Multiscale Comparison
of Nonparametric Trend Curves

Marina Khismatullina[1]   Michael Vogt[2]

Erasmus University Rotterdam   Ulm University

We develop new econometric methods for the comparison of nonparametric time trends. In many applications, practitioners are interested in whether the observed time series all have the same time trend. Moreover, they would often like to know which trends are different and in which time intervals they differ. We design a multiscale test to formally approach these questions. Specifically, we develop a test which allows to make rigorous confidence statements about which time trends are different and where (that is, in which time intervals) they differ. Based on our multiscale test, we further develop a clustering algorithm which allows to cluster the observed time series into groups with the same trend. We derive asymptotic theory for our test and clustering methods. The theory is complemented by a simulation study and two applications to GDP growth data and house pricing data.

Key words: Multiscale statistics; nonparametric regression; time series errors; shape constraints; strong approximations; anti-concentration bounds.
JEL classifications: C12; C14; C23; C38.

1 Introduction

The comparison of time trends is an important topic in time series analysis and has a wide range of applications in economics, finance and other fields such as climatology. In economics, one may wish to compare trends in real GDP (Grier and Tullock, 1989) or trends in long-term interest rates (Christiansen and Pigott, 1997) across different countries. In finance, it is of interest to compare the volatility trends of different stocks (Nyblom and Harvey, 2000). In climatology, researchers are interested in comparing the trending behaviour of temperature time series across different spatial locations (Karoly and Wu, 2005).

In this paper, we develop new econometric methods for the comparison of time trends. Classically, time trends are modelled stochastically in econometrics; see e.g. Stock and Watson (1988). Recently, however, there has been a growing interest in econometric models with deterministic time trends; see Cai (2007), Atak et al. (2011), Robinson (2012), Chen et al. (2012), Zhang et al. (2012) and Hidalgo and Lee (2014) among others. Following this recent development, we consider a general panel model with deterministic

---

[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`.

time trends. We observe a panel of n time series $\mathcal{T}_i = \{(Y_{it}, \boldsymbol{X}_{it}) : 1 \leq t \leq T\}$ for $1 \leq i \leq n$, where Y_{it} are real-valued random variables and $\boldsymbol{X}_{it} = (X_{it,1}, \ldots, X_{it,d})^\top$ 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 $1 \leq t \leq T$, 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 the rescaled time interval $[0,1]$, α_i is a fixed effect error term and $\mathcal{E}_i = \{\varepsilon_{it} : 1 \leq t \leq T\}$ is a zero-mean stationary error process. A detailed description of model (1.1) together with the technical assumptions on the various model components can be found in Section 2.

An important question in many applications is whether the nonparametric time trends m_i in model (1.1) are all the same. This question can formally be addressed by a statistical test of the null hypothesis

$$H_0 : m_1 = m_2 = \ldots = m_n$$

against the alternative H_1: $m_i(u) \neq m_j(u)$ for some $u \in [0,1]$ and $i \neq j$. The problem of testing H_0 versus H_1 is well studied in the literature. In a restricted version of model (1.1) without covariates, it has been investigated in Härdle and Marron (1990), Hall and Hart (1990), Degras et al. (2012) and Chen and Wu (2019) among many others. Test procedures in a less restricted version of model (1.1) with covariates and homogeneous parameter vectors ($\beta_i = \beta$ for all i) have been derived, for instance, in Zhang et al. (2012) and Hidalgo and Lee (2014).

Even though many different approaches to test H_0 versus H_1 have been developed over the years, most of them have a serious shortcoming: they are very uninformative. More specifically, they only allow to test *whether* the trend curves are the same or not. But they do not give any information on *which* curves are different and *where* (that is, in which time intervals) they differ. The following example illustrates the importance of this point: Suppose a group of researchers wants to test whether the GDP trend is the same in ten different countries. If the researchers use a test which is uninformative in the above sense and this test rejects H_0, they can only infer that there are at least two countries with a different trend. They do not obtain any information on *which* countries have a different GDP trend and *where* (that is, in which time periods) the trends differ. However, this is exactly the information that is relevant in practice. Without it, it is hardly possible to get a good economic understanding of the situation at hand. The main aim of the present paper is to construct a test procedure which provides the required information. More specifically, we construct a test which allows to make rigorous confidence statements about (i) *which* trend curves are different and

(ii) *where* (that is, in which time intervals) they differ.

Very roughly speaking, our approach works as follows: Let $\mathcal{I}_{u,h} := [u-h, u+h] \subseteq [0,1]$ be the rescaled time interval with midpoint $u \in [0,1]$ and length $2h$. We call u the location and h the scale of the interval $\mathcal{I}_{u,h}$. For any given location-scale pair (u,h), let

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

be the hypothesis that the two trend curves m_i and m_j are identical on the interval $\mathcal{I}_{u,h}$. Our procedure simultaneously tests $H_0^{[i,j]}(u,h)$ for a wide range of different location-scale pairs (u,h) and all pairs (i,j). As it takes into account multiple scales h, it is called a multiscale method. Our main theoretical result shows that under suitable technical conditions, the developed multiscale test controls the familywise error rate, that is, the probability of wrongly rejecting at least one null hypothesis $H_0^{[i,j]}(u,h)$. As we will see, this allows us to make simultaneous confidence statements of the following form for a given significance level $\alpha \in (0,1)$:

> *With probability at least $1-\alpha$, the trends m_i and m_j differ on any interval $\mathcal{I}_{u,h}$ and for any pair (i,j) for which $H_0^{[i,j]}(u,h)$ is rejected.* (1.2)

Hence, as desired, we can make confidence statements about *which* trend curves are different and *where* they differ.

To the best of our knowledge, there are no other test procedures available in the literature which allow to make simultaneous confidence statements of the form (1.2) in the context of the general model (1.1). We are only aware of two other multiscale tests for the comparison of nonparametric time trends, both of which are restricted to a strongly simplified version of model (1.1) without covariates. The first one is a SiZer-type test by Park et al. (2009), for which theory has only been derived in the special case of $n = 2$ time series. The second one is a multiscale test by Khismatullina and Vogt (2021), which was developed to detect differences between epidemic time trends in the context of the COVID-19 pandemic. Notably, our multiscale test can be regarded as an extension of the approach in Khismatullina and Vogt (2021). In the present paper, we go beyond the quite limited model setting in Khismatullina and Vogt (2021) and develop multiscale inference methods in the general model framework (1.1) which allows to deal with a wide range of economic and financial applications. Moreover, we develop a clustering algorithm based on our multiscale test which allows to detect groups of time series with the same trend.

Our multiscale test is constructed step by step in Section 3, whereas its theoretical properties are laid out in Section 4. The clustering algorithm is introduced and investigated in Section 5. The proofs of all theoretical results are relegated to the technical Appendix. We complement the theoretical analysis of the paper by a simulation study

and two application examples in Sections 6 and 7. In the first application example, we use our test and clustering methods to examine GDP growth data from different OECD contries. We in particular test the hypothesis that there is a common GDP trend in these countries and we cluster the countries into groups with the same trend. In the second example, we examine a long record of housing price data from different countries and compare the price trends in these countries by our methods.

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 $\|\mathbf{v}\|_q = \left(\sum_{i=1}^m v_i^q \right)^{1/q}$ 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 $\mathcal{T}_i = \{(Y_{it}, \mathbf{X}_{it}) : 1 \le t \le T\}$ of length T for $1 \le i \le n$. Each time series $\mathcal{T}_i$ satisfies the model equation

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

for $1 \le t \le T$, where $\boldsymbol{\beta}_i$ is a $d \times 1$ vector of unknown parameters, $\mathbf{X}_{it} = (X_{it,1}, \ldots, X_{it,d})^\top$ 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 of

the trend function m_i in the presence of the intercept α_i. We allow α_i to be correlated with the covariates $\boldsymbol{X}_{it}$ in an arbitrary way. Hence, α_i can be regarded as a fixed effect error term. We do not impose any restrictions on the dependence between the fixed effects α_i across i. Similarly, the covariates $\boldsymbol{X}_{it}$ are allowed to be dependent across i in an arbitrary way. As a consequence, the n time series $\mathcal{T}_i$ in our panel can be correlated with each other in various ways. In contrast to α_i and $\boldsymbol{X}_{it}$, the error terms ε_{it} 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 grows with the time series length 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 $\mathcal{F}_{it} = (\ldots, \eta_{it-2}, \eta_{it-1}, \eta_{it})$, the variables η_{it} are i.i.d. across t, and $g_i : \mathbb{R}^\infty \to \mathbb{R}$ is a measurable function such that ε_{it} is well-defined. 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, $\sum_{s \geq 0} \delta_q(g_i, s)$ is finite and $\sum_{s \geq t} \delta_q(g_i, s) = O(t^{-\gamma}(\log t)^{-A})$ 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 $\varepsilon_{it}, \varepsilon_{it+1}, \ldots$ 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). 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 $\boldsymbol{X}_{it} = (X_{it,1}, \ldots, X_{it,d})^\top$ 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 $q' > \max\{4, \theta q\}$ with q from (C1) and θ specified in (C10) below.

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

(C6) For each i and j, it holds that $\sum_{s \geq 0} \delta_{q'}(h_{ij}, s)$ is finite and $\sum_{s \geq t} \delta_{q'}(h_{ij}, s) = O(t^{-\alpha})$ for some $\alpha > 1/2 - 1/q'$ with q' from (C4).

Assumption (C4) guarantees that the process $\{\boldsymbol{X}_{it} : 1 \leq t \leq T\}$ is stationary and causal for each i. Assumption (C6) restricts the serial dependence of the process $\{\boldsymbol{X}_{it} : 1 \leq t \leq T\}$ for each i in terms of the physical dependence measure.

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 Lipschitz continuous on $[0, 1]$, that is, $|m_i(v) - m_i(w)| \leq L|v - w|$ for all $v, w \in [0, 1]$ and some constant $L < \infty$. Moreover, they are normalised such that $\int_0^1 m_i(u) du = 0$ for each i.

Remark 2.1. *Conditions (C4)–(C6) can be relaxed to cover locally stationary regressors. For example, (C4) may 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 $h_{ij}(t; \cdot) : \mathbb{R}^\infty \to \mathbb{R}$ is a measurable function for each t 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 $q' > \max\{4, \theta q\}$ with q from (C1) and θ specified in (C10) below.*

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

3 The multiscale test

In this section, we develop a multiscale test for the comparison of the trend curves m_i in model (2.2). More specifically, we construct a multiscale test of the null hypothesis

$$H_0 : m_1 = m_2 = \ldots = m_n.$$

The test is designed to be as informative as possible: It does not only allow to say whether H_0 is violated. It also gives information on which types of violation occur. In particular, it allows to infer *which* trends are different and *where* (that is, in which time intervals) they differ.

Our strategy to construct the test is as follows: For any given location-scale point (u, h), let

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

be the hypothesis that m_i and m_j are identical on the interval $\mathcal{I}_{u,h} := [u - h, u + h] \subseteq [0, 1]$. $H_0^{[i,j]}(u, h)$ can be viewed as a local null hypothesis that characterises the behaviour of two trend functions locally on the interval $\mathcal{I}_{u,h}$, whereas H_0 is the global null hypothesis that is concerned with the comparison of all trends on the whole unit interval $[0, 1]$. We consider a large family of time intervals $\mathcal{I}_{u,h}$ that fully cover the unit interval. More formally, we consider all intervals $\mathcal{I}_{u,h}$ with $(u, h) \in \mathcal{G}_T$, where $\mathcal{G}_T$ is a set of location-scale points (u, h) with the property that $\bigcup_{(u,h)\in\mathcal{G}_T} \mathcal{I}_{u,h} = [0, 1]$. Technical conditions on the set $\mathcal{G}_T$ are given below. The global null H_0 can now be formulated as

$H_0 :$ The hypothesis $H_0^{[i,j]}(u, h)$ holds true for all intervals $\mathcal{I}_{u,h}$ with $(u, h) \in \mathcal{G}_T$ and for all $1 \le i < j \le n$.

This formulation shows that we can test H_0 by a procedure which simultaneously tests the local hypothesis $H_0^{[i,j]}(u, h)$ for all $1 \le i < j \le n$ and all $(u, h) \in \mathcal{G}_T$.

In what follows, we design such a simultaneous test procedure step by step. To start with, we introduce some auxiliary estimators in Section 3.1. We then construct the test statistics in Section 3.2 and set up the test procedure in Section 3.3. Finally, Section 3.4 explains how to implement the procedure in practice.

3.1 Preliminary steps

If the fixed effects α_i and the coefficient vectors $\boldsymbol{\beta}_i$ were known, our testing problem would be greatly simplified. In particular, we could consider the model

$$Y_{it}^{\circ} = m_i\left(\frac{t}{T}\right) + \varepsilon_{it} \qquad \text{with} \qquad Y_{it}^{\circ} := Y_{it} - \alpha_i - \boldsymbol{\beta}_i^{\top} \boldsymbol{X}_{it},$$

which is a standard nonparametric regression equation. As α_i and $\boldsymbol{\beta}_i$ are not observed in practice, we construct estimators $\widehat{\alpha}_i$ and $\widehat{\boldsymbol{\beta}}_i$ and replace the unknown variables Y_{it}° by the approximations

$$\widehat{Y}_{it} := Y_{it} - \widehat{\alpha}_i - \widehat{\boldsymbol{\beta}}_i^\top \boldsymbol{X}_{it}.$$

In what follows, we explain how to construct estimators of α_i and $\boldsymbol{\beta}_i$.

In order to estimate the parameter vector $\boldsymbol{\beta}_i$ for a given 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

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

where $\Delta\varepsilon_{it} = \varepsilon_{it} - \varepsilon_{it-1}$. Since m_i is Lipschitz by Assumption (C8), we can use the fact that $|m_i(\frac{t}{T}) - m_i(\frac{t-1}{T})| = O(\frac{1}{T})$ to get that

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

This suggests to estimate $\boldsymbol{\beta}_i$ by applying least squares methods to equation (3.1), treating $\Delta\boldsymbol{X}_{it}$ as the regressors and ΔY_{it} as the response variable. As a result, we obtain the least squares 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 $\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i = O_p(T^{-1/2})$ under our assumptions. Given $\widehat{\boldsymbol{\beta}}_i$, we next estimate the fixed effect term α_i by

$$\widehat{\alpha}_i = \frac{1}{T} \sum_{t=1}^{T} \big(Y_{it} - \widehat{\boldsymbol{\beta}}_i^\top \boldsymbol{X}_{it} \big). \tag{3.3}$$

This is a reasonable estimator of α_i for the following reason: For each i, it holds that

$$\widehat{\alpha}_i - \alpha_i = \big(\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i \big)^\top \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{X}_{it} + \frac{1}{T} \sum_{t=1}^{T} m_i\Big(\frac{t}{T}\Big) + \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{it} = O_p(T^{-1/2}),$$

since $\frac{1}{T}\sum_{i=1}^{T} \varepsilon_{it} = O_p(T^{-1/2})$ by the law of large numbers, $\frac{1}{T}\sum_{i=1}^{T} m_i(t/T) = O(T^{-1})$ due to Lipschitz continuity of m_i and the normalisation $\int_0^1 m_i(u)du = 0$, $\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{X}_{it} = O_p(1)$ by Chebyshev's inequality and $\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i = O_p(T^{-1/2})$.

In order to construct our multiscale test, we do not only require the variables $\widehat{Y}_{it}$ and thus the estimators $\widehat{\boldsymbol{\beta}}_i$ and $\widehat{\alpha}_i$. We also need an estimator of the long-run error variance

$\sigma_i^2 = \sum_{\ell=-\infty}^{\infty} \text{Cov}(\varepsilon_{i0}, \varepsilon_{i\ell})$ for each i. Throughout the paper, we assume that the long-run variance σ_i^2 does not depend on i, that is $\sigma_i^2 = \sigma^2$ for all i. For technical reasons, we nevertheless use a different estimator of $\sigma_i^2 = \sigma^2$ for each i. Let $\widehat{\sigma}_i^2$ be an estimator of σ_i^2 which is computed from the i-th time series $\{\widehat{Y}_{it} : 1 \leq t \leq T\}$, that is, let $\widehat{\sigma}_i^2 = \widehat{\sigma}_i^2(\widehat{Y}_{i1}, \ldots, \widehat{Y}_{iT})$ be a function of the variables $\widehat{Y}_{it}$ for $1 \leq t \leq T$. Our theory works with any estimator $\widehat{\sigma}_i^2$ which has the property that $\widehat{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$ for each i, where ρ_T slowly converges to zero, in particular, $\rho_T = o(1/\log T)$. We now briefly discuss two possible choices of $\widehat{\sigma}_i^2$.

Following Kim (2016), we can estimate σ_i^2 by a variant of the subseries variance estimator, which was first proposed by Carlstein (1986) and then extended by Wu and Zhao (2007). Formally, we set

$$\widehat{\sigma}_i^2 = \frac{1}{2(M-1)s_T} \sum_{m=1}^{M} \left[\sum_{t=1}^{s_T} \Big(Y_{i(t+ms_T)} - Y_{i(t+(m-1)s_T)} \right.$$
$$\left. - \widehat{\boldsymbol{\beta}}_i^\top \big(\boldsymbol{X}_{i(t+ms_T)} - \boldsymbol{X}_{i(t+(m-1)s_T)} \big) \Big) \right]^2, \qquad (3.4)$$

where s_T is the length of the 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 $s_T \asymp T^{1/3}$. When implementing $\widehat{\sigma}_i^2$, we in particular choose $s_T = \lfloor T^{1/3} \rfloor$. According to Lemma A.6 in the 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 subseries estimator $\widehat{\sigma}_i^2$ is completely nonparametric: it does not presuppose a specific time series model for the error process $\mathcal{E}_i = \{\varepsilon_{it} : 1 \leq t \leq T\}$ but merely requires $\mathcal{E}_i$ to fulfill general weak dependence conditions. In practice, however, it often makes sense to impose additional structure on the error process $\mathcal{E}_i$. A very popular yet general error model is the class of $\text{AR}(\infty)$ processes. A difference-based estimator of the long-run error variance for this error model has been developed in Khismatullina and Vogt (2020) and can be easily adapted to the situation at hand. A detailed discussion of the estimator and a comparison with other long-run variance estimators can be found in Section 4 of Khismatullina and Vogt (2020).

3.2 Construction of the test statistics

A statistic to test the local null hypothesis $H_0^{[i,j]}(u, h)$ for a given location-scale point (u, h) and a given pair of time series (i, j) can be constructed as follows: Let

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

be a weighted average of the differenced variables $\widehat{Y}_{it} - \widehat{Y}_{jt}$. In particular, let $w_{t,T}(u,h)$ be local linear kernel weights of the form

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

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 with the following property:

(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 $C_K > 0$.

The kernel average $\widehat{\psi}_{ij,T}(u,h)$ can be regarded as a measure of distance between the two trend curves m_i and m_j on the interval $\mathcal{I}_{u,h} = [u - h, u + h]$. We may thus use a normalized (absolute) version of the kernel average $\widehat{\psi}_{ij,T}(u,h)$, in particular, the term

$$\Big|\frac{\widehat{\psi}_{ij,T}(u,h)}{(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2)^{1/2}}\Big| \tag{3.7}$$

as a statistic to test the local null $H_0^{[i,j]}(u,h)$. Our aim is to test $H_0^{[i,j]}(u,h)$ simultaneously for a wide range of location-scale points (u,h) and all possible pairs of time series (i,j). To take into account that we are faced with a simultaneous test problem, we replace the statistics from (3.7) by additively corrected versions of the form

$$\widehat{\psi}_{ij,T}^0(u,h) = \Big|\frac{\widehat{\psi}_{ij,T}(u,h)}{(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2)^{1/2}}\Big| - \lambda(h), \tag{3.8}$$

where $\lambda(h) = \sqrt{2\log\{1/(2h)\}}$. The scale-dependent correction term $\lambda(h)$ was first introduced in the multiscale approach of Dümbgen and Spokoiny (2001) and has been used in various contexts since then. We refer to Khismatullina and Vogt (2020) for a detailed discussion of the idea behind this additive correction.

In order to test the global null hypothesis H_0, we aggregate the test statistics $\widehat{\psi}_{ij,T}^0(u,h)$ by taking their maximum over all location-scale points $(u,h) \in \mathcal{G}_T$ and all pairs of time series (i,j) with $1 \leq i < j \leq n$. This leads to the multiscale test statistic

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

For our theoretical results, we suppose that the set of location-scale points $\mathcal{G}_T$ is a subset of $\mathcal{G}_T^{\text{full}} = \{(u,h) : \mathcal{I}_{u,h} = [u - h, u + h] \subseteq [0,1] \text{ with } u = t/T \text{ and } h = s/T \text{ for}$

some $1 \le t, s \le T$ and $h \in [h_{\min}, h_{\max}]\}$ which fulfills the following conditions:

(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 (C1) and $h_{\max} = o(1)$.

Assumptions (C10) and (C11) place relatively mild restrictions on the set $\mathcal{G}_T$: (C10) allows $\mathcal{G}_T$ to be very large compared to the sample size T. In particular, $\mathcal{G}_T$ may grow as any polynomial of T. (C11) allows $\mathcal{G}_T$ to contain intervals $[u-h, u+h]$ of many different scales h, ranging from very small intervals of scale $h_{\min}$ to substantially larger intervals of scale $h_{\max}$.

3.3 The test procedure

Let $\mathcal{M} = \{(u, h, i, j) : (u, h) \in \mathcal{G}_T \text{ and } 1 \le i < j \le n\}$ be the collection of all location-scale points (u, h) and all pairs of time series (i, j) under consideration. Moreover, let $\mathcal{M}_0 \subseteq \mathcal{M}$ be the set of tuples (u, h, i, j) for which $H_0^{[i,j]}(u, h)$ is true. For a given significance level $\alpha \in (0, 1)$, our multiscale test is carried out as follows:

(i) For each tuple $(u, h, i, j) \in \mathcal{M}$, we reject the local null hypothesis $H_0^{[i,j]}(u, h)$ if

$$\widehat{\psi}_{ij,T}^0(u, h) > q_{n,T}(\alpha),$$

where the critical value $q_{n,T}(\alpha)$ is constructed below such that the familywise error rate (FWER) is controlled at level α. As usual, the FWER is defined as the probability of wrongly rejecting $H_0^{[i,j]}(u, h)$ for at least one tuple $(u, h, i, j) \in \mathcal{M}$. More formally, it is defined as

$$\text{FWER}(\alpha) = \mathbb{P}\Big(\exists (u, h, i, j) \in \mathcal{M}_0 : \widehat{\psi}_{ij,T}^0(u, h) > q_{n,T}(\alpha)\Big)$$

for a given significance level $\alpha \in (0, 1)$ and we say that the FWER is controlled at level α if $\text{FWER}(\alpha) \le \alpha$.

(ii) We reject the global null hypothesis $H_0 : m_1 = m_2 = \ldots = m_n$ if at least one local null hypothesis is rejected. Put differently, we reject H_0 if

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

We now construct a critical value $q_{n,T}(\alpha)$ which controls the FWER at level α. Let $q_{n,T}(\alpha)$ be the $(1-\alpha)$-quantile of the multiscale statistic $\widehat{\Psi}_{n,T}$ under the global null

11

$H_0 : m_1 = m_2 = \ldots = m_n$. Since

$$\text{FWER}(\alpha) = 1 - \mathbb{P}\Big(\forall (u, h, i, j) \in \mathcal{M}_0 : \widehat{\psi}^0_{ij,T}(u, h) \leq q_{n,T}(\alpha)\Big)$$

$$= 1 - \mathbb{P}\Big(\max_{(u,h,i,j) \in \mathcal{M}_0} \widehat{\psi}^0_{ij,T}(u, h) \leq q_{n,T}(\alpha)\Big)$$

$$\leq 1 - \mathbb{P}_{H_0}\Big(\max_{(u,h,i,j) \in \mathcal{M}} \widehat{\psi}^0_{ij,T}(u, h) \leq q_{n,T}(\alpha)\Big) \leq \alpha$$

with $\mathbb{P}_{H_0}$ denoting the probability under H_0, this choice of $q_{n,T}(\alpha)$ indeed controls the FWER at the desired level. However, this choice is not computable in practice. We thus replace it by a suitable approximation: Suppose we could perfectly estimate $\boldsymbol{\beta}_i$ and σ_i^2, that is, $\widehat{\boldsymbol{\beta}}_i = \boldsymbol{\beta}_i$ and $\widehat{\sigma}_i^2 = \sigma_i^2$ for all i. Then under H_0, the test statistics $\widehat{\psi}^0_{ij,T}(u, h)$ would simplify to

$$\widehat{\psi}^0_{ij,T}(u, h) = \left| \frac{\sum_{t=1}^{T} w_{t,T}(u, h)\{(\varepsilon_{it} - \bar{\varepsilon}_i) - (\varepsilon_{jt} - \bar{\varepsilon}_j)\}}{(\sigma_i^2 + \sigma_j^2)^{1/2}} \right| - \lambda(h), \tag{3.9}$$

where $\bar{\varepsilon}_i = \bar{\varepsilon}_{i,T} := T^{-1} \sum_{t=1}^{T} \varepsilon_{it}$ denotes the empirical average of the variables $\varepsilon_{i1}, \ldots, \varepsilon_{iT}$. Replacing the error terms ε_{it} in (3.9) by normally distributed variables leads to the statistics

$$\phi^0_{ij,T}(u, h) = \left| \frac{\phi_{ij,T}(u, h)}{(\sigma_i^2 + \sigma_j^2)^{1/2}} \right| - \lambda(h) \tag{3.10}$$

with

$$\phi_{ij,T}(u, h) = \sum_{t=1}^{T} w_{t,T}(u, h) \left\{ \sigma_i(Z_{it} - \bar{Z}_i) - \sigma_j(Z_{jt} - \bar{Z}_j) \right\}, \tag{3.11}$$

where Z_{it} are independent standard normal random variables for $1 \leq t \leq T$ and $1 \leq i \leq n$ and, as before, we use the shorthand $\bar{Z}_i = \bar{Z}_{i,T} := T^{-1} \sum_{t=1}^{T} Z_{it}$. With this notation, we define

$$\Phi_{n,T} = \max_{1 \leq i < j \leq n} \max_{(u,h) \in \mathcal{G}_T} \phi^0_{ij,T}(u, h), \tag{3.12}$$

which can be regarded as a Gaussian version of the test statistic $\widehat{\Psi}_{n,T}$ under the null H_0 (in the idealized case with $\widehat{\boldsymbol{\beta}}_i = \boldsymbol{\beta}_i$ and $\widehat{\sigma}_i^2 = \sigma_i^2$ for all i). We now choose $q_{n,T}(\alpha)$ to be the $(1 - \alpha)$-quantile of $\Phi_{n,T}$.

Remark 3.1. *Importantly, this choice of $q_{n,T}(\alpha)$ can be computed by Monte Carlo simulations in practice: under 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 statistics from (3.10) as

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

This shows that the distribution of these statistics and thus the distribution of the multiscale statistic $\Phi_{n,T}$ only depend on the Gaussian variables Z_{it} for $1 \leq i \leq n$ and $1 \leq t \leq T$. Consequently, we can approximate the distribution of $\Phi_{n,T}$ – and in particular its quantiles $q_{n,T}(\alpha)$ – by simulating values of the Gaussian random variables Z_{it}. In Section 3.4, we explain in detail how to compute Monte Carlo approximations of the quantiles $q_{n,T}(\alpha)$.

3.4 Implementation of the test in practice

In practice, we implement the test procedure as follows for a given significance level $\alpha \in (0,1)$:

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

Step 2. For each (i,j) with $1 \leq i < j \leq n$ and each $(u,h) \in \mathcal{G}_T$, reject the local null hypothesis $H_0^{[i,j]}(u,h)$ if $\widehat{\psi}_{ij,T}^0(u,h) > \widehat{q}_{n,T}(\alpha)$. Reject the global null hypothesis H_0 if at least one local hypothesis $H_0^{[i,j]}(u,h)$ is rejected.

Step 3. Display the test results as follows. For each pair of time series (i,j), let $\mathcal{S}^{[i,j]}(\alpha)$ be the set of intervals $\mathcal{I}_{u,h}$ for which we reject $H_0^{[i,j]}(u,h)$. Produce a separate plot for each pair (i,j) which displays the intervals in $\mathcal{S}^{[i,j]}(\alpha)$. This gives a graphical overview over the intervals where our tests finds a deviation from the null.

Remark 3.2. *In some cases, the number of intervals in $\mathcal{S}^{[i,j]}(\alpha)$ may be quite large, rendering the visual representation of the test results outlined in Step 3 cumbersome. To overcome this drawback, we replace $\mathcal{S}^{[i,j]}(\alpha)$ with the subset of minimal intervals $\mathcal{S}_{min}^{[i,j]}(\alpha) \subseteq \mathcal{S}^{[i,j]}(\alpha)$. 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 $\mathcal{I}_{u',h'} \subset \mathcal{I}_{u,h}$. Notably, our theoretical results remain to hold true when the sets $\mathcal{S}^{[i,j]}(\alpha)$ are replaced by $\mathcal{S}_{min}^{[i,j]}(\alpha)$. See Section 4 for the details.*

4 Theoretical properties of the multiscale test

To start with, we investigate 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

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

with $\bar{\varepsilon}_i = \bar{\varepsilon}_{i,T} := T^{-1} \sum_{t=1}^{T} \varepsilon_{it}$ 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 $H_0^{[i,j]}(u,h)$. Hence, the auxiliary statistic $\widehat{\Phi}_{n,T}$ is identical to the multiscale test statistic $\widehat{\Psi}_{n,T}$ under the global null H_0. Consequently, in order to determine the distribution of the test statistic $\widehat{\Psi}_{n,T}$ under H_0, it suffices to 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 (3.12).

Theorem 4.1. *Let (C1)–(C11) be fulfilled. Moreover, assume that $\widehat{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$ with $\rho_T = o(1/\log T)$ and $\sigma_i^2 = \sigma^2$ for all i. Then*

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

Theorem 4.1 is key 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 for dependent processes (Berkes et al., 2014) and anti-concentration bounds for Gaussian random vectors (Nazarov, 2003). It can be regarded as a further development of the proof strategy in Khismatullina and Vogt (2020) who developed multiscale methods to test for local increases/decreases of the nonparametric trend function m in the univariate time series model $Y_t = m(t/T) + \varepsilon_t$. We extend their theoretical results in several directions: we consider the case of multiple time series and work with a more general model which includes covariates and a flexible fixed effect error structure.*

With the help of Theorem 4.1, we now examine the theoretical properties of our multiscale test developed in Section 3. The following proposition shows that our test of the

global null H_0 has correct (asymptotic) size.

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

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

The next proposition characterises the power of the test against 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 time series length T 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 $m_{i,T}(w) - m_{j,T}(w) \geq c_T \sqrt{\log T/(Th)}$ for all $w \in [u - h, u + h]$ or $m_{j,T}(w) - m_{i,T}(w) \geq c_T \sqrt{\log T/(Th)}$ for all $w \in [u - h, u + h]$, where $\{c_T\}$ is any sequence of positive numbers with $c_T \to \infty$. Then*

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

We now turn attention to the local null hypotheses $H_0^{[i,j]}(u, h)$. As already defined above, let $\mathcal{M} = \{(u, h, i, j) : (u, h) \in \mathcal{G}_T \text{ and } 1 \leq i < j \leq n\}$ be the collection of location-scale points (u, h) and pairs of time series (i, j) under consideration. Moreover, let $\mathcal{M}_0 \subseteq \mathcal{M}$ be the set of tuples (u, h, i, j) for which $H_0^{[i,j]}(u, h)$ is true. Since we test $H_0^{[i,j]}(u, h)$ simultaneously for all $(u, h, i, j) \in \mathcal{M}$, we would like to bound the probability of making at least one false discovery. More formally, we would like to control the family-wise error rate

$$\text{FWER}(\alpha) = \mathbb{P}\Big(\exists\, (u, h, i, j) \in \mathcal{M}_0 : \widehat{\psi}_{ij,T}^0(u, h) > q_{n,T}(\alpha)\Big)$$

$$= \mathbb{P}\Big(\exists\, (u, h, i, j) \in \mathcal{M}_0 : \mathcal{I}_{u,h} \in \mathcal{S}^{[i,j]}(\alpha)\Big)$$

at level α, where $\mathcal{S}^{[i,j]}(\alpha)$ is the set of intervals $\mathcal{I}_{u,h}$ for which the null hypothesis $H_0^{[i,j]}(u, h)$ is rejected. The following result shows that our test procedure asymptotically controls the FWER at level α.

Proposition 4.5. *Let the conditions of Theorem 4.1 be satisfied. Then for any given $\alpha \in (0, 1)$,*

$$\text{FWER}(\alpha) \leq \alpha + o(1).$$

The statement of Proposition 4.5 can obviously be reformulated as follows:

$$\mathbb{P}\Big(\forall\, (u, h, i, j) \in \mathcal{M} \text{ for which } H_0^{[i,j]}(u, h) \text{ is true, } H_0^{[i,j]}(u, h) \text{ is not rejected}\Big)$$

$$\geq 1 - \alpha + o(1).$$

Consequently, we can make the following simultaneous confidence statement:

> *With (asymptotic) probability at least $1 - \alpha$, the trends m_i and m_j differ on any interval $\mathcal{I}_{u,h}$ and for all pairs (i,j) for which $H_0^{[i,j]}(u,h)$ is rejected.* (4.2)

Put differently:

> *With (asymptotic) probability at least $1 - \alpha$, for any pair (i,j) with $1 \leq i < j \leq n$, the two trends m_i and m_j differ on all intervals $\mathcal{I}_{u,h} \in \mathcal{S}^{[i,j]}(\alpha)$.* (4.3)

Remark 4.6. *According to (4.3), the graphical device introduced in Section 3.4 which depicts the intervals in $\mathcal{S}^{[i,j]}(\alpha)$ comes with the following statistical guarantee: we can claim with (asymptotic) confidence at least $1 - \alpha$ that the trends m_i and m_j differ on all the depicted intervals. As $\mathcal{S}_{min}^{[i,j]}(\alpha) \subseteq \mathcal{S}^{[i,j]}(\alpha)$ for all i and j, this guarantee remains to hold true when the sets $\mathcal{S}^{[i,j]}(\alpha)$ are replaced by the sets of minimal intervals $\mathcal{S}_{min}^{[i,j]}(\alpha)$.*

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 $N \leq n$ and $\{1, \ldots, n\} = \dot{\bigcup}_{\ell=1}^{N} G_\ell$ such that for each $1 \leq \ell \leq N$,

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

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:

(C12) 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, $|f_{\ell,T}(v) - f_{\ell,T}(w)| \leq L|v - w|$ 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 $f_{\ell',T}$ differ in the following sense: There exists $(u,h) \in \mathcal{G}_T$ with $\mathcal{I}_{u,h} \subseteq [0,1]$ such that $f_{\ell,T}(w) - f_{\ell',T}(w) \geq c_T\sqrt{\log T/(Th)}$ for all $w \in \mathcal{I}_{u,h}$ or $f_{\ell',T}(w) - f_{\ell,T}(w) \geq c_T\sqrt{\log T/(Th)}$ 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 multiscale statistics

$\max_{(u,h)\in\mathcal{G}_T}\hat{\psi}_{ij,T}^0(u,h)$ calculated for each i and j 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

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

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

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

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

and merge them into a new cluster.

Iterating this procedure for $r = 1,\dots,n-1$ yields a tree of nested partitions $\{\widehat{G}_1^{[r]},\dots\dots,\widehat{G}_{n-r}^{[r]}\}$, which can be graphically represented by a dendrogram. Roughly speaking, the HAC algorithm merges the n singleton clusters $\widehat{G}_i^{[0]} = \{i\}$ 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 $\{G_1,\dots,G_N\}$ by the N-partition $\{\widehat{G}_1^{[n-N]},\dots,\widehat{G}_N^{[n-N]}\}$ produced by the HAC algorithm. When N is unknown, we estimate it by the $\widehat{N}$-partition $\{\widehat{G}_1^{[n-\widehat{N}]},\dots,\widehat{G}_{\widehat{N}}^{[n-\widehat{N}]}\}$, where $\widehat{N}$ is an estimator of N. The latter is defined as

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

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 $\{\widehat{G}_1,\dots,\widehat{G}_{\widehat{N}}\}$, 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 and (C12) be satisfied. Then*

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

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 at least $1-\alpha$ that the estimated group structure is identical to the true group structure. 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 $G_1,\ldots,G_N$ and their number 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 $P(\widehat{G}_\ell = G_\ell$ 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 $1 \le \ell < \ell' \le \widehat{N}$. An interval $\mathcal{I}_{u,h}$ is contained in $\mathcal{S}_{n,T}^{[\ell,\ell']}(\alpha)$ 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, $\mathcal{I}_{u,h} \in \mathcal{S}_{n,T}^{[\ell,\ell']}(\alpha)$ 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

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

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

Proposition 5.2. *Under the conditions of Proposition 5.1, the event*

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

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

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

According to Proposition 5.2, the sets $\mathcal{S}_{n,T}^{[\ell,\ell']}(\alpha)$ allow us to locate, with a pre-specified confidence, time intervals where the group-specific trend functions f_ℓ and $f_{\ell'}$ differ from each other. In particular, we can claim with asymptotic confidence at least $1 - \alpha$ that the trend functions f_ℓ and $f_{\ell'}$ differ on all intervals $\mathcal{I}_{u,h} \in \mathcal{S}_{n,T}^{[\ell,\ell']}(\alpha)$.[3]

6   Simulations

In this section, we investigate our testing and clustering methods by means of a simulation study. The simulation design is as follows: We generate data from the model $Y_{it} = m_i(\frac{t}{T}) + \beta_i X_{it} + \alpha_i + \varepsilon_{it}$, where the number of time series i is set to $n = 15$ and we consider different time series lengths T. For simplicity, we assume that the fixed effect term α_i is equal to 0 in all time series and we only include a single regressor X_{it} in the model, setting the unknown parameter β_i to 1 for all i. For each i, the errors ε_{it} follow the AR(1) model $\varepsilon_{it} = a\varepsilon_{i,t-1} + \eta_{it}$, where $a = 0.25$ and the innovations η_{it} are i.i.d. normally distributed with mean 0 and variance 0.25. Similarly, for each i, the covariates X_{it} follow an AR(1) process of the form $X_{it} = a_x X_{i,t-1} + \zeta_{it}$, where $a_x = 0.5$ and the innovations ζ_{it} are i.i.d. normally distributed with mean 0 and variance 1. We assume that the covariates X_{it} and the error terms ε_{js} are independent from each other for all $1 \leq i, j \leq n$ and $1 \leq t, s \leq T$. 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 $m_1(u) = b(u - 0.5)$ with $b \in \{0.75, 1, 1.25\}$ 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. Note that the normalisation constraint $\int_0^1 m_1(u)du = 0$ is directly satisfied in this case. For each simulation exercise, we simulate 5000 data samples.

Our multiscale test is implemented as follows: The estimators $\widehat{\beta}_i$ and $\widehat{\alpha}_i$ of the unknown parameters β_i and α_i are computed as described in Section 3.1. Since the errors ε_{it} follow an AR(1) process, we estimate the long-run error variance σ_i^2 by the difference-based estimator proposed in Khismatullina and Vogt (2020), setting the tuning parameters q and r to 25 and 10, respectively. In order to construct our test statistics, we use an Epanechnikov kernel and the grid $\mathcal{G}_T = U_T \times H_T$ with

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

---

[3]This statement remains to hold true when the sets of intervals $\mathcal{S}_{n,T}^{[\ell,\ell']}(\alpha)$ are replaced by the corresponding sets of minimal intervals.

Table 1: Size of the multiscale test for different sample sizes T and nominal sizes α.

	nominal size α		
T	0.01	0.05	0.1
100	0.009	0.045	0.087
250	0.013	0.063	0.117
500	0.013	0.057	0.112

Table 2: Power of the multiscale test for different sample sizes T and nominal sizes α. Each panel corresponds to a different slope parameter b.

(a) $b = 0.75$

	nominal size α		
T	0.01	0.05	0.1
100	0.033	0.122	0.199
250	0.209	0.434	0.549
500	0.741	0.891	0.947

(b) $b = 1.00$

	nominal size α		
T	0.01	0.05	0.1
100	0.105	0.270	0.376
250	0.635	0.840	0.901
500	0.994	0.999	0.999

(c) $b = 1.25$

	nominal size α		
T	0.01	0.05	0.1
100	0.275	0.512	0.628
250	0.933	0.986	0.993
500	1.000	1.000	1.000

We thus consider intervals $\mathcal{I}_{u,h} = [u - h, u + h]$ which contain $5, 15, 25, \ldots$ data points. The critical value $q_{n,T}(\alpha)$ of our test is computed by Monte Carlo methods as described in Section 3.4, where we set $L = 5000$.

The simulation results for our multiscale test can be found in Table 1, which reports its actual size under H_0, and in Table 2, which reports its power against different alternatives. Both the actual size and the power are computed as the number of simulations in which the test rejects the global null H_0 divided by the total number of simulations. Inspecting Table 1, one can see that the actual size is fairly close to the nominal target α in all the considered scenarios. Hence, the test has approximately the correct size. Inspecting Table 2, one can further see that the test has reasonable power properties. For the smallest slope $b = 0.75$ and the smallest time series length $T = 100$, the power is only moderate, reflecting the fact that the alternative with $b = 0.75$ is not very far away from the null. However, as we increase the slope b and the sample size T, the power increases quickly. Already for the slope value $b = 1.00$, we reach a power of 0.99 for $T = 500$ and for all nominal sizes α.

We next investigate the finite sample performance of the clustering algorithm from Section 5. To do so, we consider a very simple scenario: we generate data from the model $Y_{it} = m_i(\frac{t}{T}) + \varepsilon_{it}$, that is, we assume that there are no fixed effects and no covariates. The error terms ε_{it} are specified as above. Moreover, as before, we set the number of time series to $n = 15$ and we consider different time series lengths T. 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 $G_3 = \{11, \ldots, 15\}$, and we assume that $m_i = f_l$ for all $i \in G_l$ and all $l = 1, 2, 3$. The group-specific trend functions f_1, f_2 and f_3 are defined as $f_1(u) = 0$, $f_2(u) = 1 \cdot (u - 0.5)$ and $f_3(u) = (-1) \cdot (u - 0.5)$.

20

Table 3: 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
100	0.055	0.188	0.298
250	0.713	0.922	0.939
500	0.994	0.979	0.956

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

	nominal size α		
T	0.01	0.05	0.1
100	0.009	0.045	0.077
250	0.640	0.825	0.845
500	0.992	0.978	0.956

In order to estimate 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 simulation results are reported in Table 3. The entries in Table 3a 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 3b 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 results in Table 3 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 largest sample size $T = 500$, the empirical probabilities reported in Table 3 can indeed be seen to exceed the value $(1 - \alpha)$ as predicted by Proposition 5.1. For the smaller sample sizes $T = 100$ and $T = 250$, in contrast, the empirical probabilities are substantially smaller than $(1 - \alpha)$. This reflects the asymptotic nature of Proposition 5.1 and is not very surprising. It simply mirrors the fact that for the smaller sample sizes $T = 100$ and $T = 250$, the effective noise level in the simulated data is quite high.

Figures 1 and 2 give more insight into what happens for $T = 100$ and $T = 250$. Figure 1 shows histograms of the 5000 simulated values of $\widehat{N}$, while Figure 2 depicts histograms of the number of classification errors produced by our algorithm. By the number of classification errors, we simply mean the number of incorrectly classified time series, which is formally calculated as

$$\min_{\pi \in S_{\widehat{N}}} \left\{ |G_1 \setminus \widehat{G}_{\pi(1)}| + |G_2 \setminus \widehat{G}_{\pi(2)}| + |G_3 \setminus \widehat{G}_{\pi(3)}| \right\}$$

with $S_{\widehat{N}}$ being the set of all permutations of $\{1, 2, \ldots, \widehat{N}\}$. The histogram of Figure 1 for $T = 100$ clearly shows that our method underestimates the number of groups ($\widehat{N} = 2$ in 4055 cases out of 5000). In particular, it fails to detect the difference between two out of three groups in a large number of simulations. This is reflected in the corresponding histogram of Figure 2 which shows that there are exactly 5 classification errors in 3924 of the 5000 simulation runs. In most of these cases, the estimated group structure $\{\widehat{G}_1, \widehat{G}_2\}$ coincides with either $\{G_1 \cup G_2, G_3\}$, $\{G_1, G_2 \cup G_3\}$ or $\{G_1 \cup G_3, G_2\}$. In summary, we
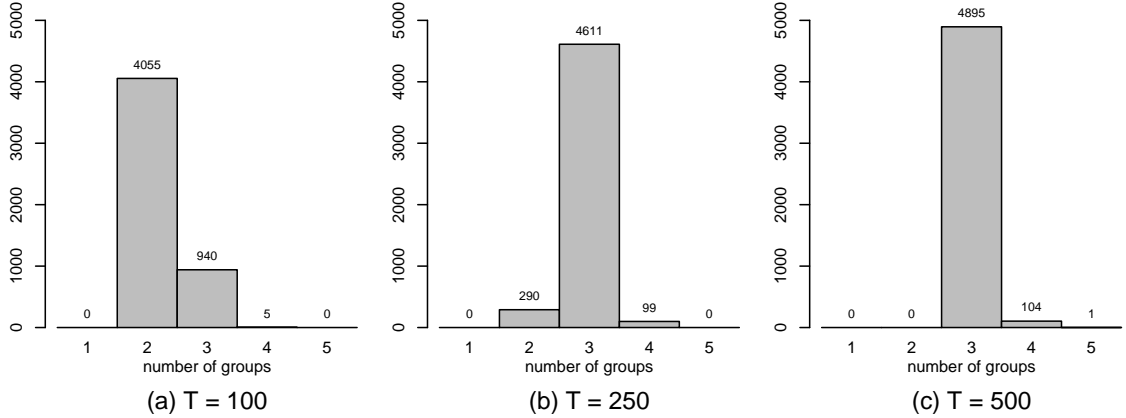
Figure 1: Estimated number of groups $\widehat{N}$ for nominal size $\alpha = 0.05$. Each panel corresponds to a different sample size T.

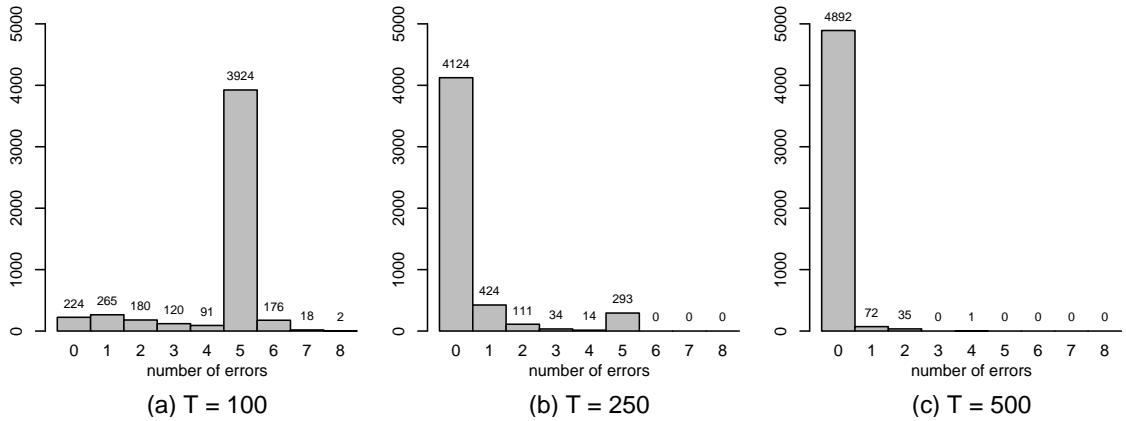

Figure 2: Number of classification errors for nominal size $\alpha = 0.05$. Each panel corresponds to a different sample size T.

can conclude that the small empirical probabilities for $T = 100$ in Table 3 are due to the algorithm underestimating the number of groups. Inspecting the histograms for $T = 250$, one can see that the performance of the estimators $\widehat{N}$ and $\{\widehat{G}_1, \ldots, \widehat{G}_{\widehat{N}}\}$ improves significantly, even though the corresponding empirical probabilities in Table 3 are still somewhat below the target $(1 - \alpha)$.

7 Applications

7.1 Analysis of GDP growth

In what follows, we revisit an application example from Zhang et al. (2012). The aim is to test the hypothesis of a common trend in the GDP growth time series of several OECD countries. Since we do not have access to the original dataset of Zhang et al. (2012) and we do not know the exact data specifications used there, we work with data from the following common sources: Refinitiv Datastream, the OECD.Stat

22

database, Federal Reserve Economics Data (FRED) and the Barro-Lee Educational Attainment dataset (Barro and Lee, 2013). We consider a data specification that is as close as possible to the one in Zhang et al. (2012) with one important distinction: In the original study, the authors examined 16 OECD countries (not specifying which ones) over the time period from the fourth quarter of 1975 up to and including the third quarter of 2010, whereas we consider only 11 countries (Australia, Austria, Canada, Finland, France, Germany, Japan, Norway, Switzerland, UK and USA) over the same time span. The reason is that we have access to data of good quality only for these 11 countries. In the following list, we specify the data for our analysis.[4]

- **Gross domestic product (GDP):** We use freely available data on *Gross Domestic Product – Expenditure Approach* from the OECD.Stat database (`https://stats.oecd.org/Index.aspx`). To be as close as possible to the specification of the data in Zhang et al. (2012), we use seasonally adjusted quarterly data on GDP expressed in millions of 2015 US dollars.[5]

- **Capital (K):** We use data on *Gross Fixed Capital Formation* from the OECD.Stat database. The data are at a quarterly frequency, seasonally adjusted, and expressed in millions of 2015 US dollars. In contrast to Zhang et al. (2012), who use data on *Capital Stock at Constant National Prices*, we choose to work with gross fixed capital formation due to data availability. 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 is standard in the literature; see e.g. Sharma and Dhakal (1994), Lee and Huang (2002) and Lee (2005).

- **Labour (L):** We collect data on the *Number of Employed People* from various sources. For most of the countries (Austria, Australia, Canada, Germany, Japan, UK and USA), we download the OECD data on *Employed Population: Aged 15 and Over* retrieved from FRED (`https://fred.stlouisfed.org/`). The data for France and Switzerland were downloaded from Refinitiv Datastream. 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, 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 in particular use the default method of the function `seas` from the R package `seasonal` (Sax and Eddelbuettel, 2018) which is an interface to X-13-ARIMA-SEATS, the seasonal adjustment software used by the US Census Bureau. For all of the countries, the data are given in thousands of persons.

---

[4]All data were accessed and downloaded on 7 December 2021.

[5]Since the publication of Zhang et al. (2012), 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.

- **Human capital (H):** We use *Educational Attainment for Population Aged 25 and Over* collected from `http://www.barrolee.com` 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.

For each of the $n = 11$ countries in our sample, we thus observe a quarterly time series $\mathcal{T}_i = \{(Y_{it}, \boldsymbol{X}_{it}) : 1 \leq t \leq T\}$ of length $T = 140$, where $Y_{it} = \Delta \ln GDP_{it} := \ln GDP_{it} - \ln GDP_{i(t-1)}$ and $\boldsymbol{X}_{it} = (\Delta \ln L_{it}, \Delta \ln K_{it}, \Delta \ln H_{it})^\top$ 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 generality, we let $\Delta \ln GDP_{i1} = \Delta \ln L_{i1} = \Delta \ln K_{i1} = \Delta \ln H_{i1} = 0$. Each time series $\mathcal{T}_i$ is assumed to follow the model $Y_{it} = m_i(t/T) + \boldsymbol{\beta}_i^\top \boldsymbol{X}_{it} + \alpha_i + \varepsilon_{it}$, or equivalently,

$$\Delta \ln GDP_{it} = m_i\left(\frac{t}{T}\right) + \beta_{i,1} \Delta \ln L_{it} + \beta_{i,2} \Delta \ln K_{it} + \beta_{i,3} \Delta \ln H_{it} + \alpha_i + \varepsilon_{it} \qquad (7.1)$$

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 country-specific fixed effect.

In order to test the null hypothesis $H_0 : m_1 = \ldots = m_n$ with $n = 11$ in model (7.1), we implement our multiscale test as follows:

- We choose K to be the Epanechnikov kernel and consider the set of location-scale points $\mathcal{G}_T = U_T \times H_T$, where

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

We thus take into account all locations u on an equidistant grid U_T with step length $4/T$ and all scales $h = 4/T, 8/T, 12/T, \ldots$ with $\log T/T \leq h \leq 1/4$. The choice of the grid $\mathcal{G}_T$ is motivated by the quarterly frequency of the data: each interval $\mathcal{I}_{u,h} \in \mathcal{G}_T$ spans $8, 16, 24, \ldots$ quarters, i.e., $2, 4, 6, \ldots$ years. The lower bound $\log T/T$ on the scales h in H_T is motivated by Assumption (C11), which requires that $\log T/T \ll h_{\min}$ (given that all moments of ε_{it} exist).

- To obtain an estimator $\hat{\sigma}_i^2$ of the long-run error variance σ_i^2 for each i, we assume that the error process $\mathcal{E}_i$ follows an $AR(p_i)$ model and apply the difference-based procedure of Khismatullina and Vogt (2020) to the augmented time series $\{\widehat{Y}_{it} : 1 \leq t \leq T\}$ with $\widehat{Y}_{it} = Y_{it} - \widehat{\boldsymbol{\beta}}_i^\top \boldsymbol{X}_{it} - \widehat{\alpha}_i$. We set the tuning parameters q and r of the procedure to 20 and 10, respectively, and choose the AR order p_i by minimizing the Bayesian Information Criterion (BIC), which yields $p_i = 3$ for Australia, Canada and the UK

and $p_i = 1$ for all other countries.[6]

- The critical values $q_{n,T}(\alpha)$ are computed by Monte Carlo methods as described in Section 3.4, where we set $L = 5000$.

Besides these choices, we construct and implement the multiscale test exactly as described in Section 3.

The thus implemented multiscale test rejects the global null hypothesis H_0 at the usual significance levels $\alpha = 0.01, 0.05, 0.1$. This result is in line with the findings in Zhang et al. (2012) where the null hypothesis of a common trend is rejected at level $\alpha = 0.1$. The main advantage of our multiscale test over the method in Zhang et al. (2012) is that it is much more informative. In particular, it provides information about *which* of the $n = 11$ countries have different trends and *where* the trends differ. This information is presented graphically in Figures 3–13. Each figure corresponds to a specific pair of countries (i, j) and 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 compared. Panel (b) presents smoothed versions of the time series from (a), in particular, it shows local linear kernel estimates of the two trend functions m_i and m_j, where the bandwidth is set to 14 quarters (that is, to 0.1 in terms of rescaled time) and an Epanechnikov kernel is used. Panel (c) presents the results produced by our test for the 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 $\mathcal{S}_{min}^{[i,j]}(\alpha) \subseteq \mathcal{S}^{[i,j]}(\alpha)$ is highlighted in black. According to (4.3), 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 the functions m_i and m_j on each of these intervals.

Out of 55 pairwise comparisons, our test detects differences for 11 pairs of countries (i, j). These 11 cases are presented in Figures 3–13. In 9 cases (Figures 3–11), one of the involved countries is Norway. Inspecting the trend estimates in panels (b) of Figures 3–11, the Norwegian trend estimate can be seen to exhibit a strong downward movement at the end of the observation period, whereas the other trend estimates show a much less pronounced downward movement (or even a slight upward movement). According to our test, this is a significant difference between the Norwegian and the other trend functions rather than an artefact of the sampling noise: In all 9 cases, the test rejects the local null for at least one interval which covers the last 10 years of the analysed time period (from the first quarter in 2000 up to the third quarter in 2010). Apart from these differences at the end of the sampling period, our test also finds differences in the beginning, however, only for part of the pairwise comparisons.

---

[6]We also calculated 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.
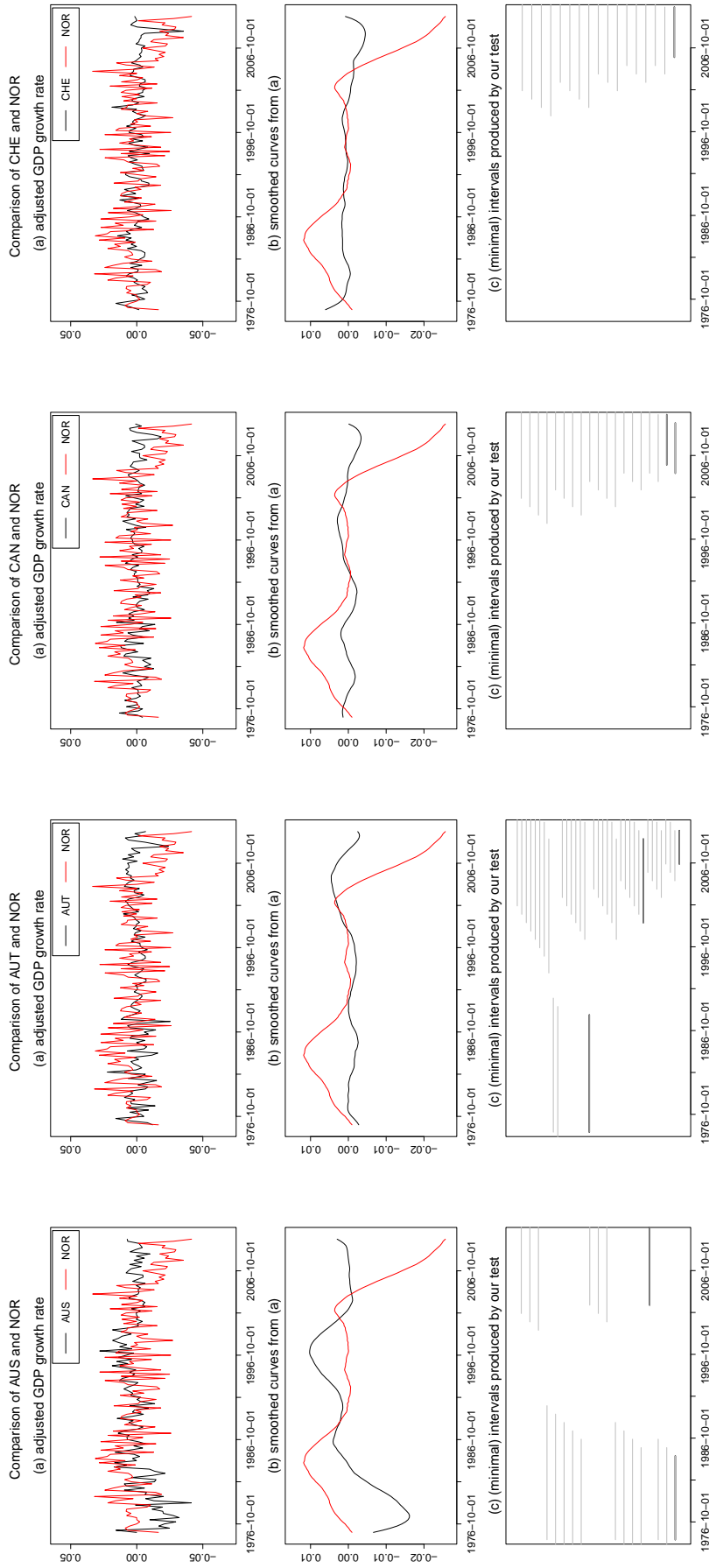
Figure 3: Test results for the comparison of Australia and Norway.

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

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 $\mathcal{S}_{min}^{[i,j]}(\alpha)$ in black.
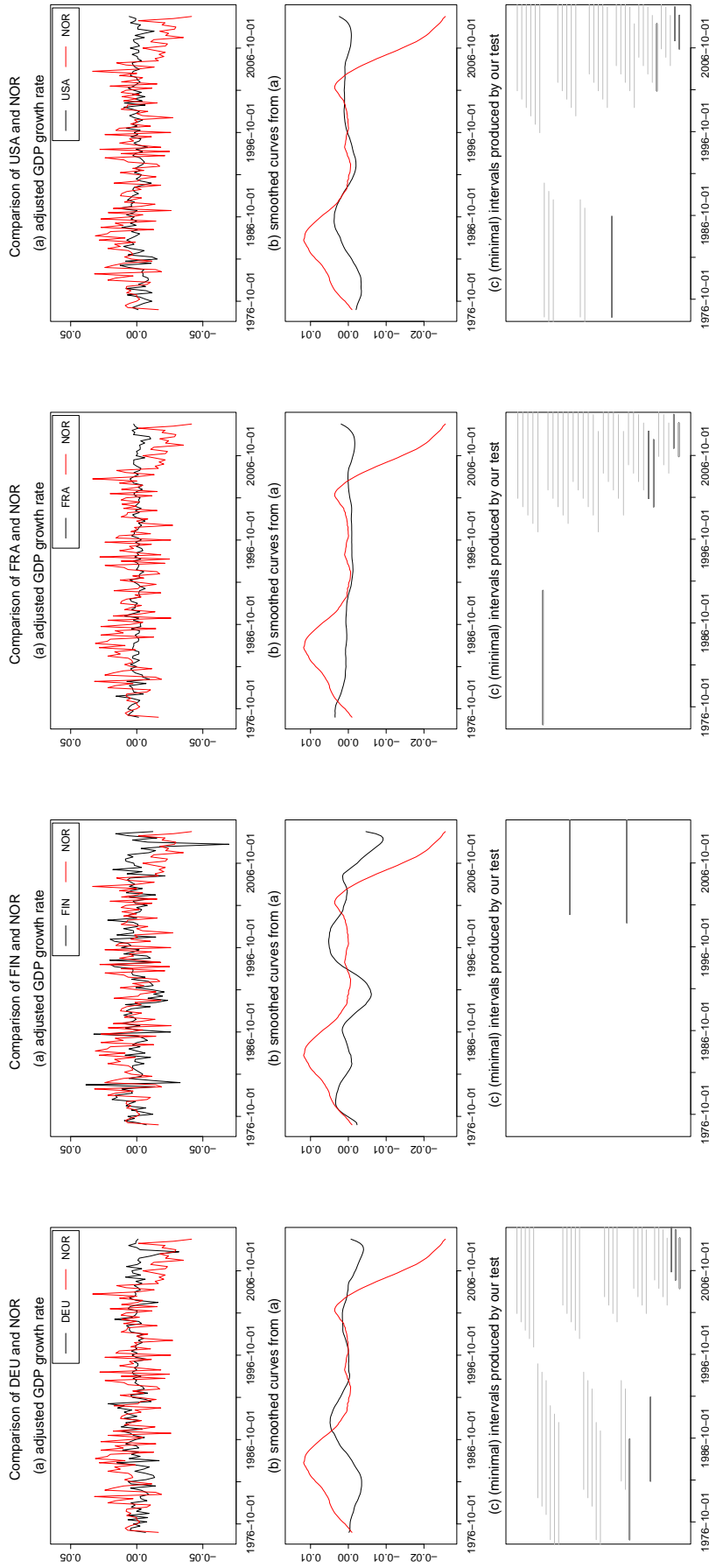
Figure 7: Test results for the comparison of Germany and Norway.

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

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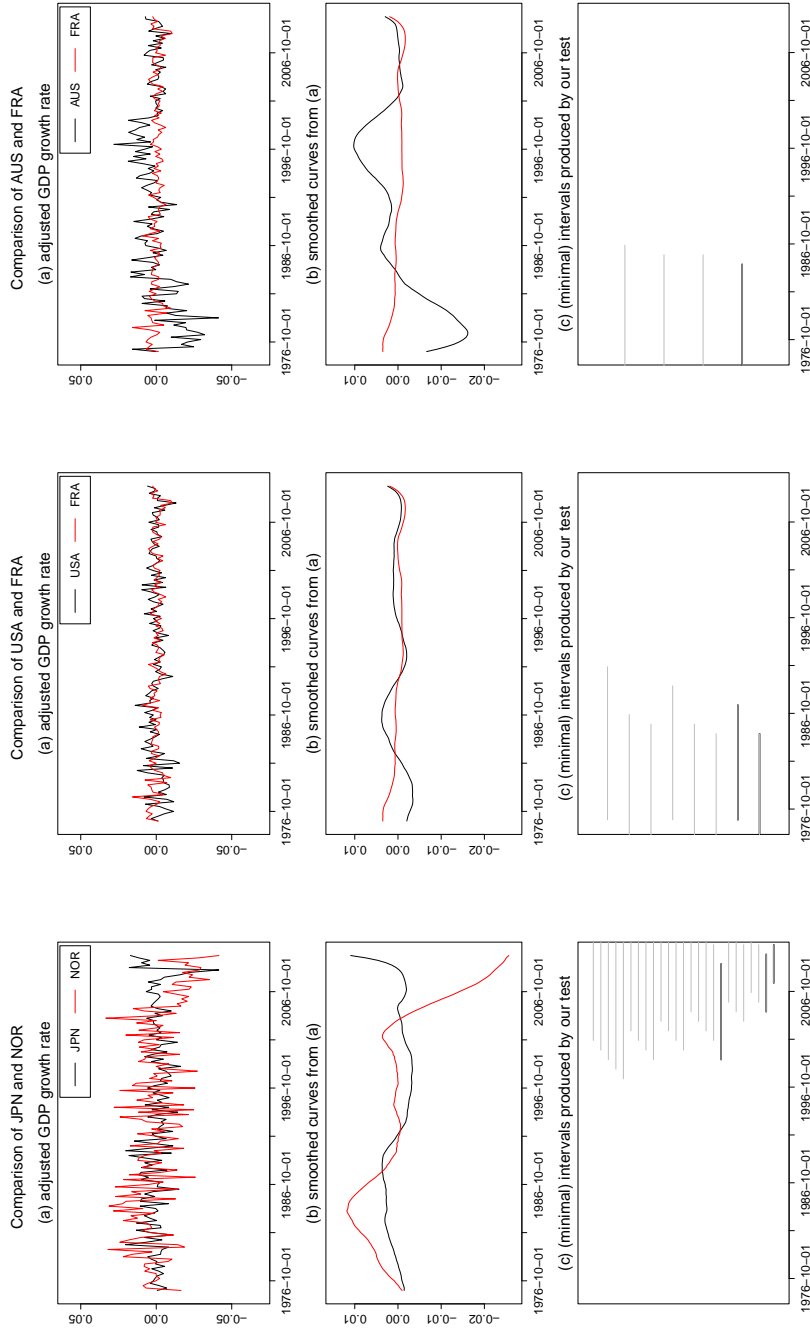
Figure 11: Test results for the comparison of Japan and Norway.

Figure 12: Test results for the comparison of the USA and France.
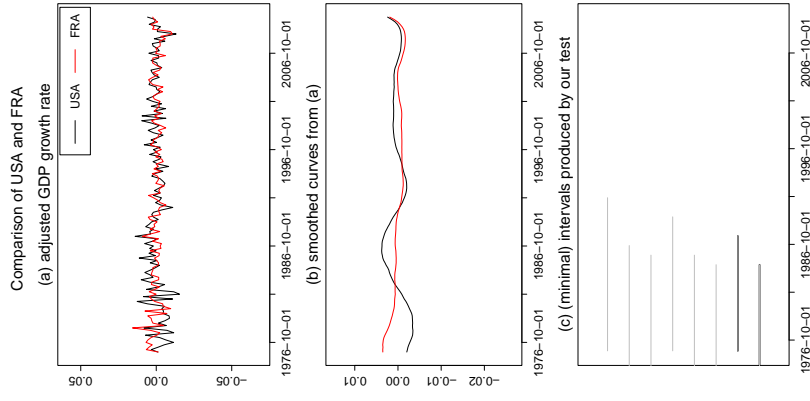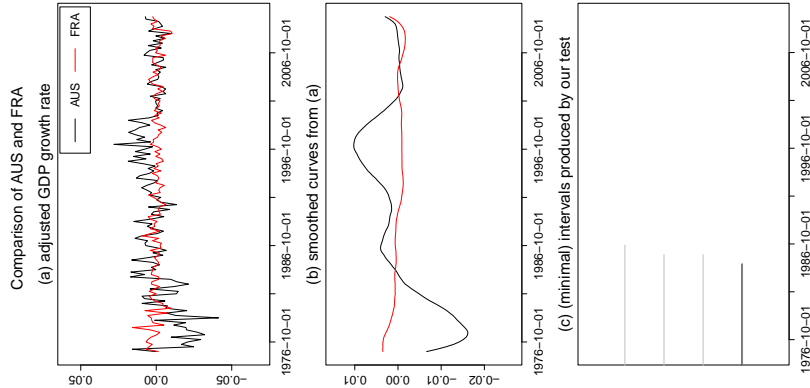
Figure 13: Test results for the comparison of Australia and France.

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.
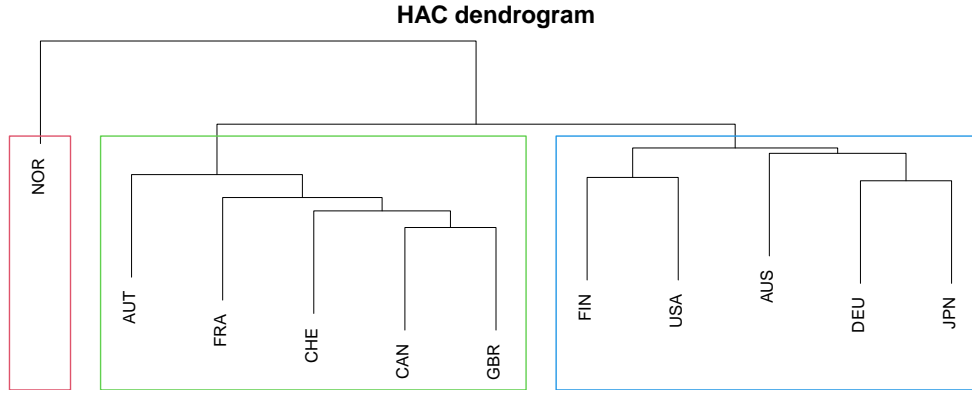
HAC dendrogram

Figure 14: Dendrogram of the HAC algorithm. Each coloured rectangle corresponds to one of the clusters.
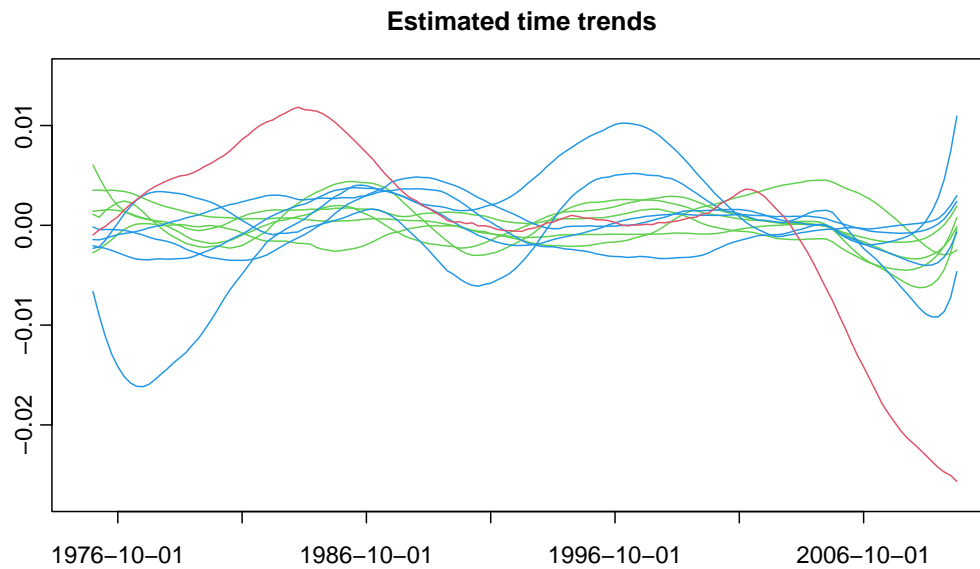
Estimated time trends

Figure 15: Local linear estimates of the $n = 11$ time trends (calculated from the augmented time series $\widehat{Y}_{it}$ with bandwidth $h = 0.1$ and Epanechnikov kernel). Each trend estimate is coloured according to the cluster that it is assigned to.

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 differences between the GDP trends only in the beginning of the considered time period. In the case of Australia and France, it is clearly visible in the raw data (panel (a) in Figure 13) that there is a difference between the trends, whereas this is not so obvious in the case of the USA and France. According to our test, there are indeed significant differences in both cases. In particular, we can claim with confidence at least 95%, that there are differences between the trends of the USA and France (of Australia and France) up to the fourth quarter in 1991 (the fourth quarter in 1986), but there is no evidence of any differences between the trends from 1992 (1987) onwards.

We next apply our clustering techniques to find groups of countries that have the same

time trend. We implement our HAC algorithm with $\alpha = 0.05$ and the same choices as detailed above. The dendrogram that depicts the clustering results is plotted in Figure 14. The number of clusters is estimated to be $\widehat{N} = 3$. The rectangles in Figure 14 indicate the $\widehat{N} = 3$ clusters. In particular, each rectangle is drawn around the branches of the dendrogram that correspond to one of the three clusters. Figure 15 depicts local linear kernel estimates of the $n = 11$ GDP time trends (calculated from the augmented time series $\widehat{Y}_{it}$ with bandwidth 0.1 and Epanechnikov kernel). Their colour indicates which cluster they belong to.

The results in Figures 14 and 15 show that there is one cluster which consists only of Norway (plotted in red). As we have already discussed above and as becomes apparent from Figure 15, the Norwegian trend exhibits a strong downward movement at the end of the sampling period, whereas the other trends show a much more moderate downward movement (if at all). This is presumably the reason why the clustering procedure puts Norway in a separate cluster. The algorithm further finds two other clusters, one consisting of the 5 countries Australia, Finland, Germany, Japan and the USA (plotted in blue in Figures 14 and 15) and the other one consisting of the 5 countries Austria, Canada, France, Switzerland and the UK (plotted in green in Figures 14 and 15). Visual inspection of the trend estimates in Figure 15 suggests that the GDP time trends in the blue cluster exhibit more pronounced decreases and increases than the GDP time trends in the green cluster. Hence, overall, the clustering procedure appears to produce a reasonable grouping of the GDP trends.

7.2 Analysis of house prices

We next analyse a historical dataset on nominal annual house prices from Knoll et al. (2017) that contains data for 14 advanced economies covering 143 years from 1870 to 2012. In our analysis, we consider 8 countries (Australia, Belgium, Denmark, France, Netherlands, Norway, Sweden and USA) over the time period 1890–2012. The data for all these countries except one (Belgium) contain no missing values, and for Belgium there are only five missing observations[7] which we impute by linear interpolation. The time series of the other 6 countries contain more than 10 missing values each, which is why we exclude them from our analysis.

We deflate the nominal house prices with the corresponding consumer price index (CPI) to obtain real house prices (HP). Variables that can potentially influence the average house prices are numerous, and there seems to be no general consensus about what the main determinants 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

---

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

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).

In summary, we 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 $i \in \{1, \ldots, 8\}$, where $Y_{it} = \ln HP_{it}$ and $\boldsymbol{X}_{it} = (\ln GDP_{it}, \ln POP_{it}, I_{it}, INFL_{it})^\top$. For each i, the time series $\mathcal{T}_i$ is assumed to follow the model $Y_{it} = m_i(t/T) + \boldsymbol{\beta}_i^\top \boldsymbol{X}_{it} + \alpha_i + \varepsilon_{it}$, or equivalently,

$$\ln HP_{it} = m_i\left(\frac{t}{T}\right) + \beta_{i,1} \ln GDP_{it} + \beta_{i,2} \ln POP_{it} + \beta_{i,3} I_{it} + \beta_{i,4} INFL_{it} + \alpha_i + \varepsilon_{it} \quad (7.2)$$

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.

The inclusion of a nonparametric trend function m_i in model (7.2) is supported by the literature. Ugarte et al. (2009), for example, model the trend in average Spanish house prices by means of splines. Winter et al. (2022) include a long-term stochastic trend component when describing the dynamic behaviour of real house prices in 8 advanced economies. Including a nonparametric trend function when modelling the evolution of house prices is also the main conclusion in Zhang et al. (2016), where it is shown that the time series of logarithmic US house prices is trend-stationary, i.e., can be transformed into a stationary series by subtracting a deterministic trend.

We implement the multiscale test from Section 3 in the same way as in the previous application example with one minor modification: we let $\mathcal{G}_T = U_T \times H_T$ with

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

We thus take into account all locations u on an equidistant grid U_T with step length $1/T$ and all scales $h = 2/T, 7/T, 12/T, \ldots$ with $\log T/T \leq h \leq 1/4$. This implies that each interval $\mathcal{I}_{u,h} = [u-h, u+h]$ with $(u, h) \in \mathcal{G}_T$ spans $5, 15, 25, \ldots$ years. The lower bound $\log T/T$ is motivated by Assumption (C11). As in Section 7.1, we assume that for each i, the error process $\mathcal{E}_i = \{\varepsilon_{it} : 1 \leq t \leq T\}$ follows an AR(p_i) model and we estimate the long-run variances σ_i^2 by the difference-based estimator from Khismatullina and Vogt

---

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

(2020) with tuning parameters q and r equal to 15 and 10, respectively. We choose p_i by minimizing the BIC. For 7 out of 8 countries the order p_i determined by BIC[9] is equal to 1. For the sake of simplicity, we thus assume that $p_i = 1$ for all i.

We are now ready to apply our test to the data. The overall null hypothesis H_0 is rejected at levels $\alpha = 0.05$ and $\alpha = 0.1$. The detailed test results for the significance level $\alpha = 0.05$ are presented in Figures 16–19. As in Section 7.1, each figure corresponds to the comparison of a pair of countries (i, j) for which our test detects differences between the trends. 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 under consideration. Panel (b) presents smoothed versions of the time series from (a) with the bandwidth window covering 15 years. Panel (c) presents the test results for the level $\alpha = 0.05$. As before, the set of intervals $\mathcal{S}^{[i,j]}(\alpha)$ for which our test rejects and the set of minimal intervals $\mathcal{S}_{min}^{[i,j]}(\alpha) \subseteq \mathcal{S}^{[i,j]}(\alpha)$ are depicted in grey and black, respectively. According to (4.3), we can make the following simultaneous confidence statement about the intervals plotted in panels (c) of Figures 16–19: we can claim, with confidence of about 95%, that there is a difference between the trends m_i and m_j on each of these intervals.

Overall, our findings are in line with the observations in Knoll et al. (2017), where the authors find considerable cross-country heterogeneity in the house price trends. The authors note that before World War II, the countries exhibit similar trends in real house prices, while the trends start to diverge sometime after World War II. This fits with our findings in Figures 16 and 17 which show that our test detects differences between the trends of Australia and the Netherlands (of Belgium and the Netherlands) starting only from 1968 (1966) onwards. Contrary to Knoll et al. (2017), however, our test also finds significant differences in the first half of the observed time period, specifically, between the time trends of Belgium and Denmark and of Belgium and the USA (Figures 18 and 19, respectively). This discrepancy in the results is most certainly due to the fact that the method used in Knoll et al. (2017) does not account for effects of other factors such as GDP or population growth, while our test allows us to include various determinants of the average house prices in model (7.2).

We next apply the clustering procedure from Section 5 to the data. As in the previous application example, we set $\alpha = 0.05$. The results are displayed in Figures 20 and 21. Specifically, Figure 20 shows the dendrogram with the results of the HAC algorithm and Figure 21 presents local linear kernel estimates of the trend curves (calculated with a bandwidth window of 15 years and an Epanechnikov kernel). The number of clusters is estimated to be $\widehat{N} = 3$. As before, the coloured rectangles in Figure 20 are drawn around the countries that belong to the same cluster and the same colours are used to display the trend estimates in Figure 21.

---

[9]Applying other information criteria such as FPE, AIC and HQ yields exactly the same results in these cases.
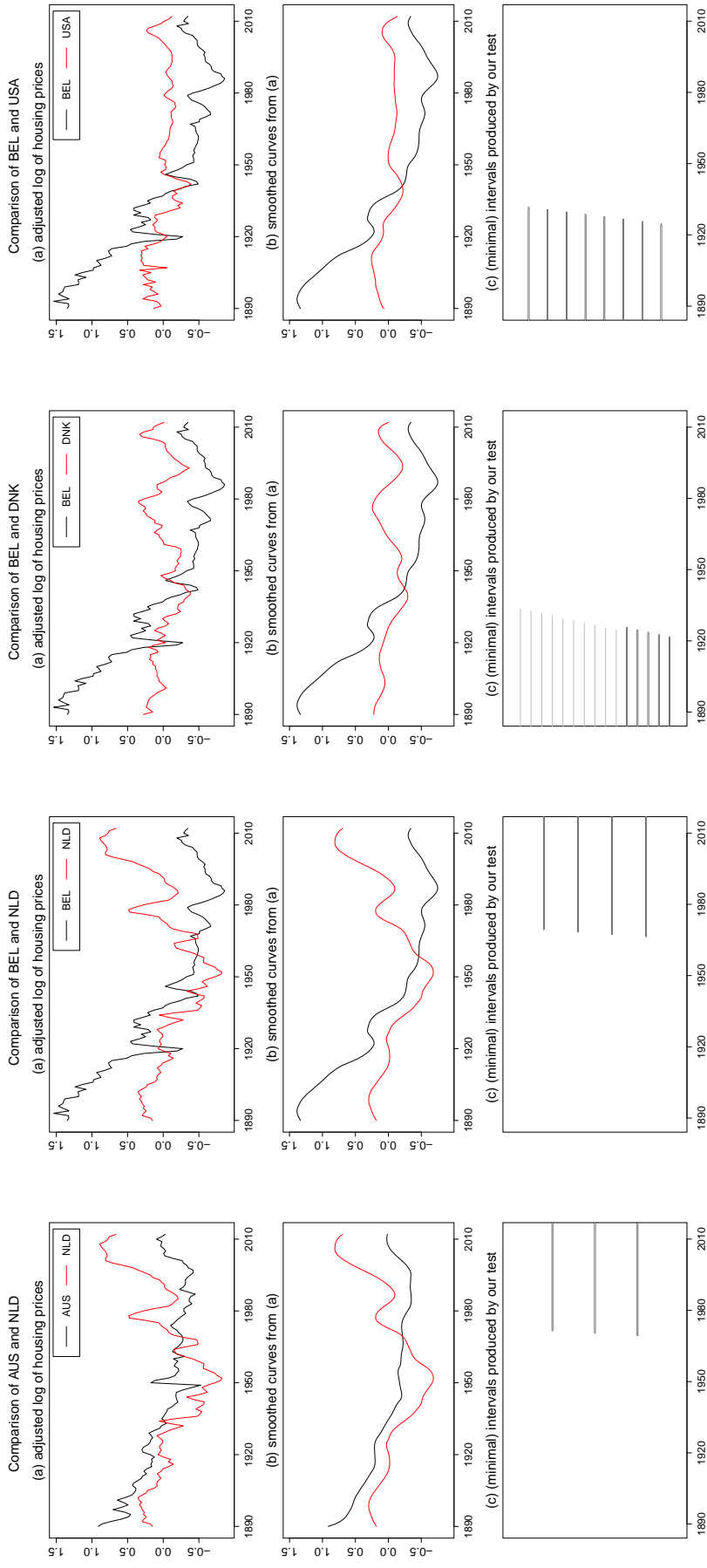
Figure 16: Test results for the comparison of the house prices in Australia and the Netherlands.

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

Figure 18: Test results for the comparison of the house prices in Belgium and Denmark.

Figure 19: Test results for the comparison of the house 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.
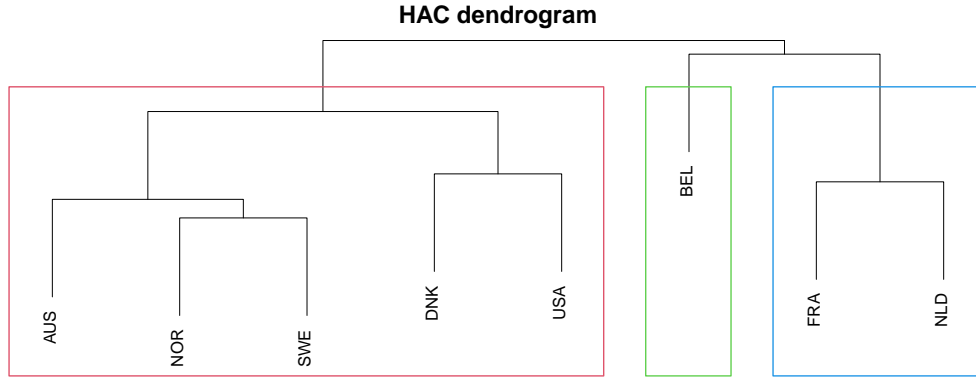
HAC dendrogram

Figure 20: Dendrogram of the HAC algorithm. Each coloured rectangle corresponds to one of the clusters.

Estimated time trends
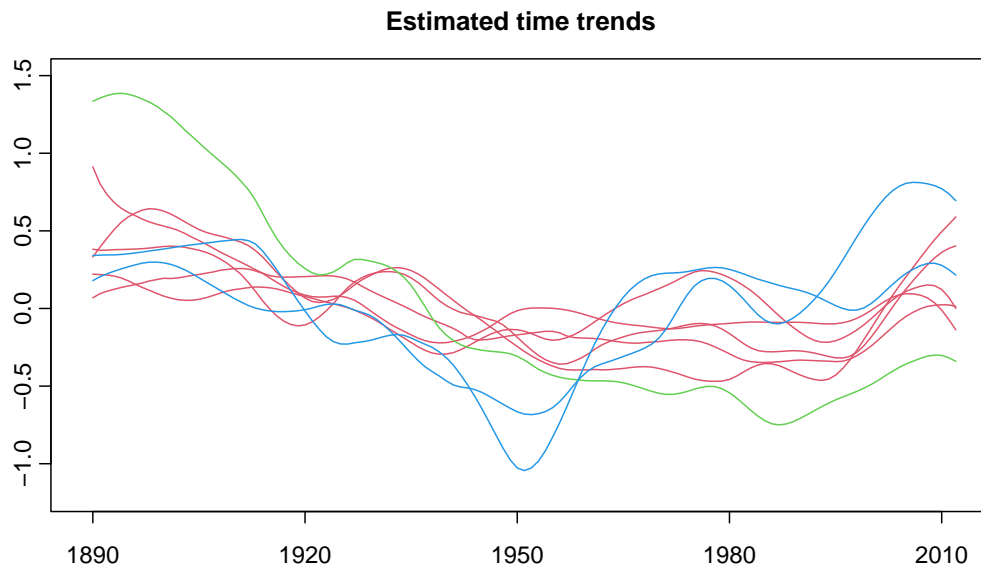
Figure 21: Local linear estimates of the $n = 8$ time trends (calculated from the augmented time series $\widehat{Y}_{it}$ with a bandwidth window covering 15 years and an Epanechikov kernel). Each trend estimate is coloured according to the cluster that it is assigned to.

Inspecting the results, we can see that there is one cluster consisting only of Belgium (plotted in green). Figure 21 suggests that the Belgium time trend indeed evolves somewhat differently from the other trends in the first 30 years of the observed time period. The algorithm further detects a cluster that consists of two countries: France and the Netherlands (plotted in blue). The time trends of these two countries display some kind of dip around 1950, which is not present in the time trends of the other countries. Overall, our algorithm thus appears to produce a reasonable clustering of the house price trends.

34

Acknowledgements

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.

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

CAI, Z. (2007). Trending time-varying coefficients time series models with serially correlated errors. *Journal of Econometrics*, **136** 163–188.

CARLSTEIN, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *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. (2019). Testing for trends in high-dimensional time series. *Journal of the American Statistical Association*, **114** 869–881.

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. *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.

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

DÜMBGEN, L. and SPOKOINY, V. G. (2001). Multiscale testing of qualitative hypotheses. *Annals of Statistics*, **29** 124–152.

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*, **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. and HUANG, B.-N. (2002). The relationship between exports and economic growth in East Asian countries: a multivariate threshold autoregressive approach. *Journal of Economic Development*, **27** 45–68.

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.

NYBLOM, J. and HARVEY, A. (2000). Tests of common stochastic trends. *Econometric Theory*, **16** 176–199.

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.

STOCK, J. H. and WATSON, M. W. (1988). Testing for common trends. *Journal of the American Statistical Association*, **83** 1097–1107.

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** 3616–3629.

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

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*, **69** 391–410.

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 prove the theoretical results from Sections 4 and 5. We use the following notation: The symbol C denotes a universal real constant which may take a different value on each occurrence. For $a, b \in \mathbb{R}$, we write $a \vee b = \max\{a, b\}$. 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 get

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

$$= \frac{\{C_1 C_Z^q\} T^{1-q/2}}{w^q} + C_2 \exp\left(-\frac{C_3 w^2}{C_Z^2}\right) \leq \eta$$

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

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

$$\boldsymbol{a}_i(\mathcal{H}_{it}) := \Delta\boldsymbol{h}_i(\mathcal{G}_{it})\Delta g_i(\mathcal{F}_{it}) \quad = \Delta\boldsymbol{X}_{it}\Delta\varepsilon_{it}$$

$$\boldsymbol{b}_i(\mathcal{G}_{it}) := \Delta\boldsymbol{h}_i(\mathcal{G}_{it})\Delta\boldsymbol{h}_i(\mathcal{G}_{it})^{\top} = \Delta\boldsymbol{X}_{it}\Delta\boldsymbol{X}_{it}^{\top},$$

where $\boldsymbol{a}_i = (a_{ij})_{j=1}^{d}$, $\boldsymbol{b}_i = (b_{ikl})_{k,l=1}^{d}$ and $\mathcal{H}_{it} = (\mathcal{H}_{it,1}, \ldots, \mathcal{H}_{it,d})^{\top}$ with $\mathcal{H}_{it,j} = (\ldots, \nu_{it-1,j}, \nu_{it,j})$ and $\nu_{it,j} = (\eta_{it}, \xi_{it,j})$. 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*

$$\sum_{s=t}^{\infty} \delta_p(a_{ij}, s) = O(t^{-\alpha}) \qquad \text{for } p = \min\{q, q'\}/2 \text{ and some } \alpha > 1/2 - 1/p$$

$$\sum_{s=t}^{\infty} \delta_p(b_{ikl}, s) = O(t^{-\alpha}) \qquad \text{for } p = q'/2 \text{ and some } \alpha > 1/2 - 1/p.$$

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

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

where $\mathcal{H}'_{it,j} = (\ldots, \nu_{i(-1),j}, \nu'_{i0,j}, \nu_{i1,j}, \ldots, \nu_{it-1,j}, \nu_{it,j})$, $\mathcal{G}'_{it,j} = (\ldots, \xi_{i(-1),j}, \xi'_{i0,j}, \xi_{i1,j}, \ldots,$

$\xi_{it-1,j}, \xi_{it,j})$ and $\mathcal{F}'_{it} = (\ldots, \eta_{i(-1)}, \eta'_{i0}, \eta_{i1}, \ldots, \eta_{it-1}, \eta_{it})$ 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 $\sum_{s=t}^{\infty} \delta_p(a_{ij}, s) = O(t^{-\alpha})$. $\qquad\square$

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,

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

In what follows, we show that

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

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

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

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 $p > 2$, $\alpha > 1/2 - 1/p$ 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

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

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,

$$\left| \frac{1}{\sqrt{T}} \sum_{t=2}^{T} \Delta X_{it,j} \Delta m_{it} \right| \leq \frac{1}{\sqrt{T}} \sum_{t=2}^{T} |\Delta X_{it,j}| \cdot |\Delta m_{it}|$$

$$\leq \frac{C}{\sqrt{T}} \cdot \frac{1}{T} \sum_{t=2}^{T} |\Delta X_{it,j}| = O_p\left(\frac{1}{\sqrt{T}}\right),$$

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

Lemma A.6. *Let $s_T \asymp T^{1/3}$. Under Assumptions (C1)–(C7),*

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

for each i, where $\widehat{\sigma}_i^2$ is the subseries variance estimator of σ_i^2 introduced in (3.4).

Proof of Lemma A.6. Let $Y_{it}^{\circ} := m_i(t/T) + \varepsilon_{it}$. Using simple arithmetic calculations, we can rewrite $\widehat{\sigma}_i^2$ as $\widehat{\sigma}_i^2 = \widehat{\sigma}_{i,A}^2 + \widehat{\sigma}_{i,B}^2 - \widehat{\sigma}_{i,C}^2$, where

$$\widehat{\sigma}_{i,A}^2 = \frac{1}{2(M-1)s_T} \sum_{m=1}^{M} \left[\sum_{t=1}^{s_T} \left(Y_{i(t+ms_T)}^{\circ} - Y_{i(t+(m-1)s_T)}^{\circ} \right) \right]^2$$

$$\widehat{\sigma}_{i,B}^2 = \frac{1}{2(M-1)s_T} \sum_{m=1}^{M} \left[\sum_{t=1}^{s_T} (\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i)^{\top} \left(\boldsymbol{X}_{i(t+ms_T)} - \boldsymbol{X}_{i(t+(m-1)s_T)} \right) \right]^2$$

$$\widehat{\sigma}_{i,C}^2 = \frac{1}{(M-1)s_T} \sum_{m=1}^{M} \left[\sum_{t=1}^{s_T} \left(Y_{i(t+ms_T)}^{\circ} - Y_{i(t+(m-1)s_T)}^{\circ} \right) \right.$$

$$\left. \times \sum_{t=1}^{s_T} (\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i)^{\top} \left(\boldsymbol{X}_{i(t+ms_T)} - \boldsymbol{X}_{i(t+(m-1)s_T)} \right) \right].$$

By Carlstein (1986) and Wu and Zhao (2007), we have $\widehat{\sigma}_{i,A}^2 = \sigma_i^2 + O_p(T^{-1/3})$. Moreover, under our assumptions, it is straightforward to see that $\widehat{\sigma}_{i,B}^2 = O_p(T^{-1/3})$ and $\widehat{\sigma}_{i,C}^2 = O_p(T^{-1/3})$. $\qquad\square$

Proof of Theorem 4.1

We first summarize the main proof strategy, which splits up into five steps, and then fill in the details. We in particular 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 $\rho_T = o(1/\log T)$. 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 (3.12). 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 (3.12).

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*

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

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

$$\sup_{x \in \mathbb{R}} \left|\mathbb{P}(\widehat{\widetilde{\Phi}}_{n,T} \leq x) - \mathbb{P}(\Phi_{n,T} \leq x)\right| = o(1). \tag{A.5}$$

Proof of (A.5). It holds that

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

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

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

by Step 2, we arrive at (A.5). □

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

$$\sup_{x\in\mathbb{R}} \left|\mathbb{P}(\widehat{\Phi}_{n,T} \le x) - \mathbb{P}(\Phi_{n,T} \le x)\right| = o(1). \tag{A.6}$$

Proof of (A.6). To start with, we verify that for any $x \in \mathbb{R}$ and any $\delta > 0$,

$$\mathbb{P}\left(\widehat{\widehat{\Phi}}_{n,T} \le x - \delta\right) - \mathbb{P}\left(\left|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\right| > \delta\right)$$
$$\le \mathbb{P}\left(\widehat{\Phi}_{n,T} \le x\right) \le \mathbb{P}\left(\widehat{\widehat{\Phi}}_{n,T} \le x + \delta\right) + \mathbb{P}\left(\left|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\right| > \delta\right). \tag{A.7}$$

It holds that

$$\mathbb{P}(\widehat{\Phi}_{n,T} \le x) = \mathbb{P}\left(\widehat{\Phi}_{n,T} \le x, \left|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\right| \le \delta\right) + \mathbb{P}\left(\widehat{\Phi}_{n,T} \le x, \left|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\right| > \delta\right)$$
$$\le \mathbb{P}\left(\widehat{\Phi}_{n,T} \le x, \widehat{\Phi}_{n,T} - \delta \le \widehat{\widehat{\Phi}}_{n,T} \le \widehat{\Phi}_{n,T} + \delta\right) + \mathbb{P}\left(\left|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\right| > \delta\right)$$
$$\le \mathbb{P}\left(\widehat{\widehat{\Phi}}_{n,T} \le x + \delta\right) + \mathbb{P}\left(\left|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\right| > \delta\right)$$

and analogously

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

Combining these two inequalities, we arrive at (A.7).
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.7), we get that

$$\left|\mathbb{P}\left(\widehat{\Phi}_{n,T} \le x\right) - \mathbb{P}\left(\Phi_{n,T} \le x\right)\right| = \mathbb{P}\left(\widehat{\Phi}_{n,T} \le x\right) - \mathbb{P}\left(\Phi_{n,T} \le x\right)$$
$$\le \mathbb{P}\left(\widehat{\widehat{\Phi}}_{n,T} \le x + \delta_T\right) + \mathbb{P}\left(\left|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\right| > \delta_T\right)$$
$$- \mathbb{P}\left(\Phi_{n,T} \le x\right)$$

45

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

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

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

Consequently,

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

Since the terms on the right-hand side are all $o(1)$ by Steps 2–4, we arrive at (A.6). $\quad\square$

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

$$\max_{1 \le t \le T} \Big| \sum_{s=1}^{t} \widetilde{\varepsilon}_{is} - \sigma_i \mathbb{B}_i(t) \Big| = o\big(T^{1/q}\big) \quad \text{a.s.,} \tag{A.8}$$

where $\sigma_i^2 = \sum_{k \in \mathbb{Z}} \text{Cov}(\varepsilon_{i0}, \varepsilon_{ik})$ denotes the long-run error variance. We apply this result 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

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

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 $\widehat{\widetilde{\sigma}}_i^2$ is constructed from $\mathcal{E}_i$. Since $[\widetilde{\varepsilon}_{i1}, \ldots, \widetilde{\varepsilon}_{iT}] \overset{\mathcal{D}}{=} [\varepsilon_{i1}, \ldots, \varepsilon_{iT}]$ and $\widehat{\widetilde{\sigma}}_i^2 = \sigma_i^2 + o_p(\rho_T)$, 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

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

and the auxiliary statistic

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

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

$$\big|\widetilde{\Phi}_{n,T} - \Phi_{n,T}^{\diamond}\big| = o_p\Big(\frac{T^{1/q}}{\sqrt{Th_{\min}}}\Big) \tag{A.9}$$

$$\big|\Phi_{n,T}^{\diamond} - \Phi_{n,T}\big| = o_p(\rho_T \sqrt{\log T}), \tag{A.10}$$

which completes the proof.

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

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

$$= O_p(1) \cdot \max_{1 \le i < j \le n} \max_{(u,h) \in \mathcal{G}_T} \big|\widetilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h)\big|, \tag{A.11}$$

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

$$\big|\widetilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h)\big|$$

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

$$= \Big| \sum_{t=1}^{T-1} A_{ij,t}\big(w_{t,T}(u,h) - w_{t+1,T}(u,h)\big) + A_{ij,T} w_{T,T}(u,h) \Big|,$$

47

where

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

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

$$\left| \widetilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h) \right| \leq W_T(u,h) \max_{1 \leq t \leq T} |A_{ij,t}| \tag{A.12}$$

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

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

Applying the strong approximation result (A.8), we can infer that

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

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.12), we obtain that

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

which in view of (A.11) yields that $|\widetilde{\Phi}_{n,T} - \Phi^\diamond_{n,T}| = o_p(T^{1/q}/\sqrt{Th_{\min}})$. This completes the proof of (A.9).

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

$$
\begin{aligned}
\left|\Phi^\diamond_{n,T} - \Phi_{n,T}\right| &\leq \max_{1\leq i<j\leq n} \max_{(u,h)\in\mathcal{G}_T} \left| \frac{\phi_{ij,T}(u,h)}{\{\widetilde{\sigma}_i^2 + \widetilde{\sigma}_j^2\}^{1/2}} - \frac{\phi_{ij,T}(u,h)}{\{\sigma_i^2 + \sigma_j^2\}^{1/2}} \right| \\
&\leq \max_{1\leq i<j\leq n} \left\{ \left| \left(\widetilde{\sigma}_i^2 + \widetilde{\sigma}_j^2\right)^{-1/2} - \left(\sigma_i^2 + \sigma_j^2\right)^{-1/2} \right| \right\} \max_{1\leq i<j\leq n} \max_{(u,h)\in\mathcal{G}_T} |\phi_{ij,T}(u,h)| \\
&= o_p(\rho_T) \max_{1\leq i<j\leq n} \max_{(u,h)\in\mathcal{G}_T} |\phi_{ij,T}(u,h)|, \tag{A.13}
\end{aligned}
$$

where the last line is due to the fact that $\widetilde{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$. 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 $c_T(u,h) = \{\sum_{t=1}^{T} w_{t,T}(u,h)\}^2/T \leq C < \infty$ for all $(u,h) \in \mathcal{G}_T$ and $1 \leq i < j \leq n$. 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 $1 \leq i < j \leq n$. Hence, standard results on the maximum of Gaussian random variables yield that

$$
\max_{1\leq i<j\leq n} \max_{(u,h)\in\mathcal{G}_T} \left|\phi_{ij,T}(u,h)\right| = O_p(\sqrt{\log T}), \tag{A.14}
$$

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.13) yields $|\Phi^\diamond_{n,T} - \Phi_{n,T}| = o_p(\rho_T\sqrt{\log T})$, which completes the proof of (A.10). $\qquad\square$

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

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

where the constant C only depends on b.

To apply this result, we introduce the following notation: We write $x = (u,h)$ and $\mathcal{G}_T = \{x_1,\ldots,x_p\}$, where $p := |\mathcal{G}_T| \leq O(T^\theta)$ for some large but fixed $\theta > 0$ by our

assumptions. For $k = 1, \ldots, p$ and $1 \leq i < j \leq n$, 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 $\mathbb{E}[Z_{ij,l}^2] \geq b > 0$ for all i, j and l. We next construct the random vector $\boldsymbol{Z} = (Z_{ij,l} : 1 \leq i < j \leq n, 1 \leq l \leq 2p)$ by stacking the variables $Z_{ij,l}$ in a certain order (which can be chosen freely) and construct the vector $\boldsymbol{\lambda} = (\lambda_{ij,l} : 1 \leq i < j \leq n, 1 \leq l \leq 2p)$ 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 $\mathbb{P}(\Phi_{n,T} \leq q)$ as follows for each $q \in \mathbb{R}$:

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

Consequently,

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

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

Proof of Proposition A.9. Straightforward calculations yield that

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

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

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

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

$$= o_p(\rho_T) \max_{1 \le i < j \le n} \max_{(u,h) \in \mathcal{G}_T} \left| \widehat{\widehat{\phi}}_{ij,T}(u,h) \right|$$

and

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

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

$$= O_p(1) \max_{1 \le i < j \le n} \max_{(u,h) \in \mathcal{G}_T} \left| \widehat{\widehat{\phi}}_{ij,T}(u,h) - \widehat{\phi}_{ij,T}(u,h) \right|,$$

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

$$\left| \widehat{\widehat{\phi}}_{ij,T}(u,h) - \widehat{\phi}_{ij,T}(u,h) \right|$$

$$= \left| \sum_{t=1}^{T} w_{t,T}(u,h) \left\{ (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top (\boldsymbol{X}_{it} - \bar{\boldsymbol{X}}_i) - (\boldsymbol{\beta}_j - \widehat{\boldsymbol{\beta}}_j)^\top (\boldsymbol{X}_{jt} - \bar{\boldsymbol{X}}_j) \right\} \right|$$

$$\le \left| (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top \sum_{t=1}^{T} w_{t,T}(u,h) \boldsymbol{X}_{it} \right| + \left| (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top \bar{\boldsymbol{X}}_i \right| \left| \sum_{t=1}^{T} w_{t,T}(u,h) \right|$$

$$+ \left| (\boldsymbol{\beta}_j - \widehat{\boldsymbol{\beta}}_j)^\top \sum_{t=1}^{T} w_{t,T}(u,h) \boldsymbol{X}_{jt} \right| + \left| (\boldsymbol{\beta}_j - \widehat{\boldsymbol{\beta}}_j)^\top \bar{\boldsymbol{X}}_j \right| \left| \sum_{t=1}^{T} w_{t,T}(u,h) \right|.$$

Hence,

$$\left| \widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T} \right| \le o_p(\rho_T) A_{n,T} + O_p(1) \left\{ 2 B_{n,T} + 2 C_{n,T} \right\}, \tag{A.15}$$

where

$$A_{n,T} = \max_{1 \le i < j \le n} \max_{(u,h) \in \mathcal{G}_T} \left| \widehat{\widehat{\phi}}_{ij,T}(u,h) \right|$$

$$B_{n,T} = \max_{1 \le i \le n} \max_{(u,h) \in \mathcal{G}_T} \left| (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top \sum_{t=1}^{T} w_{t,T}(u,h) \boldsymbol{X}_{it} \right|$$

$$C_{n,T} = \max_{1 \le i \le n} \left| (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top \bar{\boldsymbol{X}}_i \right| \max_{(u,h) \in \mathcal{G}_T} \left| \sum_{t=1}^{T} w_{t,T}(u,h) \right|.$$

We examine these three terms separately.

We first prove that

$$A_{n,T} = \max_{1 \leq i < j \leq n} \max_{(u,h) \in \mathcal{G}_T} \left| \widehat{\widetilde{\phi}}_{ij,T}(u,h) \right| = O_p\big(\sqrt{\log T}\big). \tag{A.16}$$

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

$$\max_{1 \leq i < j \leq n} \max_{(u,h) \in \mathcal{G}_T} \left| \widetilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h) \right| = o_p\Big(\frac{T^{1/q}}{\sqrt{T h_{\min}}} \Big). \tag{A.17}$$

Instead of (A.16), it thus suffices to show that

$$\max_{1 \leq i < j \leq n} \max_{(u,h) \in \mathcal{G}_T} \left| \widetilde{\phi}_{ij,T}(u,h) \right| = O_p\big(\sqrt{\log T}\big). \tag{A.18}$$

Since for any constant $c > 0$,

$$\mathbb{P}\left(\max_{i,j,(u,h)} |\phi_{ij,T}(u,h)| \leq \frac{c\sqrt{\log T}}{2} \right)$$
$$\leq \mathbb{P}\left(\max_{i,j,(u,h)} \left| \widetilde{\phi}_{ij,T}(u,h) \right| \leq c\sqrt{\log T} \right)$$
$$+ \mathbb{P}\left(\left| \max_{i,j,(u,h)} \left| \widetilde{\phi}_{ij,T}(u,h) \right| - \max_{i,j,(u,h)} |\phi_{ij,T}(u,h)| \right| > \frac{c\sqrt{\log T}}{2} \right)$$
$$\leq \mathbb{P}\left(\max_{i,j,(u,h)} \left| \widetilde{\phi}_{ij,T}(u,h) \right| \leq c\sqrt{\log T} \right)$$
$$+ \mathbb{P}\left(\max_{i,j,(u,h)} \left| \widetilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h) \right| > \frac{c\sqrt{\log T}}{2} \right)$$

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.17), we get that

$$\mathbb{P}\left(\max_{i,j,(u,h)} \left| \widetilde{\phi}_{ij,T}(u,h) \right| \leq c\sqrt{\log T} \right)$$
$$\geq \mathbb{P}\left(\max_{i,j,(u,h)} |\phi_{ij,T}(u,h)| \leq \frac{c\sqrt{\log T}}{2} \right) - o(1). \tag{A.19}$$

Moreover, since $\max_{i,j,(u,h)} |\phi_{ij,T}(u,h)| = O_p(\sqrt{\log T})$ as already proven in (A.14), we can make the probability $\mathbb{P}(\max_{i,j,(u,h)} |\phi_{ij,T}(u,h)| \leq c\sqrt{\log T}/2)$ on the right-hand side of (A.19) 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 $\mathbb{P}(\max_{i,j,(u,h)} |\widetilde{\phi}_{ij,T}(u,h)| \leq c\sqrt{\log T}) \geq 1 - \delta$ for sufficiently large T. This proves (A.18), which in turn yields (A.16).

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 $T(u-h) \leq t \leq T(u+h)$. We can use this fact to write

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

Note that

$$\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} w_{t,T}^2(u,h) = \sum_{t=1}^{T} w_{t,T}^2(u,h) = \sum_{t=1}^{T} \frac{\Lambda_{t,T}^2(u,h)}{\sum_{s=1}^{T} \Lambda_{s,T}^2(u,h)} = 1. \qquad \text{(A.20)}$$

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 $2Th \leq D_{T,u,h} \leq 2Th+2$) and using (A.20), we can normalize the kernel weights as follows:

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

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

$$\mathbb{P}\left(\left| \sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} \sqrt{D_{T,u,h}} \cdot w_{t,T}(u,h)\boldsymbol{X}_{it} \right| \geq x \right)$$

$$\leq C_1 \frac{\left(\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} |\sqrt{D_{T,u,h}} \cdot w_{t,T}(u,h)|^{q'} \right) \|\boldsymbol{X}_{i\cdot}\|_{q',\alpha}^{q'}}{x^{q'}}$$

$$+ C_2 \exp\left(-\frac{C_3 x^2}{D_{T,u,h}\|\boldsymbol{X}_{i\cdot}\|_{2,\alpha}^2} \right) \qquad \text{(A.21)}$$

for any $x > 0$, where $\|\boldsymbol{X}_{i\cdot}\|_{q',\alpha}^{q'} = \sup_{t\geq 0}(t+1)^\alpha \sum_{s=t}^{\infty} \delta_{q'}(\boldsymbol{h}_i,s)$ is the dependence adjusted norm introduced in Definition A.1 and $\|\boldsymbol{X}_{i\cdot}\|_{q',\alpha}^{q'} < \infty$ by Assumption (C6). From (A.21), it follows that for any $\delta > 0$,

$$\mathbb{P}\left(\max_{(u,h)\in\mathcal{G}_T} \left| \sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} w_{t,T}(u,h)\boldsymbol{X}_{it} \right| \geq \delta T^{1/q} \right)$$

$$\leq \sum_{(u,h)\in\mathcal{G}_T} \mathbb{P}\left(\left| \sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} w_{t,T}(u,h)\boldsymbol{X}_{it} \right| \geq \delta T^{1/q} \right)$$

$$= \sum_{(u,h)\in\mathcal{G}_T} \mathbb{P}\left(\left| \sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} \sqrt{D_{T,u,h}}\cdot w_{t,T}(u,h)\boldsymbol{X}_{it} \right| \geq \delta\sqrt{D_{T,u,h}}T^{1/q} \right)$$

$$\leq \sum_{(u,h)\in\mathcal{G}_T} \left[C_1 \frac{(\sqrt{D_{T,u,h}})^{q'}\left(\sum |w_{t,T}(u,h)|^{q'}\right)\|\boldsymbol{X}_{i\cdot}\|_{q',\alpha}^{q'}}{\left(\delta\sqrt{D_{T,u,h}}T^{1/q}\right)^{q'}} + C_2\exp\left(-\frac{C_3\left(\delta\sqrt{D_{T,u,h}}T^{1/q}\right)^2}{D_{T,u,h}\|\boldsymbol{X}_{i\cdot}\|_{2,\alpha}^2} \right) \right]$$

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

$$\leq C_1 \frac{T^\theta\|\boldsymbol{X}_{i\cdot}\|_{q',\alpha}^{q'}}{\delta^{q'}T^{q'/q}} \max_{(u,h)\in\mathcal{G}_T}\left(\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} |w_{t,T}(u,h)|^{q'} \right) + C_2 T^\theta \exp\left(-\frac{C_3\delta^2 T^{2/q}}{\|\boldsymbol{X}_{i\cdot}\|_{2,\alpha}^2} \right)$$

$$= C\frac{T^{\theta-q'/q}}{\delta^{q'}} + CT^\theta \exp\left(-CT^{2/q}\delta^2 \right), \tag{A.22}$$

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.20), 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,

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

Since $\theta - q'/q < 0$ by Assumption (C4), the bound in (A.22) converges to 0 as $T \to \infty$ for any fixed $\delta > 0$. Consequently, we obtain that

$$\max_{(u,h)\in\mathcal{G}_T}\left| \sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} w_{t,T}(u,h)\boldsymbol{X}_{it} \right| = o_p(T^{1/q}). \tag{A.23}$$

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

$$B_{n,T} = \max_{1\leq i\leq n} \max_{(u,h)\in\mathcal{G}_T} \left| (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top \sum_{t=1}^T w_{t,T}(u,h)\boldsymbol{X}_{it} \right| = o_p\left(\frac{T^{1/q}}{\sqrt{T}} \right).$$

We finally turn to $C_{n,T}$. Straightforward calculations yield that $|\sum_{t=1}^T w_{t,T}(u,h)| \leq C\sqrt{T}h_{\max} = o(\sqrt{T})$. 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

$$C_{n,T} = \max_{1 \leq i \leq n} \left| (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top \bar{\boldsymbol{X}}_i \right| \max_{(u,h) \in \mathcal{G}_T} \left| \sum_{t=1}^{T} w_{t,T}(u,h) \right| = o_p\left(\frac{1}{\sqrt{T}}\right).$$

To summarize, we have shown that

$$
\begin{aligned}
\left| \widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T} \right| &\leq o_p(\rho_T) A_{n,T} + O_p(1)\{2B_{n,T} + 2C_{n,T}\} \\
&= o_p(\rho_T) O_p(\sqrt{\log T}) + o_p\left(\frac{T^{1/q}}{\sqrt{T}}\right) + o_p\left(\frac{1}{\sqrt{T}}\right).
\end{aligned}
$$

This immediately implies the desired result. $\qquad\square$

Proof of Proposition 4.3

We first show that

$$\mathbb{P}(\Phi_{n,T} \leq q_{n,T}(\alpha)) = 1 - \alpha. \tag{A.24}$$

We proceed by contradiction. Suppose that (A.24) does not hold true. Since $\mathbb{P}(\Phi_{n,T} \leq q_{n,T}(\alpha)) \geq 1 - \alpha$ by definition of the quantile $q_{n,T}(\alpha)$, there exists $\xi > 0$ such that $\mathbb{P}(\Phi_{n,T} \leq q_{n,T}(\alpha)) = 1 - \alpha + \xi$. From the proof of Proposition A.8, we know that for any $\delta > 0$,

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

Hence,

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

for $\delta > 0$ small enough. This contradicts the definition of the quantile $q_{n,T}(\alpha)$ according to which $q_{n,T}(\alpha) = \inf_{q \in \mathbb{R}}\{\mathbb{P}(\Phi_{n,T} \leq q) \geq 1 - \alpha\}$. We thus arrive at (A.24). Proposition 4.3 is a simple consequence of Theorem 4.1 and equation (A.24). Specifically, we obtain that under H_0,

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

Proof of Proposition 4.4

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

$$\lambda(h) = \sqrt{2\log\{1/(2h)\}} \le \sqrt{2\log\{1/(2h_{\min})\}} \le C\sqrt{\log T}. \tag{A.25}$$

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

$$\widehat{\psi}_{ij,T}^A(u,h) = \sum_{t=1}^T w_{t,T}(u,h)\big\{(\varepsilon_{it} - \bar{\varepsilon}_i) + (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top(\boldsymbol{X}_{it} - \bar{\boldsymbol{X}}_i) - \bar{m}_{i,T}$$
$$- (\varepsilon_{jt} - \bar{\varepsilon}_j) - (\boldsymbol{\beta}_j - \widehat{\boldsymbol{\beta}}_j)^\top(\boldsymbol{X}_{jt} - \bar{\boldsymbol{X}}_j) + \bar{m}_{j,T}\big\}$$
$$\widehat{\psi}_{ij,T}^B(u,h) = \sum_{t=1}^T w_{t,T}(u,h)\Big(m_{i,T}\Big(\frac{t}{T}\Big) - m_{j,T}\Big(\frac{t}{T}\Big)\Big),$$

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

$$m_{i,T}(w) - m_{j,T}(w) \ge c_T\sqrt{\log T/(Th_0)} \tag{A.26}$$

for all $w \in [u_0 - h_0, u_0 + h_0]$.
We first derive a lower bound on the term $\widehat{\psi}_{ij,T}^B(u_0, h_0)$. 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,

$$w_{t,T}(u_0, h_0) = \frac{K\Big(\frac{\frac{t}{T} - u_0}{h_0}\Big)S_{T,2}(u_0, h_0)}{\Big\{\sum_{t=1}^T K^2\Big(\frac{\frac{t}{T} - u_0}{h_0}\Big)S_{T,2}^2(u_0, h_0)\Big\}^{1/2}}$$
$$= \frac{K\Big(\frac{\frac{t}{T} - u_0}{h_0}\Big)}{\Big\{\sum_{t=1}^T K^2\Big(\frac{\frac{t}{T} - u_0}{h_0}\Big)\Big\}^{1/2}} \ge 0.$$

Together with (A.26), this implies that

$$\widehat{\psi}_{ij,T}^B(u_0, h_0) \ge c_T\sqrt{\frac{\log T}{Th_0}}\sum_{t=1}^T w_{t,T}(u_0, h_0). \tag{A.27}$$

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,

$$\Big|\frac{1}{Th}\sum_{t=1}^T K\Big(\frac{\frac{t}{T} - u}{h}\Big)\Big(\frac{\frac{t}{T} - u}{h}\Big)^\ell - \int_0^1 \frac{1}{h}K\Big(\frac{w - u}{h}\Big)\Big(\frac{w - u}{h}\Big)^\ell dw\Big| \le \frac{C}{Th}, \tag{A.28}$$

where the constant C does not depend on u, h and T. With the help of (A.28), we

obtain that for any $(u, h) \in \mathcal{G}_T$ with $[u - h, u + h] \subseteq [0, 1]$,

$$\Big| \sum_{t=1}^{T} w_{t,T}(u, h) - \frac{\sqrt{Th}}{\kappa} \Big| \leq \frac{C}{\sqrt{Th}}, \qquad (A.29)$$

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.29), 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.27) allows us to infer that

$$\widehat{\psi}_{ij,T}^{B}(u_0, h_0) \geq \frac{c_T \sqrt{\log T}}{2\kappa} \qquad (A.30)$$

for sufficiently large T.

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

$$\max_{1 \leq i < j \leq n} \max_{(u,h) \in \mathcal{G}_T} \Big| \widehat{\psi}_{ij,T}^{A}(u, h) \Big|$$
$$= \max_{1 \leq i < j \leq n} \max_{(u,h) \in \mathcal{G}_T} \Big| \widehat{\phi}_{ij,T}(u, h) + (\bar{m}_{j,T} - \bar{m}_{i,T}) \sum_{t=1}^{T} w_{t,T}(u, h) \Big| = O_p(\sqrt{\log T}). \quad (A.31)$$

With the help of (A.30), (A.31), (A.25) and the assumption that $\widehat{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$, we finally arrive at

$$\begin{aligned} \widehat{\Psi}_{n,T} &\geq \max_{1 \leq i < j \leq n} \max_{(u,h) \in \mathcal{G}_T} \frac{|\widehat{\psi}_{ij,T}^{B}(u, h)|}{\{\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\}^{1/2}} - \max_{1 \leq i < j \leq n} \max_{(u,h) \in \mathcal{G}_T} \left\{ \frac{|\widehat{\psi}_{ij,T}^{A}(u, h)|}{\{\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\}^{1/2}} + \lambda(h) \right\} \\ &= \max_{1 \leq i < j \leq n} \max_{(u,h) \in \mathcal{G}_T} \frac{|\widehat{\psi}_{ij,T}^{B}(u, h)|}{\{\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\}^{1/2}} + O_p(\sqrt{\log T}) \\ &\geq \frac{c_T \sqrt{\log T}}{2\kappa} \min_{1 \leq i < j \leq n} \{\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\}^{-1/2} + O_p(\sqrt{\log T}) \\ &= \frac{c_T \sqrt{\log T}}{2\kappa} O_p(1) + O_p(\sqrt{\log T}). \end{aligned} \qquad (A.32)$$

Since $q_{n,T}(\alpha) = O(\sqrt{\log T})$ for any fixed $\alpha \in (0, 1)$ and $c_T \to \infty$, (A.32) 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

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

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

Proof of Proposition 5.1

For the sake of brevity, we introduce the following notation. For each i and j, we define the statistic $\widehat{\Psi}_{ij,T} := \max_{(u,h) \in \mathcal{G}_T} \hat{\psi}_{ij,T}^0(u, h)$ 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

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

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

$$
\max_{1 \leq \ell \leq N} \max_{i,j \in G_\ell} \widehat{\Psi}_{ij,T} < \min_{1 \leq \ell < \ell' \leq N} \min_{\substack{i \in G_\ell, \\ j \in G_{\ell'}}} \widehat{\Psi}_{ij,T}. \tag{A.33}
$$

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

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

Moreover, the same arguments as for Proposition 4.4 show that

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

Taken together, these two statements imply that

$$\mathbb{P}\big(B_{n,T}\big)\geq(1-\alpha)+o(1). \tag{A.34}$$

In what follows, we show that on the event $B_{n,T}$, (i) $\{\widehat{G}_1^{[n-N]},\ldots,\widehat{G}_N^{[n-N]}\}=\{G_1,\ldots\ldots,G_N\}$ and (ii) $\widehat{N}=N$. From (i), (ii) and (A.34), the statements of Proposition 5.1 follow immediately.

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 $\widehat{G}_i^{[0]}=\{i\}$ and $\widehat{G}_j^{[0]}=\{j\}$ with i and j belonging to the same group G_k. This is a direct consequence of (A.33). The algorithm thus produces a partition $\{\widehat{G}_1^{[1]},\ldots,\widehat{G}_{n-1}^{[1]}\}$ 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.33), 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 $r\leq n-N$, 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 $\{\widehat{G}_1^{[n-N]},\ldots,\widehat{G}_N^{[n-N]}\}=\{G_1,\ldots,G_N\}$. $\qquad\square$

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 $B_{n,T}$, $\max_{1\leq\ell\leq r}\widehat{\Delta}(\widehat{G}_\ell^{[n-r]})>q_{n,T}(\alpha)$ 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, $\widehat{\Delta}(G_\ell)\leq q_{n,T}(\alpha)$ for any ℓ. Hence, we obtain that $\max_{1\leq\ell\leq N}\widehat{\Delta}(\widehat{G}_\ell^{[n-N]})=\max_{1\leq\ell\leq N}\widehat{\Delta}(G_\ell)\leq q_{n,T}(\alpha)$.

Putting everything together, we can conclude that on the event $B_{n,T}$,

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

that is, $\widehat{N} = N$. □

Proof of Proposition 5.2

We consider the event

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

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

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

with $\widehat{\psi}_{ij,T}^{\text{trend}}(u,h) = \sum_{t=1}^T w_{t,T}(u,h)\{(m_{i,T}(t/T) - \bar{m}_{i,T}) - (m_{j,T}(t/T) - \bar{m}_{j,T})\}$ and $\bar{m}_{i,T} = T^{-1} \sum_{t=1}^T m_{i,T}(t/T)$. 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 5.1. In particular, analogous to (A.34) and statements (i) and (ii) in this proof, we can show that

$$\mathbb{P}\big(D_{n,T}\big) \ge (1 - \alpha) + o(1) \tag{A.35}$$

and

$$D_{n,T} \subseteq \big\{ \widehat{N} = N \text{ and } \widehat{G}_\ell = G_\ell \text{ for all } \ell \big\}. \tag{A.36}$$

Moreover, we have that

$$D_{n,T} \subseteq \bigcap_{1 \le \ell < \ell' \le \widehat{N}} E_{n,T}^{[\ell,\ell']}(\alpha), \tag{A.37}$$

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

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

it holds that $\widehat{\psi}_{ij,T}^{\text{trend}}(u,h) \ne 0$, which in turn implies that $m_i(v) - m_j(v) \ne 0$ for some $v \in I_{u,h}$. From (A.36) and (A.37), we obtain that

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

This together with (A.35) implies that $\mathbb{P}(E_{n,T}(\alpha)) \ge (1 - \alpha) + o(1)$.