as

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

Hence,

$$\left| \widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T} \right| \le o_p(\rho_T) A_{n,T} + O_p(1) \big\{ 2B_{n,T} + 2C_{n,T} \big\}, \tag{A.19}$$

where

$$\begin{aligned}
A_{n,T} &= \max_{1\le i<j\le n} \max_{(u,h)\in\mathcal{G}_T} \left| \widehat{\widehat{\phi}}_{ij,T}(u,h) \right| \\
B_{n,T} &= \max_{1\le i\le n} \max_{(u,h)\in\mathcal{G}_T} \left| (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top \sum_{t=1}^T w_{t,T}(u,h) \boldsymbol{X}_{it} \right| \\
C_{n,T} &= \max_{1\le i\le n} \left| (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top \bar{\boldsymbol{X}}_i \right| \max_{(u,h)\in\mathcal{G}_T} \left| \sum_{t=1}^T w_{t,T}(u,h) \right|.
\end{aligned}$$

We examine each of these three terms separately.

We first prove that

$$A_{n,T} = \max_{1 \le i < j \le n} \max_{(u,h) \in \mathcal{G}_T} \Big| \widehat{\widehat{\phi}}_{ij,T}(u,h) \Big| = O_p\big(\sqrt{\log T}\big). \tag{A.20}$$

From the proof of Proposition A.7, we know that there exist identically distributed versions $\widetilde{\phi}_{ij,T}(u,h)$ of the statistics $\widehat{\widehat{\phi}}_{ij,T}(u,h)$ with the property that

$$\max_{1 \le i < j \le n} \max_{(u,h) \in \mathcal{G}_T} \big| \widetilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h) \big| = o_p\Big(\frac{T^{1/q}}{\sqrt{T h_{\min}}} \Big). \tag{A.21}$$

Instead of (A.20), it thus suffices to show that

$$\max_{1 \le i < j \le n} \max_{(u,h) \in \mathcal{G}_T} \Big| \widetilde{\phi}_{ij,T}(u,h) \Big| = O_p\big(\sqrt{\log T}\big). \tag{A.22}$$

Since for any constant $c > 0$,

$$\begin{aligned}
&\mathbb{P}\left(\max_{i,j,(u,h)} |\phi_{ij,T}(u,h)| \le \frac{c\sqrt{\log T}}{2} \right) \\
&\le \mathbb{P}\left(\max_{i,j,(u,h)} \Big| \widetilde{\phi}_{ij,T}(u,h) \Big| \le c\sqrt{\log T} \right) \\
&\quad + \mathbb{P}\left(\Big| \max_{i,j,(u,h)} \Big| \widetilde{\phi}_{ij,T}(u,h) \Big| - \max_{i,j,(u,h)} |\phi_{ij,T}(u,h)| \Big| > \frac{c\sqrt{\log T}}{2} \right) \\
&\le \mathbb{P}\left(\max_{i,j,(u,h)} \Big| \widetilde{\phi}_{ij,T}(u,h) \Big| \le c\sqrt{\log T} \right) \\
&\quad + \mathbb{P}\left(\max_{i,j,(u,h)} \Big| \widetilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h) \Big| > \frac{c\sqrt{\log T}}{2} \right)
\end{aligned}$$

and $\mathbb{P}(\max_{i,j,(u,h)} |\widetilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h)| > c\sqrt{\log T}/2) = o(1)$ by (A.21), we obtain that

$$\begin{aligned}
\mathbb{P}\Big(\max_{i,j,(u,h)} \Big| \widetilde{\phi}_{ij,T}(u,h) \Big| \le c\sqrt{\log T} \Big) \\
\ge \mathbb{P}\left(\max_{i,j,(u,h)} |\phi_{ij,T}(u,h)| \le \frac{c\sqrt{\log T}}{2} \right) - o(1).
\end{aligned} \tag{A.23}$$

Moreover, since $\max_{i,j,(u,h)} |\phi_{ij,T}(u,h)| = O_p(\sqrt{\log T})$ as already proven in (A.18), we can make the probability $\mathbb{P}(\max_{i,j,(u,h)} |\phi_{ij,T}(u,h)| \le c\sqrt{\log T}/2)$ on the right-hand side of (A.23) arbitrarily close to 1 by choosing the constant c sufficiently large. Hence, for any $\delta > 0$, we can find a constant $c > 0$ such that $\mathbb{P}(\max_{i,j,(u,h)} |\widetilde{\phi}_{ij,T}(u,h)| \le c\sqrt{\log T}) \ge 1 - \delta$ for sufficiently large T. This proves (A.22), which in turn yields (A.20).

We next turn to $B_{n,T}$. Without loss of generality, we assume that $\boldsymbol{X}_{it}$ is real-valued. The vector-valued case can be handled analogously. To start with, we have a closer look at the term $\sum_{t=1}^T w_{t,T}(u,h) \boldsymbol{X}_{it}$. By construction, the kernel weights $w_{t,T}(u,h)$ are unequal to 0 if and only if $T(u-h) \le t \le T(u+h)$. We can use this fact to write

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

Note that

$$\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} w_{t,T}^2(u,h) = \sum_{t=1}^{T} w_{t,T}^2(u,h) = \sum_{t=1}^{T} \frac{\Lambda_{t,T}^2(u,h)}{\sum_{s=1}^{T}\Lambda_{s,T}^2(u,h)} = 1. \tag{A.24}$$

Denoting by $D_{T,u,h}$ the number of integers between $\lfloor T(u-h)\rfloor$ and $\lceil T(u+h)\rceil$ (with the obvious bounds $2Th \le D_{T,u,h} \le 2Th+2$) and using (A.24), we can normalize the kernel weights as follows:

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

Next, we apply Proposition A.2 with the weights $a_t = \sqrt{D_{T,u,h}}\cdot w_{t,T}(u,h)$ to obtain that

$$\begin{aligned}\mathbb{P}\bigg(\Big|\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} \sqrt{D_{T,u,h}}\cdot w_{t,T}(u,h)\boldsymbol{X}_{it}\Big| \ge x\bigg) \\ \le C_1 \frac{\big(\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} |\sqrt{D_{T,u,h}}\cdot w_{t,T}(u,h)|^{q'}\big)\|\boldsymbol{X}_{i\cdot}\|_{q',\alpha}^{q'}}{x^{q'}} \\ + C_2 \exp\bigg(-\frac{C_3 x^2}{D_{T,u,h}\|\boldsymbol{X}_{i\cdot}\|_{2,\alpha}^2}\bigg)\end{aligned} \tag{A.25}$$

for any $x > 0$, where $\|\boldsymbol{X}_{i\cdot}\|_{q',\alpha}^{q'} = \sup_{t\ge 0}(t+1)^{\alpha}\sum_{s=t}^{\infty}\delta_{q'}(\boldsymbol{h}_i, s)$ is the dependence adjusted norm introduced in Definition A.1 and $\|\boldsymbol{X}_{i\cdot}\|_{q',\alpha}^{q'} < \infty$ by Assumption (C6). From (A.25), it follows that for any $\delta > 0$,

$$\begin{aligned}&\mathbb{P}\bigg(\max_{(u,h)\in\mathcal{G}_T}\Big|\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} w_{t,T}(u,h)\boldsymbol{X}_{it}\Big| \ge \delta T^{1/q}\bigg) \\ &\le \sum_{(u,h)\in\mathcal{G}_T}\mathbb{P}\bigg(\Big|\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} w_{t,T}(u,h)\boldsymbol{X}_{it}\Big| \ge \delta T^{1/q}\bigg) \\ &= \sum_{(u,h)\in\mathcal{G}_T}\mathbb{P}\bigg(\Big|\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} \sqrt{D_{T,u,h}}\cdot w_{t,T}(u,h)\boldsymbol{X}_{it}\Big| \ge \delta\sqrt{D_{T,u,h}}T^{1/q}\bigg) \\ &\le \sum_{(u,h)\in\mathcal{G}_T}\bigg[C_1\frac{(\sqrt{D_{T,u,h}})^{q'}\big(\sum|w_{t,T}(u,h)|^{q'}\big)\|\boldsymbol{X}_{i\cdot}\|_{q',\alpha}^{q'}}{(\delta\sqrt{D_{T,u,h}}T^{1/q})^{q'}} + C_2\exp\bigg(-\frac{C_3(\delta\sqrt{D_{T,u,h}}T^{1/q})^2}{D_{T,u,h}\|\boldsymbol{X}_{i\cdot}\|_{2,\alpha}^2}\bigg)\bigg] \\ &= \sum_{(u,h)\in\mathcal{G}_T}\bigg[C_1\frac{\big(\sum|w_{t,T}(u,h)|^{q'}\big)\|\boldsymbol{X}_{i\cdot}\|_{q',\alpha}^{q'}}{\delta^{q'}T^{q'/q}} + C_2\exp\bigg(-\frac{C_3\delta^2T^{2/q}}{\|\boldsymbol{X}_{i\cdot}\|_{2,\alpha}^2}\bigg)\bigg]\end{aligned}$$

$$\leq C_1 \frac{T^\theta \|\boldsymbol{X}_{i\cdot}\|_{q',\alpha}^{q'}}{\delta^{q'} T^{q'/q}} \max_{(u,h)\in\mathcal{G}_T} \Big(\sum\nolimits_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} |w_{t,T}(u,h)|^{q'} \Big) + C_2 T^\theta \exp\Big(- \frac{C_3 \delta^2 T^{2/q}}{\|\boldsymbol{X}_{i\cdot}\|_{2,\alpha}^2} \Big)$$
$$= C \frac{T^{\theta - q'/q}}{\delta^{q'}} + C T^\theta \exp\big(-C T^{2/q} \delta^2 \big), \tag{A.26}$$

where the constant C depends neither on T nor on δ. In the last equality of the above display, we have used the following facts:

(i) $\|\boldsymbol{X}_{i\cdot}\|_{q',\alpha}^{q'} < \infty$ by Assumption (C6).

(ii) $\|\boldsymbol{X}_{i\cdot}\|_{2,\alpha}^2 < \infty$ (which follows from (i)).

(iii) $\max_{(u,h)\in\mathcal{G}_T} (\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} |w_{t,T}(u,h)|^{q'}) \leq 1$ for the following reason: By (A.24), it holds that $\sum_{t=1}^T w_{t,T}^2(u,h) = 1$ and thus $0 \leq w_{t,T}^2(u,h) \leq 1$ for all t, T and (u,h). This implies that $0 \leq |w_{t,T}(u,h)|^{q'} = (w_{t,T}^2(u,h))^{q'/2} \leq w_{t,T}^2(u,h) \leq 1$ for all t, T and (u,h). As a result,
$$\max_{(u,h)\in\mathcal{G}_T} \Bigg(\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} |w_{t,T}(u,h)|^{q'} \Bigg) \leq \max_{(u,h)\in\mathcal{G}_T} \Bigg(\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} w_{t,T}^2(u,h) \Bigg) = 1.$$

Since $\theta - q'/q < 0$ by Assumption (C6), the bound in (A.26) converges to 0 as $T \to \infty$ for any fixed $\delta > 0$. Consequently, we obtain that
$$\max_{(u,h)\in\mathcal{G}_T} \Bigg| \sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} w_{t,T}(u,h) \boldsymbol{X}_{it} \Bigg| = o_p(T^{1/q}). \tag{A.27}$$

Using this together with the fact that $\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i = O_p(1/\sqrt{T})$ (which is the statement of Lemma A.5), we arrive at the bound
$$B_{n,T} = \max_{1\leq i\leq n} \max_{(u,h)\in\mathcal{G}_T} \Big| (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top \sum_{t=1}^T w_{t,T}(u,h) \boldsymbol{X}_{it} \Big| = o_p\Big(\frac{T^{1/q}}{\sqrt{T}} \Big).$$

We finally turn to $C_{n,T}$. Straightforward calculations yield that $|\sum_{t=1}^T w_{t,T}(u,h)| \leq C\sqrt{Th_{\max}} = o(\sqrt{T})$. Moreover, $\bar{\boldsymbol{X}}_i = O_p(1/\sqrt{T})$ by Lemma A.3 and $\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i = O_p(1/\sqrt{T})$ by Lemma A.5. This immediately yields that
$$C_{n,T} = \max_{1\leq i\leq n} \big| (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top \bar{\boldsymbol{X}}_i \big| \max_{(u,h)\in\mathcal{G}_T} \Big| \sum_{t=1}^T w_{t,T}(u,h) \Big| = o_p\Big(\frac{1}{\sqrt{T}} \Big).$$

To summarize, we have shown that
$$\begin{aligned} \big| \widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T} \big| &\leq o_p(\rho_T) A_{n,T} + O_p(1) \big\{ 2B_{n,T} + 2C_{n,T} \big\} \\ &= o_p(\rho_T) O_p(\sqrt{\log T}) + o_p\Big(\frac{T^{1/q}}{\sqrt{T}} \Big) + o_p\Big(\frac{1}{\sqrt{T}} \Big), \end{aligned}$$

where $\{c_T\}$ is any sequence of positive real numbers with $\sqrt{\log T} \ll c_T$. This immediately implies the desired result. □

Proof of Proposition 4.3

We first show that

$$\mathbb{P}(\Phi_{n,T} \le q_{n,T}(\alpha)) = 1 - \alpha. \tag{A.28}$$

We proceed by contradiction. Suppose that (A.28) does not hold true. Since $\mathbb{P}(\Phi_{n,T} \le q_{n,T}(\alpha)) \ge 1 - \alpha$ by definition of the quantile $q_{n,T}(\alpha)$, there exists $\xi > 0$ such that $\mathbb{P}(\Phi_{n,T} \le q_{n,T}(\alpha)) = 1 - \alpha + \xi$. From the proof of Proposition A.8, we know that for any $\delta > 0$,

$$\begin{aligned}
&\mathbb{P}\big(\Phi_{n,T} \le q_{n,T}(\alpha)\big) - \mathbb{P}\big(\Phi_{n,T} \le q_{n,T}(\alpha) - \delta\big) \\
&\qquad \le \sup_{x \in \mathbb{R}} \mathbb{P}\big(|\Phi_{n,T} - x| \le \delta\big) \le 2C\delta\sqrt{\log((n-1)np)}.
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathbb{P}\big(\Phi_{n,T} \le q_{n,T}(\alpha) - \delta\big) &\ge \mathbb{P}\big(\Phi_{n,T} \le q_{n,T}(\alpha)\big) - 2C\delta\sqrt{\log((n-1)np)} \\
&= 1 - \alpha + \xi - 2C\delta\sqrt{\log((n-1)np)} > 1 - \alpha
\end{aligned}$$

for $\delta > 0$ small enough. This contradicts the definition of the quantile $q_{n,T}(\alpha)$ according to which $q_{n,T}(\alpha) = \inf_{q \in \mathbb{R}}\{\mathbb{P}(\Phi_{n,T} \le q) \ge 1 - \alpha\}$. We thus arrive at (A.28).
Proposition 4.3 is a simple consequence of Theorem 4.1 and Equation (A.28). Specifically, we obtain that under H_0,

$$\begin{aligned}
\big|\mathbb{P}(\widehat{\Psi}_{n,T} \le q_{n,T}(\alpha)) - (1-\alpha)\big| &= \big|\mathbb{P}(\widehat{\Phi}_{n,T} \le q_{n,T}(\alpha)) - (1-\alpha)\big| \\
&= \big|\mathbb{P}(\widehat{\Phi}_{n,T} \le q_{n,T}(\alpha)) - \mathbb{P}(\Phi_{n,T} \le q_{n,T}(\alpha))\big| \\
&\le \sup_{x \in \mathbb{R}} \big|\mathbb{P}(\widehat{\Phi}_{n,T} \le x) - \mathbb{P}(\Phi_{n,T} \le x)\big| = o(1).
\end{aligned}$$

Proof of Proposition 4.4

To start with, note that for some sufficiently large constant C we have

$$\lambda(h) = \sqrt{2\log\{1/(2h)\}} \le \sqrt{2\log\{1/(2h_{\min})\}} \le C\sqrt{\log T}. \tag{A.29}$$

Write $\widehat{\psi}_{ij,T}(u,h) = \widehat{\psi}^A_{ij,T}(u,h) + \widehat{\psi}^B_{ij,T}(u,h)$ with

$$\begin{aligned}
\widehat{\psi}^A_{ij,T}(u,h) &= \sum_{t=1}^T w_{t,T}(u,h)\big\{(\varepsilon_{it} - \bar{\varepsilon}_i) + (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top(\boldsymbol{X}_{it} - \bar{\boldsymbol{X}}_i) - \bar{m}_{i,T} \\
&\qquad - (\varepsilon_{jt} - \bar{\varepsilon}_j) - (\boldsymbol{\beta}_j - \widehat{\boldsymbol{\beta}}_j)^\top(\boldsymbol{X}_{jt} - \bar{\boldsymbol{X}}_j) + \bar{m}_{j,T}\big\} \\
\widehat{\psi}^B_{ij,T}(u,h) &= \sum\nolimits_{t=1}^T w_{t,T}(u,h)\bigg(m_{i,T}\Big(\frac{t}{T}\Big) - m_{j,T}\Big(\frac{t}{T}\Big)\bigg),
\end{aligned}$$

where $\bar{m}_{i,T} = T^{-1}\sum_{t=1}^T m_{i,T}(t/T)$. Without loss of generality, consider the following scenario: there exists $(u_0, h_0) \in \mathcal{G}_T$ with $[u_0 - h_0, u_0 + h_0] \subseteq [0,1]$ such that

$$m_{i,T}(w) - m_{j,T}(w) \ge c_T\sqrt{\log T/(Th_0)} \tag{A.30}$$

for all $w \in [u_0 - h_0, u_0 + h_0]$.

We first derive a lower bound on the term $\widehat{\psi}^B_{ij,T}(u_0, h_0)$. Since the kernel K is symmetric and $u_0 = t/T$ for some t, it holds that $S_{T,1}(u_0, h_0) = 0$ and thus,

$$\begin{aligned} w_{t,T}(u_0, h_0) &= \frac{K\Big(\frac{\frac{t}{T} - u_0}{h_0}\Big) S_{T,2}(u_0, h_0)}{\Big\{\sum_{t=1}^T K^2\Big(\frac{\frac{t}{T} - u_0}{h_0}\Big) S^2_{T,2}(u_0, h_0)\Big\}^{1/2}} \\ &= \frac{K\Big(\frac{\frac{t}{T} - u_0}{h_0}\Big)}{\Big\{\sum_{t=1}^T K^2\Big(\frac{\frac{t}{T} - u_0}{h_0}\Big)\Big\}^{1/2}} \geq 0. \end{aligned}$$

Together with (A.30), this implies that

$$\widehat{\psi}^B_{ij,T}(u_0, h_0) \geq c_T \sqrt{\frac{\log T}{T h_0}} \sum_{t=1}^T w_{t,T}(u_0, h_0). \tag{A.31}$$

Using the Lipschitz continuity of the kernel K, we can show by straightforward calculations that for any $(u, h) \in \mathcal{G}_T$ and any natural number ℓ,

$$\Big| \frac{1}{Th} \sum_{t=1}^T K\Big(\frac{\frac{t}{T} - u}{h}\Big)\Big(\frac{\frac{t}{T} - u}{h}\Big)^\ell - \int_0^1 \frac{1}{h} K\Big(\frac{w - u}{h}\Big)\Big(\frac{w - u}{h}\Big)^\ell dw \Big| \leq \frac{C}{Th}, \tag{A.32}$$

where the constant C does not depend on u, h and T. With the help of (A.32), we obtain that for any $(u, h) \in \mathcal{G}_T$ with $[u - h, u + h] \subseteq [0, 1]$,

$$\Big| \sum_{t=1}^T w_{t,T}(u, h) - \frac{\sqrt{Th}}{\kappa} \Big| \leq \frac{C}{\sqrt{Th}}, \tag{A.33}$$

where $\kappa = (\int K^2(\varphi) d\varphi)^{1/2}$ and the constant C does once again not depend on u, h and T. From (A.33), it follows that $\sum_{t=1}^T w_{t,T}(u, h) \geq \sqrt{Th}/(2\kappa)$ for sufficiently large T and any $(u, h) \in \mathcal{G}_T$ with $[u - h, u + h] \subseteq [0, 1]$. This together with (A.31) allows us to infer that

$$\widehat{\psi}^B_{ij,T}(u_0, h_0) \geq \frac{c_T \sqrt{\log T}}{2\kappa} \tag{A.34}$$

for sufficiently large T.

We next analyze $\widehat{\psi}^A_{ij,T}(u_0, h_0)$, which can be expressed as $\widehat{\psi}^A_{ij,T}(u_0, h_0) = \widehat{\phi}_{ij,T}(u, h) + (\bar{m}_{j,T} - \bar{m}_{i,T}) \sum_{t=1}^T w_{t,T}(u, h)$. The proof of Proposition A.9 shows that

$$\max_{1 \leq i < j \leq n} \max_{(u,h) \in \mathcal{G}_T} \Big| \widehat{\phi}_{ij,T}(u, h) \Big| = O_p(\sqrt{\log T}).$$

Using this together with the bounds $\bar{m}_{i,T} \leq C/T$ and $\sum_{t=1}^T w_{t,T}(u, h) \leq C\sqrt{T}$, we can infer that

$$\begin{aligned} &\max_{1 \leq i < j \leq n} \max_{(u,h) \in \mathcal{G}_T} \Big| \widehat{\psi}^A_{ij,T}(u, h) \Big| \\ &= \max_{1 \leq i < j \leq n} \max_{(u,h) \in \mathcal{G}_T} \Big| \widehat{\phi}_{ij,T}(u, h) + (\bar{m}_{j,T} - \bar{m}_{i,T}) \sum_{t=1}^T w_{t,T}(u, h) \Big| = O_p(\sqrt{\log T}). \end{aligned} \tag{A.35}$$

With the help of (A.34), (A.35), (A.29) and the assumption that $\widehat{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$, we finally arrive at

$$\begin{aligned}
\widehat{\Psi}_{n,T} &\geq \max_{1\leq i<j\leq n} \max_{(u,h)\in\mathcal{G}_T} \frac{|\widehat{\psi}^B_{ij,T}(u,h)|}{\{\widehat{\sigma}_i^2+\widehat{\sigma}_j^2\}^{1/2}} - \max_{1\leq i<j\leq n} \max_{(u,h)\in\mathcal{G}_T} \left\{ \frac{|\widehat{\psi}^A_{ij,T}(u,h)|}{\{\widehat{\sigma}_i^2+\widehat{\sigma}_j^2\}^{1/2}} + \lambda(h) \right\} \\
&= \max_{1\leq i<j\leq n} \max_{(u,h)\in\mathcal{G}_T} \frac{|\widehat{\psi}^B_{ij,T}(u,h)|}{\{\widehat{\sigma}_i^2+\widehat{\sigma}_j^2\}^{1/2}} + O_p(\sqrt{\log T}) \\
&\geq \frac{c_T\sqrt{\log T}}{2\kappa} \min_{1\leq i<j\leq n} \{\widehat{\sigma}_i^2+\widehat{\sigma}_j^2\}^{-1/2} + O_p(\sqrt{\log T}) \\
&\geq C\frac{c_T\sqrt{\log T}}{2\kappa} + O_p(\sqrt{\log T}).
\end{aligned} \tag{A.36}$$

Since $q_{n,T}(\alpha) = O(\sqrt{\log T})$ for any fixed $\alpha \in (0,1)$ and $c_T \to \infty$, (A.36) immediately implies that $\mathbb{P}(\widehat{\Psi}_{n,T} \leq q_{n,T}(\alpha)) = o(1)$.

Proof of Proposition 4.5

Denote by $\mathcal{M}_0$ the set of quadruples $(i,j,u,h) \in \{1\ldots,n\}^2 \times \mathcal{G}_T$ for which $H_0^{[i,j]}(u,h)$ is true. Then we can write the FWER as

$$\begin{aligned}
\text{FWER}(\alpha) &= \mathbb{P}\Big(\exists (i,j,u,h) \in \mathcal{M}_0 : \widehat{\psi}^0_{ij,T}(u,h) > q_{n,T}(\alpha)\Big) \\
&= \mathbb{P}\Big(\max_{(i,j,u,h)\in\mathcal{M}_0} \widehat{\psi}^0_{ij,T}(u,h) > q_{n,T}(\alpha)\Big) \\
&= \mathbb{P}\Big(\max_{(i,j,u,h)\in\mathcal{M}_0} \widehat{\phi}^0_{ij,T}(u,h) > q_{n,T}(\alpha)\Big) \\
&\leq \mathbb{P}\Big(\max_{1\leq i<j\leq n} \max_{(u,h)\in\mathcal{G}_T} \widehat{\phi}^0_{ij,T}(u,h) > q_{n,T}(\alpha)\Big) \\
&= \mathbb{P}\big(\widehat{\Phi}_{n,T} > q_{n,T}(\alpha)\big) = \alpha + o(1),
\end{aligned}$$

where the third equality uses that $\hat{\psi}^0_{ijk,T} = \hat{\phi}^0_{ijk,T}$ under $H_0^{[i,j]}(u,h)$.

Proof of Corollary 4.6

By Proposition 4.5,

$$\begin{aligned}
1 - \alpha + o(1) &\leq 1 - \text{FWER}(\alpha) \\
&= \mathbb{P}\Big(\nexists (i,j,u,h) \in \mathcal{M}_0 : \hat{\psi}^0_{ij,T}(u,h) > q_{n,T}(\alpha)\Big) \\
&= \mathbb{P}\Big(\forall (i,j,u,h) \in \mathcal{M}_0 : \hat{\psi}^0_{ij,T}(u,h) \leq q_{n,T}(\alpha)\Big) \\
&= \mathbb{P}\Big(\forall\, i,j \in \{1,\ldots,n\}, (u,h) \in \mathcal{G}_T \text{ such that} \\
&\qquad\qquad H_0^{[i,j]}(u,h) \text{ is true} : \hat{\psi}^0_{ij,T}(u,h) \leq q_{n,T}(\alpha)\Big),
\end{aligned}$$

which is the statement of Corollary 4.6.

Proof of Proposition 5.1

For the sake of brevity, we introduce the following notation. For each i and j, we define the statistics $\widehat{\Psi}_{ij,T} := \max_{(u,h)\in\mathcal{G}_T} \hat{\psi}^0_{ij,T}(u,h)$ which can be interpreted as a distance measure between the two curves m_i and m_j on the whole interval $[0,1]$. Using this notation, we can rewrite the dissimilarity measure defined in (5.1) as

$$\widehat{\Delta}(S,S') = \max_{\substack{i\in S,\\ j\in S'}} \widehat{\Psi}_{ij,T}.$$

Now consider the event

$$B_{n,T} = \Big\{ \max_{1\le\ell\le N} \max_{i,j\in G_\ell} \widehat{\Psi}_{ij,T} \le q_{n,T}(\alpha) \ \text{ and } \ \min_{1\le\ell<\ell'\le N} \min_{\substack{i\in G_\ell,\\ j\in G_{\ell'}}} \widehat{\Psi}_{ij,T} > q_{n,T}(\alpha) \Big\}.$$

The term $\max_{1\le\ell\le N}\max_{i,j\in G_\ell} \widehat{\Psi}_{ij,T}$ is the largest multiscale distance between two time series i and j from the same group, whereas $\min_{1\le\ell<\ell'\le N}\min_{i\in G_\ell, j\in G_{\ell'}} \widehat{\Psi}_{ij,T}$ is the smallest multiscale distance between two time series from two different groups. On the event $B_{n,T}$, it obviously holds that

$$\max_{1\le\ell\le N} \max_{i,j\in G_\ell} \widehat{\Psi}_{ij,T} < \min_{1\le\ell<\ell'\le N} \min_{\substack{i\in G_\ell,\\ j\in G_{\ell'}}} \widehat{\Psi}_{ij,T}. \tag{A.37}$$

Hence, any two time series from the same class have a smaller distance than any two time series from two different classes. With the help of Proposition 4.3, it is easy to see that

$$\mathbb{P}\Big(\max_{1\le\ell\le N} \max_{i,j\in G_\ell} \widehat{\Psi}_{ij,T} \le q_{n,T}(\alpha)\Big) \ge (1-\alpha) + o(1).$$

Moreover, the same arguments as for Proposition 4.4 show that

$$\mathbb{P}\Big(\min_{1\le\ell<\ell'\le N} \min_{\substack{i\in G_\ell,\\ j\in G_{\ell'}}} \widehat{\Psi}_{ij,T} \le q_{n,T}(\alpha)\Big) = o(1).$$

Taken together, these two statements imply that

$$\mathbb{P}\big(B_{n,T}\big) \ge (1-\alpha) + o(1). \tag{A.38}$$

In what follows, we show that on the event $B_{n,T}$, (i) $\{\widehat{G}^{[n-N]}_1, \ldots, \widehat{G}^{[n-N]}_N\} = \{G_1, \ldots \ldots, G_N\}$ and (ii) $\widehat{N} = N$. From (i), (ii) and (A.38), the statements of Proposition 5.1 easily follow.

Proof of (i). Suppose we are on the event $B_{n,T}$. The proof proceeds by induction on the iteration steps r of the HAC algorithm.

Base case ($r = 0$): In the first iteration step, the HAC algorithm merges two singleton clusters $\widehat{G}^{[0]}_i = \{i\}$ and $\widehat{G}^{[0]}_j = \{j\}$ with i and j belonging to the same group G_k. This is a direct consequence of (A.37). The algorithm thus produces a partition $\{\widehat{G}^{[1]}_1, \ldots, \widehat{G}^{[1]}_{n-1}\}$

whose elements $\widehat{G}_\ell^{[1]}$ all have the following property: $\widehat{G}_\ell^{[1]} \subseteq G_k$ for some k, that is, each cluster $\widehat{G}_\ell^{[1]}$ contains elements from only one group.

Induction step ($r \curvearrowright r+1$): Now suppose we are in the r-th iteration step for some $r < n-N$. Assume that the partition $\{\widehat{G}_1^{[r]}, \ldots, \widehat{G}_{n-r}^{[r]}\}$ is such that for any ℓ, $\widehat{G}_\ell^{[r]} \subseteq G_k$ for some k. Because of (A.37), the dissimilarity $\widehat{\Delta}(\widehat{G}_\ell^{[r]}, \widehat{G}_{\ell'}^{[r]})$ gets minimal for two clusters $\widehat{G}_\ell^{[r]}$ and $\widehat{G}_{\ell'}^{[r]}$ with the property that $\widehat{G}_\ell^{[r]} \cup \widehat{G}_{\ell'}^{[r]} \subseteq G_k$ for some k. Hence, the HAC algorithm produces a partition $\{\widehat{G}_1^{[r+1]}, \ldots, \widehat{G}_{n-(r+1)}^{[r+1]}\}$ whose elements $\widehat{G}_\ell^{[r+1]}$ are all such that $\widehat{G}_\ell^{[r+1]} \subseteq G_k$ for some k.

The above induction argument shows the following: For any $r \le n-N$, the partition $\{\widehat{G}_1^{[r]}, \ldots, \widehat{G}_{n-r}^{[r]}\}$ consists of clusters $\widehat{G}_\ell^{[r]}$ which all have the property that $\widehat{G}_\ell^{[r]} \subseteq G_k$ for some k. This in particular holds for the partition $\{\widehat{G}_1^{[n-N]}, \ldots, \widehat{G}_N^{[n-N]}\}$, which implies that $\{\widehat{G}_1^{[n-N]}, \ldots, \widehat{G}_N^{[n-N]}\} = \{G_1, \ldots, G_N\}$. $\square$

Proof of (ii). To start with, consider any partition $\{\widehat{G}_1^{[n-r]}, \ldots, \widehat{G}_r^{[n-r]}\}$ with $r < N$ elements. Such a partition must contain at least one element $\widehat{G}_\ell^{[n-r]}$ with the following property: $\widehat{G}_\ell^{[n-r]} \cap G_k \neq \emptyset$ and $\widehat{G}_\ell^{[n-r]} \cap G_{k'} \neq \emptyset$ for some $k \neq k'$. On the event $B_{n,T}$, it obviously holds that $\widehat{\Delta}(S) > q_{n,T}(\alpha)$ for any S with the property that $S \cap G_k \neq \emptyset$ and $S \cap G_{k'} \neq \emptyset$ for some $k \neq k'$. Hence, we can infer that on the event $B_{n,T}$, $\max_{1 \le \ell \le r} \widehat{\Delta}(\widehat{G}_\ell^{[n-r]}) > q_{n,T}(\alpha)$ for any $r < N$.
Next consider the partition $\{\widehat{G}_1^{[n-r]}, \ldots, \widehat{G}_r^{[n-r]}\}$ with $r = N$ and suppose we are on the event $B_{n,T}$. From (i), we already know that $\{\widehat{G}_1^{[n-N]}, \ldots, \widehat{G}_N^{[n-N]}\} = \{G_1, \ldots, G_N\}$. Moreover, it is easy to see that $\widehat{\Delta}(G_\ell) \le q_{n,T}(\alpha)$ for any ℓ. Hence, we obtain that $\max_{1 \le \ell \le N} \widehat{\Delta}(\widehat{G}_\ell^{[n-N]}) = \max_{1 \le \ell \le N} \widehat{\Delta}(G_\ell) \le q_{n,T}(\alpha)$.
Putting everything together, we can conclude that on the event $B_{n,T}$,

$$\min \Big\{ r = 1, 2, \ldots \,\Big|\, \max_{1 \le \ell \le r} \widehat{\Delta}\big(\widehat{G}_\ell^{[n-r]}\big) \le q_{n,T}(\alpha) \Big\} = N,$$

that is, $\widehat{N} = N$. $\square$

Proof of Proposition 5.2

We consider the event

$$D_{n,T} = \Big\{ \widehat{\Phi}_{n,T} \le q_{n,T}(\alpha) \text{ and } \min_{1 \le \ell < \ell' \le N} \min_{\substack{i \in G_\ell, \\ j \in G_{\ell'}}} \widehat{\Psi}_{ij,T} > q_{n,T}(\alpha) \Big\},$$

where we write the statistic $\widehat{\Phi}_{n,T}$ as

$$\widehat{\Phi}_{n,T} = \max_{1 \le i < j \le n} \max_{(u,h) \in \mathcal{G}_T} \Big\{ \Big| \frac{\widehat{\psi}_{ij,T}(u,h) - \mathbb{E}\widehat{\psi}_{ij,T}(u,h)}{(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2)^{1/2}} \Big| - \lambda(h) \Big\}.$$

The event $D_{n,T}$ can be analysed by the same arguments as those applied to the event $B_{n,T}$ in the proof of Proposition **??**. In particular, analogous to (A.38) and statements (i) and (ii) in this proof, we can show that

$$\mathbb{P}\big(D_{n,T}\big) \ge (1-\alpha) + o(1) \tag{A.39}$$

and

$$D_{n,T} \subseteq \big\{\widehat{N} = N \text{ and } \widehat{G}_\ell = G_\ell \text{ for all } \ell\big\}. \tag{A.40}$$

Moreover, we have that

$$D_{n,T} \subseteq \bigcap_{1\le \ell<\ell'\le \widehat{N}} E_{n,T}(\ell,\ell'), \tag{A.41}$$

which is a consequence of the following observation: For all i, j and $(u,h) \in \mathcal{G}_T$ with

$$\Big|\frac{\widehat{\psi}_{ij,T}(u,h) - \mathbb{E}\widehat{\psi}_{ij,T}(u,h)}{(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2)^{1/2}}\Big| - \lambda(h) \le q_{n,T}(\alpha) \quad \text{and} \quad \Big|\frac{\widehat{\psi}_{ij,T}(u,h)}{(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2)^{1/2}}\Big| - \lambda(h) > q_{n,T}(\alpha),$$

it holds that $\mathbb{E}[\widehat{\psi}_{ij,T}(u,h)] \neq 0$, which in turn implies that $m_i(v) - m_j(v) \neq 0$ for some $v \in I_{u,h}$. From (A.40) and (A.41), we obtain that

$$D_{n,T} \subseteq \Big\{\bigcap_{1\le \ell<\ell'\le \widehat{N}} E_{n,T}(\ell,\ell')\Big\} \cap \big\{\widehat{N} = N \text{ and } \widehat{G}_\ell = G_\ell \text{ for all } \ell\big\} = E_{n,T}.$$

This together with (A.39) implies that $\mathbb{P}(E_{n,T}) \ge (1-\alpha) + o(1)$, thus completing the proof.