# Supplement to "Multiscale Comparison of Nonparametric Trend Curves"

## A Simulations

### A.1 Size and power calculations

We now investigate our testing and clustering methods by means of a simulation study. The simulation design is as follows: We generate data from the model $Y_{it} = m_i(\frac{t}{T}) + \boldsymbol{\beta}_i^\top \boldsymbol{X}_{it} + \alpha_i + \varepsilon_{it}$, where the number of time series is set to $n = 15$ and we consider different time series lengths $T$. We include 3 covariates and model them by the following VAR(3) process:

$$\underbrace{\begin{pmatrix} X_{it,1} \\ X_{it,2} \\ X_{it,3} \end{pmatrix}}_{=:\boldsymbol{X}_{it}} = \begin{pmatrix} a_1 & 0 & 0 \\ 0 & a_2 & 0 \\ 0 & 0 & a_3 \end{pmatrix} \begin{pmatrix} X_{it-1,1} \\ X_{it-1,2} \\ X_{it-1,3} \end{pmatrix} + \underbrace{\begin{pmatrix} \nu_{it,1} \\ \nu_{it,2} \\ \nu_{it,3} \end{pmatrix}}_{=:\boldsymbol{\nu}_{it}}.$$

We choose $a_1 = a_2 = a_3 = 0.25$. The innovations $\boldsymbol{\nu}_{it}$ are drawn i.i.d. from a multivariate normal distribution $N(0, \Phi)$ with

$$\Phi = \begin{pmatrix} 1 & \varphi & \varphi \\ \varphi & 1 & \varphi \\ \varphi & \varphi & 1 \end{pmatrix},$$

where $\varphi \in \{0.1, 0.25\}$ specifies the correlation between the 3 covariates. We set $\boldsymbol{\beta}_i = (\beta_{i,1}, \beta_{i,2}, \beta_{i,3}) = (1, 1, 1)$ for all $i$. As for the fixed effect term, we let $\alpha = (\alpha_1, \ldots, \alpha_n)$ be a

normally distributed random vector. In particular, $\alpha \sim N(0, \Sigma)$ with

$$\Sigma = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho & \cdots & \rho & 1 \end{pmatrix},$$

where $\rho \in \{0.1, 0.25\}$ gives the correlation across $i$.

For each $i$, the errors $\varepsilon_{it}$ follow the AR(1) model $\varepsilon_{it} = a\varepsilon_{i,t-1} + \eta_{it}$, where $a = 0.25$ and the innovations $\eta_{it}$ are i.i.d. normally distributed with mean 0 and standard deviation 0.25. We assume that the covariates $\boldsymbol{X}_{it}$ and the error terms $\varepsilon_{js}$ 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 use the bump function

$$\begin{aligned} m_1(u) = b \cdot \mathbb{1}(u \in [0.2, 0.4]) \cdot \left(1 - \left\{\frac{u - 0.3}{0.1}\right\}^2\right)^2 \\ - b \cdot \mathbb{1}(u \in [0.6, 0.8]) \cdot \left(1 - \left\{\frac{u - 0.7}{0.1}\right\}^2\right)^2 \end{aligned}$$

for $b \in \{0.25, 0.5, 0.75\}$ (depicted in Figure A.1) and $m_i = 0$ for $i \neq 1$. 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 samples.

Our multiscale test is implemented as follows: The estimators $\widehat{\boldsymbol{\beta}}_i$ and $\widehat{\alpha}_i$ of the unknown parameters $\boldsymbol{\beta}_i$ and $\alpha_i$ are computed as described in Section 3.1. Since the errors $\varepsilon_{it}$ follow an AR(1) process, we estimate the long-run error variance $\sigma_i^2$ 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 $U_T = \left\{u \in [0,1] : u = \frac{5t}{T} \text{ for some } t \in \mathbb{N}\right\}$ and $H_T = \left\{h \in \left[\frac{\log T}{T}, \frac{1}{4}\right] : h = \frac{5t-3}{T} \text{ for some } t \in \mathbb{N}\right\}$. We thus 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

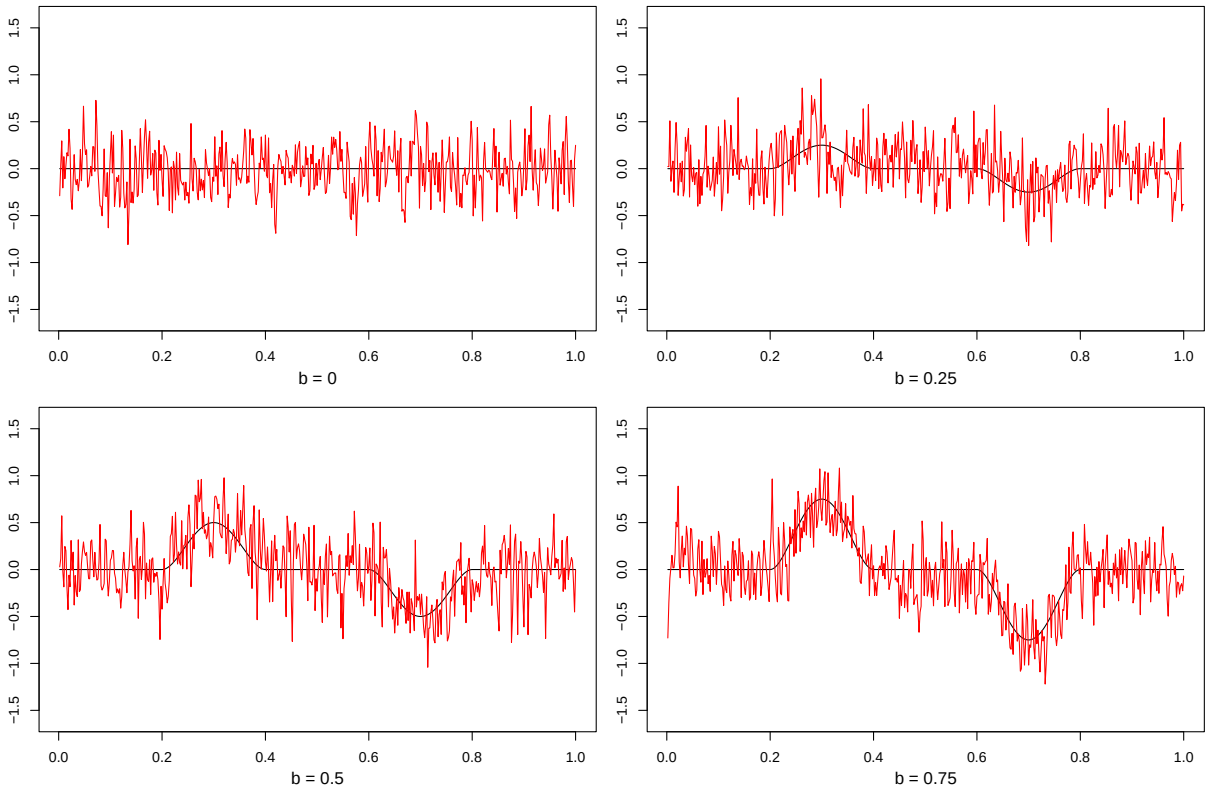


Figure A.1: In black, the bump function $m_1$ is plotted for different heights of the bump: $b \in \{0, 0.25, 0.5, 0.75\}$ with $b = 0$ corresponding to the null $H_0$. In red, we depict the time series $\{m_1(t/T) + \varepsilon_{1t} : 1 \leq t \leq T\}$ for one simulation run and $T = 500$.

by Monte Carlo methods as described in Section 3.4, where we set $L = 5000$.

The results of the simulation study are presented in Tables **??** and **??** for $\phi = 0.1$ and $\rho = 0.1$ and in Tables **??** and **??** for $\phi = 0.25$ and $\rho = 0.25$. We do not present the results for $\phi = 0.1, \rho = 0.25$ and $\phi = 0.25, \rho = 0.1$ as they are fairly similar and do not bring additional insights. The actual size of the mutliscale test is reported in Tables **??** and **??**, while Tables **??** and **??** report 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 Tables **??** and **??**, one can see that the empirical size gives a reasonable approximation to the target $\alpha$ in all scenarios under investigation, even though the size numbers have a slight upward bias. This bias gets smaller as the sample size $T$ increases, which

reflects the fact that we get more and more information for each of the tests that we need to carry out. We can also see that the upward bias become slightly more pronounced in case of stronger degrees of correlation across the covariates and the time series, i.e., with higher values of $\phi$ and $\rho$, but the change in test performance is almost negligible and can be attributed to random sampling. To summarize, even though slightly liberal, the test controls the FWER quite accurately in the simulation settings that we consider.

Inspecting Tables **??** and **??**, one can further see that the test has reasonable power properties. For the smallest height of the bump function $b = 0.25$ and the smallest time series length $T = 100$, the power is only moderate, reflecting the fact that the alternative with $b = 0.25$ is not very far away from the null. However, as we increase the height $b$ and the sample size $T$, the power increases quickly. Already for the height value $b = 0.5$ in all simulation scenarios, we reach a power of 0.98 for $T = 250$ and for all nominal sizes $\alpha$.

## A.2 Clustering algorithm

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 $Y_{it} = m_i(\frac{t}{T}) + \varepsilon_{it}$, i.e., we assume that there are no fixed effects and no covariates. The error terms $\varepsilon_{it}$ 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 $G_1 = \{1, \dots, 5\}$, $G_2 = \{6, \dots, 10\}$ and $G_3 = \{11, \dots, 15\}$, 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 $f_1(u) = 0$, $f_2(u) = 0.35 \cdot \mathbb{1}\Big\{\frac{|u-0.25|}{0.25} \leq 1\Big\}\big(1 - \frac{(u-0.25)^2}{0.25^2}\big)^2 - 0.35 \cdot \mathbb{1}\Big\{\frac{|u-0.75|}{0.25} \leq 1\Big\}\big(1 - \frac{(u-0.75)^2}{0.25^2}\big)^2$ and $f_3(u) = \mathbb{1}\Big\{\frac{|u-0.75|}{0.025} \leq 1\Big\}\big(1 - \frac{(u-0.75)^2}{0.025^2}\big)^2 - \mathbb{1}\Big\{\frac{|u-0.25|}{0.025} \leq 1\Big\}\big(1 - \frac{(u-0.25)^2}{0.025^2}\big)^2$ (depicted in Figure A.2). 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.

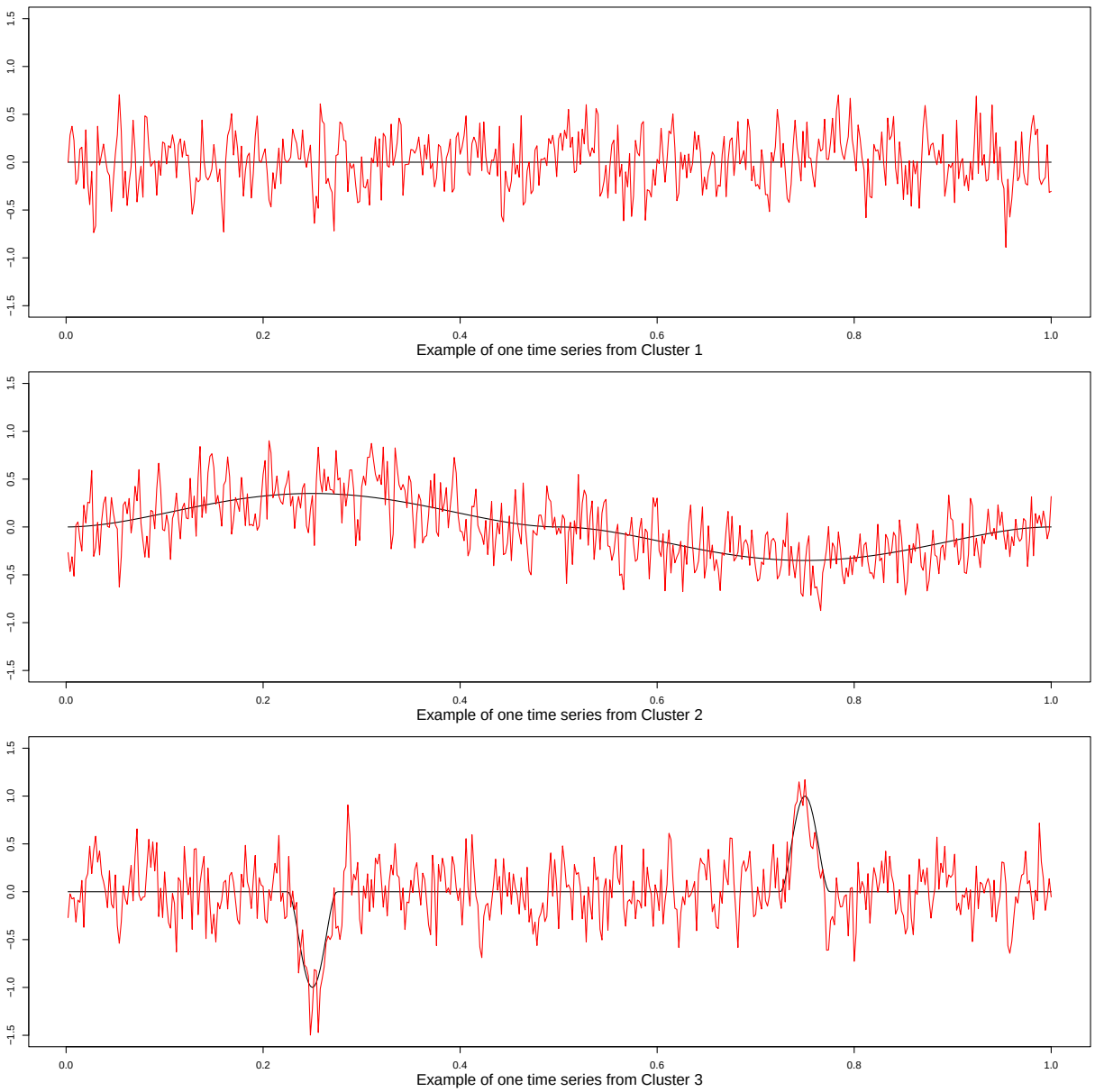


Figure A.2: In black, the group-specific trend functions $f_1, f_2$ and $f_3$ are plotted. In red, we depict the time series $\{f_l(t/T) + \varepsilon_{it} : 1 \leq t \leq T\}$ for $l = 1, 2, 3$ for one simulation run and $T = 500$.

The simulation results are reported in Table A.1. The entries in Table A.1a are computed as the number of simulations for which $\widehat{N} = N$ divided by the total number of simulations. They thus specify the empirical probabilities with which the estimate $\widehat{N}$ is equal to the true number of groups $N = 3$. Analogously, the entries of Table A.1b give the empirical probabilities with which the estimated group structure $\{\widehat{G}_1, \ldots, \widehat{G}_{\widehat{N}}\}$ equals the true one $\{G_1, G_2, G_3\}$. The results in Table A.1 nicely illustrate the theoretical properties of our

Table A.1: Clustering results for different sample sizes $T$ and nominal sizes $\alpha$.

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

| | nominal size $\alpha$ | | |
|---|---|---|---|
| $T$ | 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 $\alpha$ | | |
|---|---|---|---|
| $T$ | 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 |

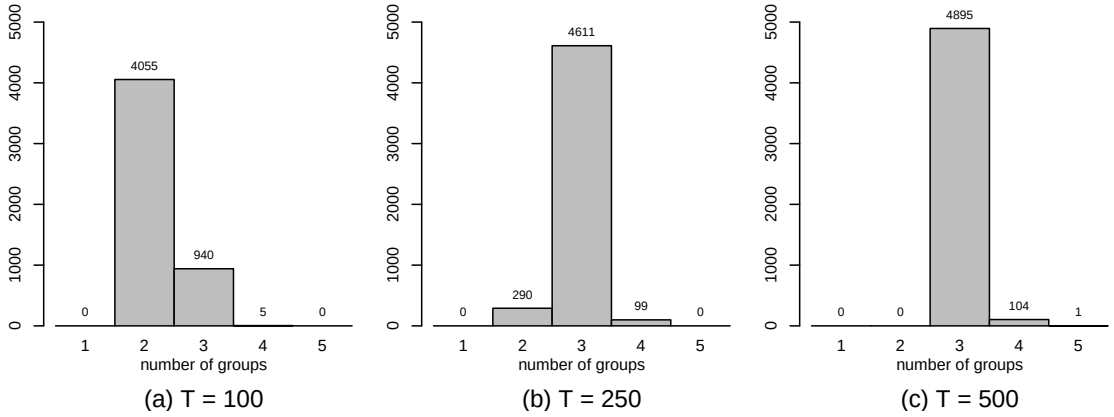


Figure A.3: Estimated number of groups $\widehat{N}$ for nominal size $\alpha = 0.05$.

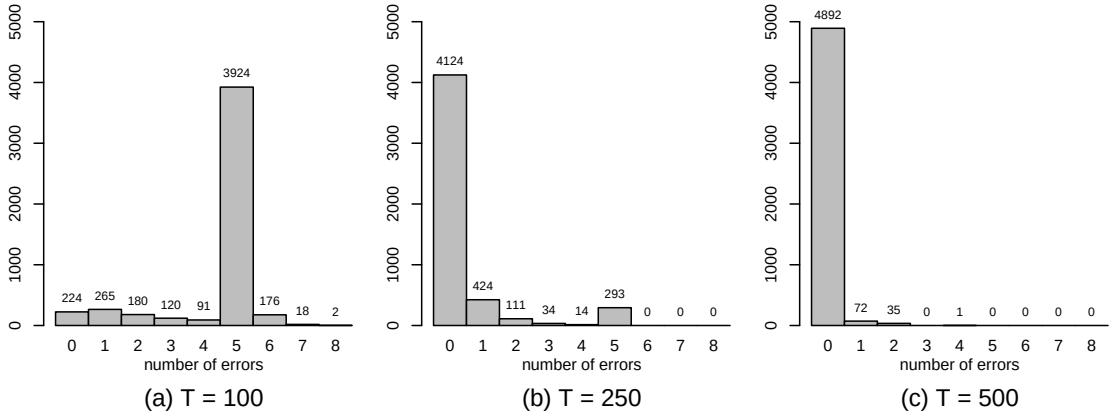


Figure A.4: Number of classification errors for nominal size $\alpha = 0.05$.

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 A.1 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 A.3 and A.4 give more insight into what happens for $T = 100$ and $T = 250$. Figure A.3 shows histograms of the 5000 simulated values of $\widehat{N}$, while Figure A.4 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 A.3 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 A.4 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 can conclude that the small empirical probabilities for $T = 100$ in Table A.1 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 A.1 are still somewhat below the target $(1-\alpha)$.

## A.3 Results for larger $n$

In order to examine the performance of our method for larger $n$, we rerun our simulation exercises from Section A.1 for $n \in \{25, 50, 100\}$ and $T$ as before, i.e., $T \in \{100, 250, 500\}$. We work with precisely the same simulation setup as specified in Section A.1 except for the fact that we use a smaller grid $\mathcal{G}_T^{\text{sparse}}$ to decrease the computational burden. Specifically,

we take the grid

$$\mathcal{G}_T^{\text{sparse}} = \big\{(u,h) \subseteq [0,1] : (u,h) = ((2s+1)h, h) \text{ for } s = 0, \dots, \Big\lfloor \frac{h^{-1}-1}{2} \Big\rfloor \text{ and } h \in H_T^{\text{sparse}} \big\},$$

that is, we work with a dyadic scheme (as in Wavelet analysis) with scales in the set

$$H_T^{\text{sparse}} = \big\{h = 2^k h_{\min} \text{ for } k = 0, \dots, K\big\},$$

where $h_{\min} = \frac{\lceil \log T \rceil}{T}$ and $K$ is such that $2^K h_{\min} \leq \frac{1}{4}$, i.e.,

$$K \leq \Big\lfloor \log\Big(\frac{T}{4\lceil \log T \rceil}\Big) \Big\rfloor \frac{1}{\log(2)}.$$

Overall, the simulation results show that the method performs well for larger $n$. The results are presented in Tables A.2 - A.6 (we report the results for the case $\phi = 0.25$ and $\rho = 0.25$ since the results for other combination of values of $\phi$ and $\rho$ are fairly similar). As can be seen in Table A.2, the empirical size provides an appropriate approximation to the target alpha across all scenarios under consideration. As in Section A.1, we observe a slight upward bias for $T = 250$ and $T = 500$ which becomes somewhat more pronounced with larger values of $n$. In regards to the empirical power calculations, the test demonstrates significant power across all simulated scenarios, as illustrated in Tables A.3 - A.6. Notably, despite slight variations from the original simulated scenario discussed in Section A.1, the observed changes are not substantial which affirms the good performance of the proposed test in case of large number of time series $n$.

## A.4 Robustness checks on the choice of tuning parameters

The main tuning parameter of our method is the grid $\mathcal{G}_T$ which in particular specifies the bandwidths that are taken into account by the procedure. In order to asses the robustness of the proposed multiscale test with respect to the grid, we repeat the analysis from Section A.1 with a different choice of a grid. Specifically, we make the grid denser: $\mathcal{G}_T^{\text{dense}} = U_T^{\text{dense}} \times$

Table A.2: Size of the multiscale test for different number of time series $n$, sample sizes $T$ and nominal sizes $\alpha$ for the dyadic grid. Each panel corresponds to a different number of time series $n$.

(a) $n = 15$

| | nominal size $\alpha$ | | |
|---|---|---|---|
| $T$ | 0.01 | 0.05 | 0.1 |
| 100 | 0.008 | 0.033 | 0.068 |
| 250 | 0.012 | 0.060 | 0.105 |
| 500 | 0.012 | 0.053 | 0.108 |

(b) $n = 25$

| | nominal size $\alpha$ | | |
|---|---|---|---|
| $T$ | 0.01 | 0.05 | 0.1 |
| 100 | 0.005 | 0.032 | 0.065 |
| 250 | 0.012 | 0.054 | 0.098 |
| 500 | 0.018 | 0.059 | 0.108 |

(c) $n = 50$

| | nominal size $\alpha$ | | |
|---|---|---|---|
| $T$ | 0.01 | 0.05 | 0.1 |
| 100 | 0.005 | 0.034 | 0.066 |
| 250 | 0.015 | 0.061 | 0.110 |
| 500 | 0.010 | 0.068 | 0.128 |

(d) $n = 100$

| | nominal size $\alpha$ | | |
|---|---|---|---|
| $T$ | 0.01 | 0.05 | 0.1 |
| 100 | 0.004 | 0.033 | 0.057 |
| 250 | 0.015 | 0.069 | 0.114 |
| 500 | 0.017 | 0.069 | 0.119 |

$H_T^{\text{dense}}$ with $U_T^{\text{dense}} = \{u \in [0,1] : u = \frac{t}{T} \text{ for some } t \in \mathbb{N}\}$ and $H_T^{\text{dense}} = \{h \in [\frac{\log T}{T}, \frac{1}{4}] : h = \frac{2t}{T} \text{ for some } t \in \mathbb{N}\}$. We thus consider intervals $\mathcal{I}_{u,h} = [u-h, u+h] = [t/T - h, t/T + h]$ for all possible $t \in \{1, \ldots, T-1\}$ which contain $5, 9, 13, \ldots$ data points. Other than the grid, we work precisely with the same simulation setup as in Section A.1.

The results of the simulation study with a finer grid are presented in Tables ?? and ?? (in Tables ?? and ??). We report the results for the simulation scenario with stronger degrees of correlation across the time series and between the covariates, i.e., with $\rho = 0.25$ and $\phi = 0.25$, since this scenario is less favourable for our multiscale procedure and the results with different values of $\rho$ and $\phi$ are fairly similar. Although there are minor deviations from the results for the original simulated scenario presented in Section A.1, these variations are not substantial which confirms the robust performance of the proposed test when applied to a finer grid.

Furthermore, as we have seen in Section A.3, the behaviour of our test does not substantially change in case of a much sparser grid $\mathcal{G}_T^{\text{sparse}}$. Hence, we can make a conclusion that as long as the set of pairs of location-bandwidths is chosen sufficiently rich (in the sense of including a variety of bandwidth values ranging from very small to very large), the particular choice

Table A.3: Power of the multiscale test for $n = 15$ for different sample sizes $T$ and nominal sizes $\alpha$ for the dyadic grid. Each panel corresponds to a different height parameter $b$ of the bump function.

(a) $b = 0.25$

| | nominal size $\alpha$ | | |
|---|---|---|---|
| $T$ | 0.01 | 0.05 | 0.1 |
| 100 | 0.010 | 0.038 | 0.082 |
| 250 | 0.072 | 0.193 | 0.280 |
| 500 | 0.188 | 0.417 | 0.571 |

(b) $b = 0.50$

| | nominal size $\alpha$ | | |
|---|---|---|---|
| $T$ | 0.01 | 0.05 | 0.1 |
| 100 | 0.012 | 0.048 | 0.101 |
| 250 | 0.479 | 0.712 | 0.801 |
| 500 | 0.995 | 1.000 | 1.000 |

(c) $b = 0.75$

| | nominal size $\alpha$ | | |
|---|---|---|---|
| $T$ | 0.01 | 0.05 | 0.1 |
| 100 | 0.007 | 0.029 | 0.066 |
| 250 | 0.742 | 0.903 | 0.943 |
| 500 | 1.000 | 1.000 | 1.000 |

Table A.4: Power of the multiscale test for $n = 25$ for different sample sizes $T$ and nominal sizes $\alpha$ for the dyadic grid. Each panel corresponds to a different height parameter $b$ of the bump function.

(a) $b = 0.25$

| | nominal size $\alpha$ | | |
|---|---|---|---|
| $T$ | 0.01 | 0.05 | 0.1 |
| 100 | 0.008 | 0.042 | 0.080 |
| 250 | 0.056 | 0.159 | 0.237 |
| 500 | 0.181 | 0.378 | 0.512 |

(b) $b = 0.50$

| | nominal size $\alpha$ | | |
|---|---|---|---|
| $T$ | 0.01 | 0.05 | 0.1 |
| 100 | 0.008 | 0.045 | 0.089 |
| 250 | 0.410 | 0.634 | 0.739 |
| 500 | 0.994 | 0.999 | 1.000 |

(c) $b = 0.75$

| | nominal size $\alpha$ | | |
|---|---|---|---|
| $T$ | 0.01 | 0.05 | 0.1 |
| 100 | 0.005 | 0.032 | 0.062 |
| 250 | 0.656 | 0.848 | 0.907 |
| 500 | 1.000 | 1.000 | 1.000 |

Table A.5: Power of the multiscale test for $n = 50$ for different sample sizes $T$ and nominal sizes $\alpha$ for the dyadic grid . Each panel corresponds to a different height parameter $b$ of the bump function.

(a) $b = 0.25$

| | nominal size $\alpha$ | | |
|---|---|---|---|
| $T$ | 0.01 | 0.05 | 0.1 |
| 100 | 0.008 | 0.039 | 0.078 |
| 250 | 0.046 | 0.134 | 0.215 |
| 500 | 0.106 | 0.329 | 0.453 |

(b) $b = 0.50$

| | nominal size $\alpha$ | | |
|---|---|---|---|
| $T$ | 0.01 | 0.05 | 0.1 |
| 100 | 0.008 | 0.041 | 0.077 |
| 250 | 0.364 | 0.589 | 0.691 |
| 500 | 0.983 | 0.999 | 1.000 |

(c) $b = 0.75$

| | nominal size $\alpha$ | | |
|---|---|---|---|
| $T$ | 0.01 | 0.05 | 0.1 |
| 100 | 0.006 | 0.033 | 0.063 |
| 250 | 0.615 | 0.798 | 0.874 |
| 500 | 1.000 | 1.000 | 1.000 |

Table A.6: Power of the multiscale test for $n = 100$ for different sample sizes $T$ and nominal sizes $\alpha$ for the dyadic grid. Each panel corresponds to a different height parameter $b$ of the bump function.

(a) $b = 0.25$

| | nominal size $\alpha$ | | |
|---|---|---|---|
| $T$ | 0.01 | 0.05 | 0.1 |
| 100 | 0.006 | 0.032 | 0.063 |
| 250 | 0.040 | 0.127 | 0.197 |
| 500 | 0.102 | 0.261 | 0.374 |

(b) $b = 0.50$

| | nominal size $\alpha$ | | |
|---|---|---|---|
| $T$ | 0.01 | 0.05 | 0.1 |
| 100 | 0.008 | 0.033 | 0.064 |
| 250 | 0.305 | 0.506 | 0.618 |
| 500 | 0.982 | 0.997 | 1.000 |

(c) $b = 0.75$

| | nominal size $\alpha$ | | |
|---|---|---|---|
| $T$ | 0.01 | 0.05 | 0.1 |
| 100 | 0.006 | 0.032 | 0.058 |
| 250 | 0.513 | 0.714 | 0.801 |
| 500 | 1.000 | 1.000 | 1.000 |

of a grid can be expected to have a negligible effect on the procedure.

Apart from the grid, our method depends on (i) the number of bootstrap samples $L$ to compute the Gaussian quantile, (ii) the kernel $K$, and (iii) secondary tuning parameters for the computation of the long-run error variance. As long as $L$ is chosen large enough (say $L \geq 1000$), the exact value of $L$ should have a negligible effect. We use $L = 5000$ throughout the paper. As a robustness check, we have rerun everything with $L = 1000$, which (as expected) yields almost identical results.

Furthermore, classical nonparametric theory suggests that the choice of kernel is significantly less critical than the choice of bandwidth. Thus, we have not carried out robustness checks with respect to the choice of kernel.

For the estimation of the long-run error variance, we work with the estimator from Khismatullina and Vogt (2020) where extensive robustness checks with respect to the choice of tuning parameters have been carried out. Therefore, we did not conduct additional robustness checks concerning the secondary tuning parameters for computing the long-run error variance.

## A.5 Comparison of our test with SiZer

To the best of our knowledge, the only other test for comparing trend curves with similar properties has been developed in Park et al. (2009) (SiZer). In this section, we compare our method with SiZer. However, we would like to note that the analysis in Park et al. (2009) is mainly methodological and the theory is developed only for the case of $n = 2$ time series. In Park et al. (2009), the authors propose a possible extension to their approach in case of more than 2 time series, but the extension does not allow for pairwise comparison of the time series. Moreover, it is not clear how to calculate the actual size and power in this case. Furthermore, the model considered in Park et al. (2009) is much simpler than ours: it does not include neither the covariates, nor the fixed effects. Hence, in order to allow for fair

comparison between two methods, we consider a simplified version of the simulation setup from Section A.1. In particular, we do the following. We consider a model $Y_{it} = m_i\big(\frac{t}{T}\big) + \varepsilon_{it}$ that does not include neither the covariates, nor the fixed effects as they are not part of the model in Park et al. (2009). We choose the number of time series to be equal to $n = 2$ and consider different time series lengths $T = 100, 250, 500, 750, 1000$. The error terms $\varepsilon_{it}$ are specified as in Section A.1. To generate data under the null $H_0 : m_1 = m_2$, we let $m_i = 0$ for $i = 1, 2$ as before. As before, to produce data under the alternative, we use the bump function

$$m_1(u) = b \cdot \mathbb{1}(u \in [0.2, 0.4]) \cdot \left(1 - \left\{\frac{u - 0.3}{0.1}\right\}^2\right)^2 - b \cdot \mathbb{1}(u \in [0.6, 0.8]) \cdot \left(1 - \left\{\frac{u - 0.7}{0.1}\right\}^2\right)^2$$

for $b \in \{0.25, 0.5\}$ (depicted in Figure A.1) and $m_2 = 0$. We implement our multiscale procedure as in Section A.1 with one important change. In order to make the comparison between the methods as fair as possible, we do not estimate the long-run variance $\sigma_i^2$ from the data but consider $\sigma_i^2$ as known. Specifically, we use the theoretical value of the long-run variance calculated based on the true parameter values:

$$\sigma_i^2 = \frac{E[\eta_{it}^2]}{(1-a)^2} = \frac{1}{9} \text{ for } i = 1, 2.$$

Furthermore, since SiZer depends not on the long-run variance, but on the autocovariance functions of the time series $\gamma_i(\cdot)$ for $i = 1, 2$, in the calculation of the critical values of SiZer we use the following formula for the autocovariance function:

$$\gamma_i(k) = \frac{E[\eta_{it}^2]a^{|k|}}{1 - a^2} = \frac{0.25^{|k|}}{15}.$$

All of the details of the exact implementation of SiZer can be found in the Appendix in Khismatullina and Vogt (2020).

In what follows, we denote the proposed multiscale procedure and SiZer (Park et al. (2009)) by $\mathcal{T}_{\text{MS}}$ and $\mathcal{T}_{\text{SiZer}}$, respectively. The results of the size and power simulation studies compar-

Table A.7: Size comparison of the proposed multiscale test ($\mathcal{T}_{\text{MS}}$) and SiZer ($\mathcal{T}_{\text{SiZer}}$) for different sample sizes $T$ and various significance levels $\alpha$.

| | $\alpha = 0.01$ | | $\alpha = 0.05$ | | $\alpha = 0.1$ | |
|---|---|---|---|---|---|---|
| | $\mathcal{T}_{\text{MS}}$ | $\mathcal{T}_{\text{SiZer}}$ | $\mathcal{T}_{\text{MS}}$ | $\mathcal{T}_{\text{SiZer}}$ | $\mathcal{T}_{\text{MS}}$ | $\mathcal{T}_{\text{SiZer}}$ |
| $T = 100$ | 0.010 | 0.099 | 0.047 | 0.332 | 0.106 | 0.524 |
| $T = 250$ | 0.008 | 0.162 | 0.041 | 0.510 | 0.102 | 0.699 |
| $T = 500$ | 0.009 | 0.214 | 0.044 | 0.598 | 0.089 | 0.787 |
| $T = 750$ | 0.008 | 0.252 | 0.045 | 0.657 | 0.095 | 0.844 |
| $T = 1000$ | 0.012 | 0.271 | 0.053 | 0.691 | 0.102 | 0.873 |
| bottomrule | | | | | | |

Table A.8: Power comparison of the proposed multiscale test ($\mathcal{T}_{\text{MS}}$) and SiZer ($\mathcal{T}_{\text{SiZer}}$) for different sample sizes $T$ and various significance levels $\alpha$ for the bump function with $b = 0.25$.

| | $\alpha = 0.01$ | | $\alpha = 0.05$ | | $\alpha = 0.1$ | |
|---|---|---|---|---|---|---|
| | $\mathcal{T}_{\text{MS}}$ | $\mathcal{T}_{\text{SiZer}}$ | $\mathcal{T}_{\text{MS}}$ | $\mathcal{T}_{\text{SiZer}}$ | $\mathcal{T}_{\text{MS}}$ | $\mathcal{T}_{\text{SiZer}}$ |
| $T = 100$ | 0.061 | 0.279 | 0.172 | 0.592 | 0.300 | 0.762 |
| $T = 250$ | 0.213 | 0.655 | 0.451 | 0.906 | 0.608 | 0.964 |
| $T = 500$ | 0.577 | 0.925 | 0.804 | 0.992 | 0.884 | 0.998 |
| $T = 750$ | 0.830 | 0.988 | 0.955 | 0.999 | 0.981 | 1.000 |
| $T = 1000$ | 0.960 | 0.999 | 0.993 | 1.000 | 0.998 | 1.000 |

Table A.9: Power comparison of the proposed multiscale test ($\mathcal{T}_{\text{MS}}$) and SiZer ($\mathcal{T}_{\text{SiZer}}$) for different sample sizes $T$ and various significance levels $\alpha$ for the bump function with $b = 0.5$.

| | $\alpha = 0.01$ | | $\alpha = 0.05$ | | $\alpha = 0.1$ | |
|---|---|---|---|---|---|---|
| | $\mathcal{T}_{\text{MS}}$ | $\mathcal{T}_{\text{SiZer}}$ | $\mathcal{T}_{\text{MS}}$ | $\mathcal{T}_{\text{SiZer}}$ | $\mathcal{T}_{\text{MS}}$ | $\mathcal{T}_{\text{SiZer}}$ |
| $T = 100$ | 0.380 | 0.759 | 0.653 | 0.939 | 0.795 | 0.977 |
| $T = 250$ | 0.954 | 0.999 | 0.993 | 1.000 | 0.999 | 1.000 |
| $T = 500$ | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $T = 750$ | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $T = 1000$ | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

ing $\mathcal{T}_{\text{MS}}$ and $\mathcal{T}_{\text{SiZer}}$ are presented in Tables A.7 and Tables A.8, A.9 respectively. There is an important difference between $\mathcal{T}_{\text{MS}}$ and $\mathcal{T}_{\text{SiZer}}$ to be noticed: $\mathcal{T}_{\text{MS}}$ is a *global* test procedures, whereas $\mathcal{T}_{\text{SiZer}}$ is a scale-wise in essence. This means that $\mathcal{T}_{\text{MS}}$ tests $H_0(u, h)$ simultaneously for all locations $u \in U_T$ and scales $h \in H_T$ and controls the size simultaneously over both locations $u$ and scales $h$, whereas $\mathcal{T}_{\text{SiZer}}$ tests the hypothesis $H_0(u, h)$ simultaneously for all $u \in U_T$ but separately for each scale $h \in H_T$ and, hence, controls the size for each scale $h \in H_T$ separately. This results in $\mathcal{T}_{\text{SiZer}}$ exhibiting a much more liberal nature, as Table

A.7 illustrates. Specifically, the size numbers of the multiscale test $\mathcal{T}_{\text{MS}}$ closely align with the target nominal size levels $\alpha$. Contrarily, the actual size numbers of $\mathcal{T}_{\text{SiZer}}$ are much larger than the target $\alpha$. Since the number of scales $h$ in the grid $\mathcal{G}_T$ increases with $T$, the actual sizes significantly deviate from the target $\alpha$ as the sample size $T$ increases. To summarize, as expected, the global test $\mathcal{T}_{\text{MS}}$ holds the size reasonably well, whereas the row-wise method $\mathcal{T}_{\text{SiZer}}$ is much too liberal.

Regarding the power simulation studies presented in Tables A.8 and A.9, it is worth mentioning that in all simulation scenarios, $\mathcal{T}_{\text{SiZer}}$ is more powerful than $\mathcal{T}_{\text{MS}}$. This gain of power is presumably due to the fact that $\mathcal{T}_{\text{SiZer}}$ is in general too liberal in terms of size as observed in Table A.7.

Mentioning other competitors?

## A.6 Comparison of our clustering algorithm with a benchmark

In this section, we compare the performance of our clustering algorithm against a natural benchmark using the simulation setup described in Section A.2. Specifically, for the benchmark procedure we do the following.

1. Estimate the trends $m_i$ by a local linear estimator $\hat{m}_i$ with a fixed bandwidth. We consider two choices for the bandwidth: $h_1 = \min\{h \in H_t\}$ and $h_2 = \max\{h \in H_T\}$. As we will see further, the choice of bandwidth significantly affects the results.

2. Compute a simple $\mathcal{L}_2$ distance measure $d_{ij}$ between $\hat{m}_i$ and $\hat{m}_j$, e.g.
$$d_{ij} = \int_0^1 (\hat{m}_i(w) - \hat{m}_j(w))^2 dw.$$

3. Construct the following dissimilarity measure from these distances:
$$\hat{\Delta}(S, S') = \max_{i \in S, j \in S'} d_{ij}.$$

4. Run a HAC algorithm with the computed dissimilarities.

Table A.10: Empirical probabilities that $\{\widehat{G}_1, \ldots, \widehat{G}_{\widehat{N}}\} = \{G_1, G_2, G_3\}$ for the proposed clustering approach with estimated number of clusters ($\mathcal{T}_{\text{MS}}$) for $\alpha = 0.05$, the proposed clustering approach with known number of clusters ($\mathcal{T}_{\text{MS, known } N}$), the benchmark procedure with small bandwidth $h_1$ ($\mathcal{T}_{\text{bm, 1}}$), and the benchmark procedure with large bandwidth $h_2$ ($\mathcal{T}_{\text{bm, 2}}$) for different sample sizes $T$.

| | $\mathcal{T}_{\text{MS}}$ | $\mathcal{T}_{\text{MS, known } N}$ | $\mathcal{T}_{\text{bm, 1}}$ | $\mathcal{T}_{\text{bm, 2}}$ |
|---|---|---|---|---|
| $T = 100$ | 0.007 | 0.029 | 0.066 | 0.066 |
| $T = 250$ | 0.742 | 0.903 | 0.943 | 0.066 |
| $T = 500$ | 1.000 | 1.000 | 1.000 | 0.066 |

This procedure serves as a straightforward and intuitive benchmark, with our approach being a refinement of it. In particular, our approach replaces the simple distance measure $d_{ij}$ by a more advanced multiscale distance measure and provides a way to estimate the number of clusters—a feature not included in the benchmark procedure. For the sake of fair comparison, we assume that the number of clusters $N = 3$ is known, and thus, do not provide an estimate of the number of groups $\hat{N}$.

We implement the clustering algorithm as in Section A.2 with one small difference. Computational time of the benchmark procedure exceeds significantly the computational time needed for our method (presumably, due to the lack of the discretization of the grid), hence, in order to lighten the computational burden, instead of 5000 samples, we simulate 1000 samples for this simulation exercise.

The simulation results are reported in Table A.10. As before, the entries in Table A.10 present 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 highlight a significant distinction: the benchmark procedure's outcomes are heavily influenced by the selection of bandwidth, a limitation not observed in our clustering approach. Furthermore, with large sample sizes, the empirical probabilities of success using our clustering approach—even with an estimated number of clusters—are quite high. In contrast, the benchmark procedure lacks a method for estimating the number of clusters. Finally, our procedure, which relies on simple arithmetic averages instead of integrals, is computationally much more efficient,

offering a significant advantage for practitioners.

# B Application to GDP growth data

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 database, Federal Reserve Economics Data (FRED) and the Barro-Lee Educational Attainment dataset (**?**). 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.[1]

- **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.[2]

- **Capital (*K*):** We use data on *Gross Fixed Capital Formation* from the OECD.Stat

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

[2] 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.

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. **?**, **?** and **?**.

- **Labour ($\boldsymbol{L}$):** 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` (**?**) 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.

- **Human capital ($\boldsymbol{H}$):** 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,

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

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 $\alpha_i$ 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 (B.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 $U_T = \big\{u \in [0,1] : u = \frac{8t+1}{2T} \text{ for some } t \in \mathbb{N}\big\}$ and $H_T = \big\{h \in \big[\frac{\log T}{T}, \frac{1}{4}\big] : h = \frac{4t}{T} \text{ for some } t \in \mathbb{N}\big\}$. We thus take into account all locations $u$ on an equidistant grid $U_T$ with step length $4/T$ and all 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 $\mathcal{I}_{u,h}$ with $(u,h) \in \mathcal{G}_T$ 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 $\varepsilon_{it}$ exist).

- To obtain an estimator $\hat{\sigma}_i^2$ of the long-run error variance $\sigma_i^2$ 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 BIC, which yields $p_i = 3$ for Australia, Canada and the UK and $p_i = 1$ for all other countries.[3]

[3]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 $p_i$.

- 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 B.1–B.11. 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 $\mathcal{S}^{[i,j]}_{min}(\alpha) \subseteq \mathcal{S}^{[i,j]}(\alpha)$ 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 B.1–B.11: 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 B.1–B.11. In 9 cases (Figures B.1–B.9), one of the involved countries is Norway. Inspecting the trend estimates in panels (b) of Figures B.1–

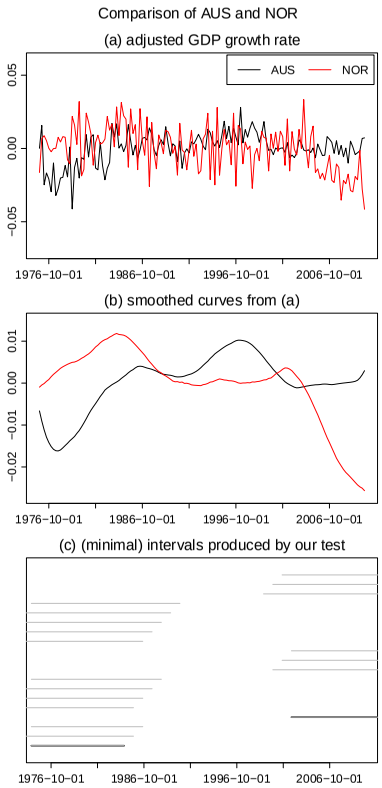


Figure B.1: Test results for the comparison of Australia and Norway.

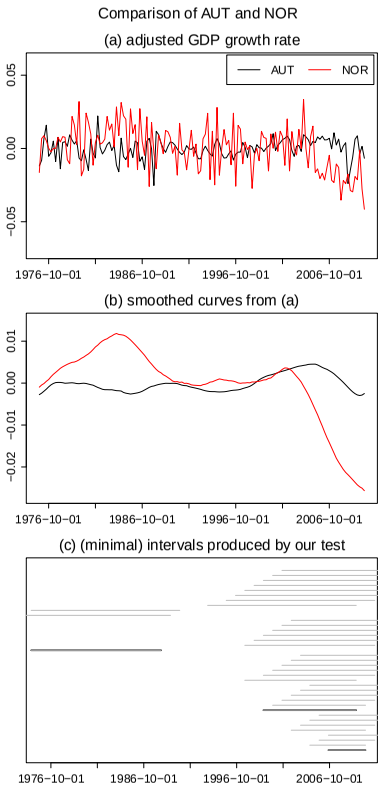


Figure B.2: Test results for the comparison of Austria and Norway.

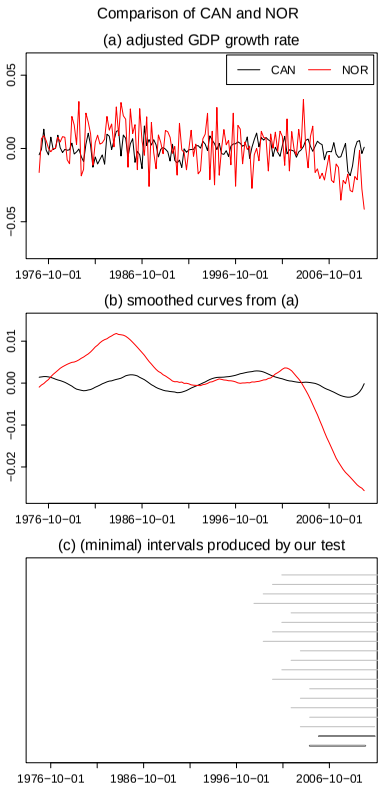


Figure B.3: Test results for the comparison of Canada and Norway.

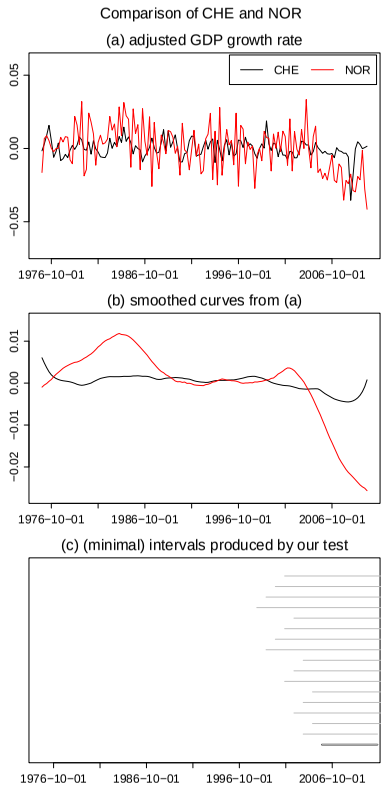


Figure B.4: Test results for the comparison of Switzerland and Norway.

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

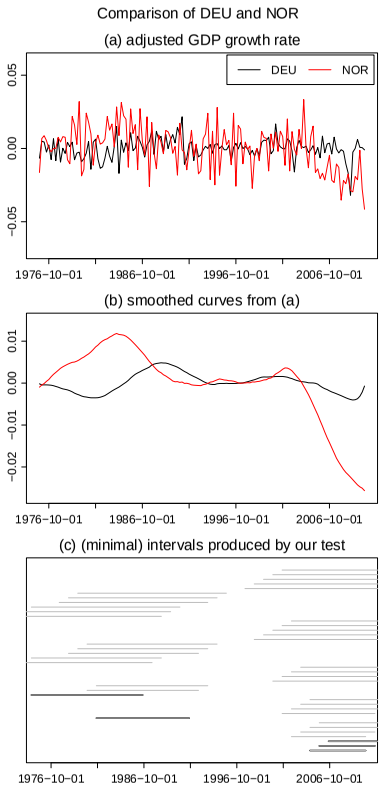


Figure B.5: Test results for the comparison of Germany and Norway.

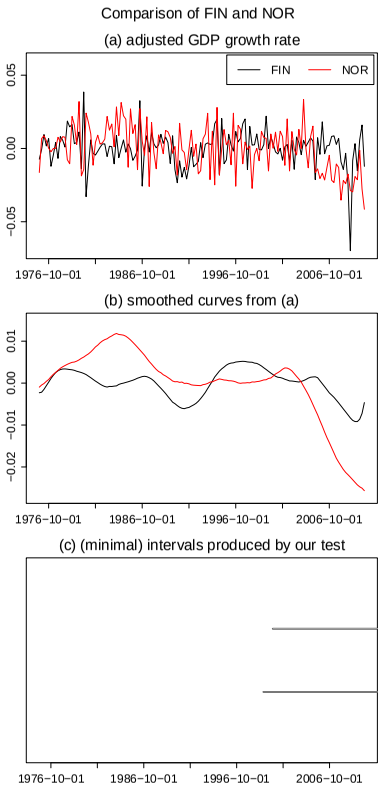


Figure B.6: Test results for the comparison of Finland and Norway.

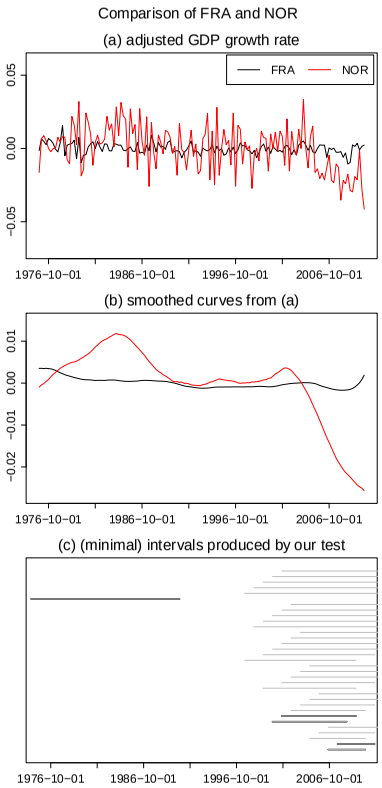


Figure B.7: Test results for the comparison of France and Norway.

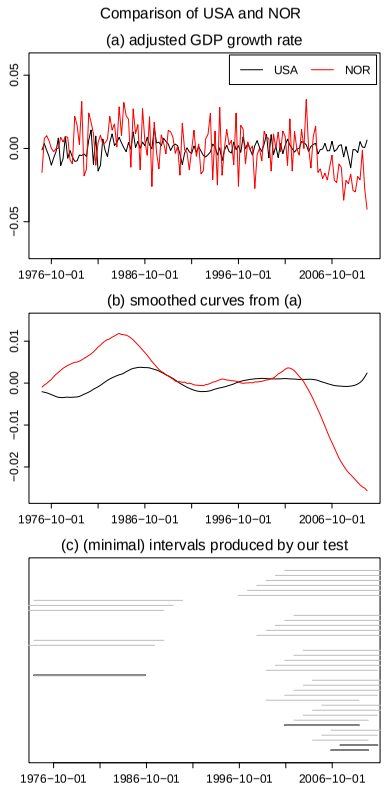


Figure B.8: Test results for the comparison of the USA and Norway.

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.

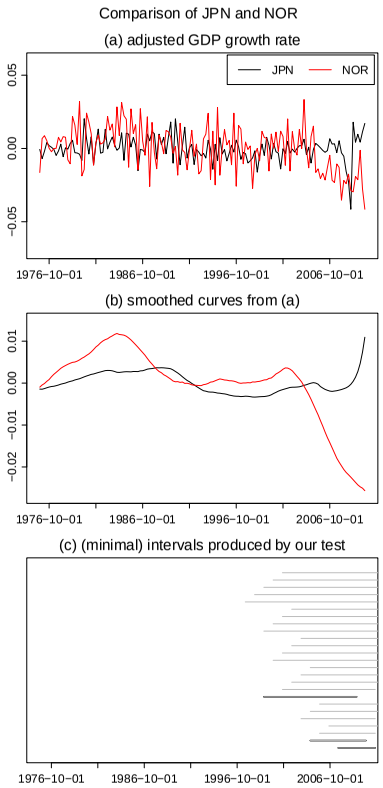


Figure B.9: Test results for the comparison of Japan and Norway.

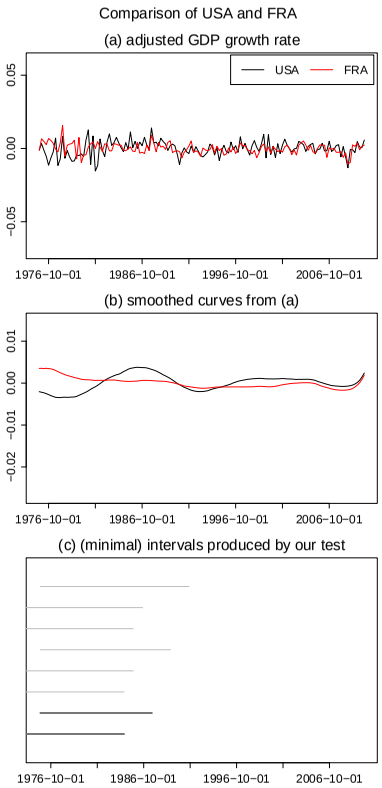


Figure B.10: Test results for the comparison of the USA and France.

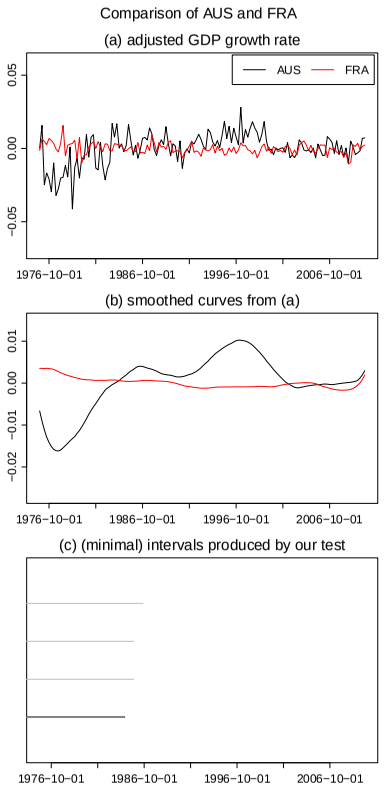


Figure B.11: Test results for the comparison of Australia and France.

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.

B.9, 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.

Figures B.10 and B.11 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 B.11) 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 B.12. The number of clusters is estimated to be $\widehat{N} = 3$. The rectangles in Figure B.12 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 B.13 depicts local linear kernel estimates of the $n = 11$ GDP time trends (calculated from the augmented

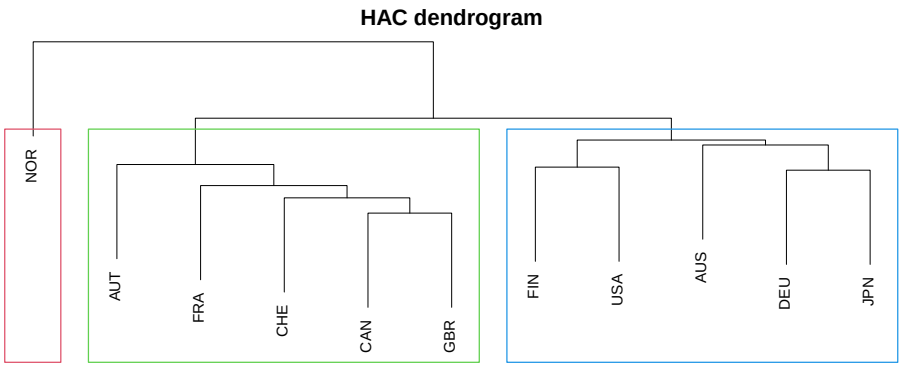


Figure B.12: Dendrogram of the HAC algorithm. Each coloured rectangle corresponds to one of the clusters.

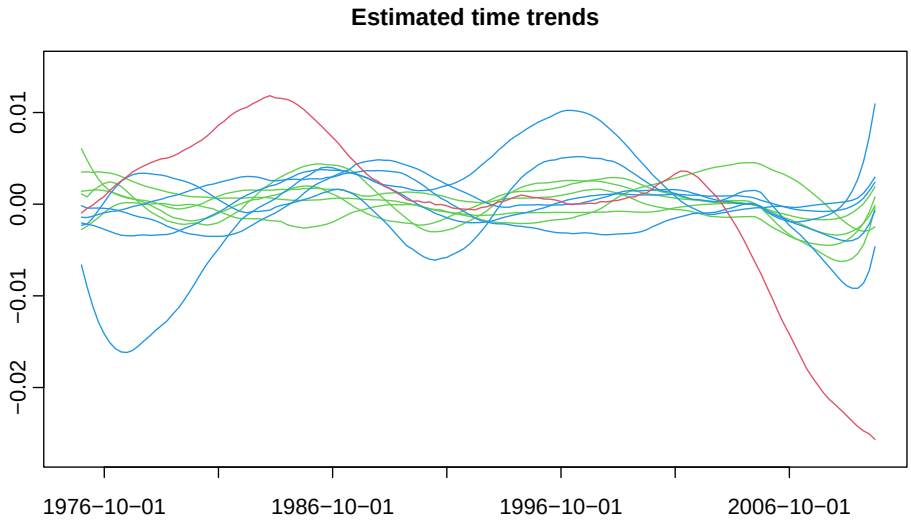


Figure B.13: Local linear estimates of the $n = 11$ time trends (calculated from the augmented time series $\widehat{Y}_{it}$ with bandwidth $h = 0.1$ and Epanechnikov kernel). Each trend estimate is coloured according to the cluster that it is assigned to.

time series $\widehat{Y}_{it}$ with bandwidth 0.1 and Epanechnikov kernel). Their colour indicates which cluster they belong to.

The results in Figures B.12 and B.13 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 B.13, 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 B.12 and B.13) and the other one consisting of the 5 countries Austria, Canada, France, Switzerland and the UK (plotted in green in Figures B.12 and B.13). Visual inspection of the trend estimates in Figure B.13 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.

# C Technical details

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

**Proposition B.2** (Wu and Wu (2016), Theorem 2)**.** *Assume that* $\|Z_\cdot\|_{q,\alpha} < \infty$ *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* $w > 0$*,*

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

*where* $C_1, C_2, C_3$ *are constants that only depend on* $q$ *and* $\alpha$*.*

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

**Lemma B.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 B.3.** Let $\eta > 0$ be a fixed number. We apply Proposition B.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 $\|Z_\cdot\|_{2,\alpha} \leq \|Z_\cdot\|_{q,\alpha} \leq C_Z < \infty$. Hence, for $w$ chosen sufficiently large, we get

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

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

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

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

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

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

where $\mathcal{H}'_{it,j} = (\ldots, \nu_{i(-1),j}, \nu'_{i0,j}, \nu_{i1,j}, \ldots, \nu_{it-1,j}, \nu_{it,j})$, $\mathcal{G}'_{it,j} = (\ldots, \xi_{i(-1),j}, \xi'_{i0,j}, \xi_{i1,j}, \ldots, \xi_{it-1,j}, \xi_{it,j})$ 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 $\eta_{i0}$. From this and Assumptions (C1), (C3), (C4) and (C6), it immediately follows that $\sum_{s=t}^{\infty} \delta_p(a_{ij}, s) = O(t^{-\alpha})$. □

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

**Lemma B.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 B.5.** The estimator $\widehat{\boldsymbol{\beta}}_i$ can be written as

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

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

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

In what follows, we show that

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

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

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

Lemma B.5 follows from applying these three statements together with standard arguments to formula (C.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 B.4, the claim (C.2) follows upon applying Lemma B.3. Another

application of Lemma B.3 yields that

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

As $\mathbb{E}[\Delta \boldsymbol{X}_{it}\Delta \boldsymbol{X}_{it}^{\top}]$ is invertible, we can invoke Slutsky's lemma to obtain (C.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 $t \in \{1,\dots,T\}$ and some constant $C > 0$. Hence,

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

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

**Lemma B.6.** *Let*

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

*be the subseries estimator of $\sigma_i^2$, where $s_T \asymp T^{1/3}$ is the length of the subseries and $M = \lfloor T/s_T \rfloor$ is the largest integer not exceeding $T/s_T$. Under Assumptions (C1)–(C7),*

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

*for each $i$.*

**Proof of Lemma B.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

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

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

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

## 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 $\sigma_i^2$ 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 (3.10). More formally, we prove the following result.

**Proposition B.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)*

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

*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 (3.10).*

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 B.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 $\delta_T$ 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

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

**Proof of (C.5).** It holds that

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

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

Moreover, since

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

by Step 1 and

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

by Step 2, we arrive at (C.5). □

**Step 4**

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

**Proposition B.9.** *It holds that*

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

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

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

**Step 5**

We finally show that

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

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

$$\mathbb{P}\Big(\widehat{\widehat{\Phi}}_{n,T} \leq x - \delta\Big) - \mathbb{P}\Big(\big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| > \delta\Big) \\ \leq \mathbb{P}\big(\widehat{\Phi}_{n,T} \leq x\big) \leq \mathbb{P}\Big(\widehat{\widehat{\Phi}}_{n,T} \leq x + \delta\Big) + \mathbb{P}\Big(\big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| > \delta\Big). \tag{C.7}$$

It holds that

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

and analogously

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

Combining these two inequalities, we arrive at (C.7).

Now let $x \in \mathbb{R}$ be any point such that $\mathbb{P}(\widehat{\Phi}_{n,T} \leq x) \geq \mathbb{P}(\Phi_{n,T} \leq x)$. With the help of (C.7), we get that

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

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

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

Consequently,

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

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

**Details on Steps 1–5**

**Proof of Proposition B.7**. Consider the stationary process $\mathcal{E}_i=\{\varepsilon_{it}:1\le t\le T\}$ for some fixed $i\in\{1,\dots,n\}$. 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 $[\widetilde{\varepsilon}_{i1},\dots,\widetilde{\varepsilon}_{iT}]\overset{\mathcal{D}}{=}[\varepsilon_{i1},\dots,\varepsilon_{iT}]$ for each $T$ and

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

where $\sigma_i^2=\sum_{k\in\mathbb{Z}}\mathrm{Cov}(\varepsilon_{i0},\varepsilon_{ik})$ denotes the long-run error variance. We apply this result separately for each $i\in\{1,\dots,n\}$. 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 $\widehat{\widehat{\sigma}}_i^2$ is constructed

from $\mathcal{E}_i$. Since $[\widetilde{\varepsilon}_{i1}, \ldots, \widetilde{\varepsilon}_{iT}] \stackrel{\mathcal{D}}{=} [\varepsilon_{i1}, \ldots, \varepsilon_{iT}]$ 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

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

and the auxiliary statistic

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

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

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

In what follows, we prove that

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

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

which completes the proof.

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

$$\begin{aligned}\left|\widetilde{\Phi}_{n,T} - \Phi_{n,T}^{\diamond}\right| &\le \max_{1 \le i < j \le n} \left(\widetilde{\sigma}_i^2 + \widetilde{\sigma}_j^2\right)^{-1/2} \max_{1 \le i < j \le n} \max_{(u,h) \in \mathcal{G}_T} \left|\widetilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h)\right| \Big\} \\ &= O_p(1) \cdot \max_{1 \le i < j \le n} \max_{(u,h) \in \mathcal{G}_T} \left|\widetilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h)\right|, \end{aligned} \tag{C.11}$$

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

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

where

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

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

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

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

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

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

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

Moreover, standard arguments show that $\max_{(u,h)\in\mathcal{G}_T} W_T(u,h) = O(1/\sqrt{Th_{\min}})$. Plugging these two results into (C.12), we obtain that

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

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

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

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

$$\leq \max_{1\leq i<j\leq n}\left\{\left|\left(\widetilde{\sigma}_i^2+\widetilde{\sigma}_j^2\right)^{-1/2}-\left(\sigma_i^2+\sigma_j^2\right)^{-1/2}\right|\right\}\max_{1\leq i<j\leq n}\max_{(u,h)\in\mathcal{G}_T}|\phi_{ij,T}(u,h)|$$

$$= o_p(\rho_T)\max_{1\leq i<j\leq n}\max_{(u,h)\in\mathcal{G}_T}|\phi_{ij,T}(u,h)|\,, \tag{C.13}$$

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

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

$$\phi_{ij,T}^{(II)}(u,h)=\sum_{t=1}^{T}w_{t,T}(u,h)\left(\sigma_i \bar{Z}_{i}-\sigma_j \bar{Z}_{j}\right)\sim N\big(0,(\sigma_i^2+\sigma_j^2)c_T(u,h)\big)$$

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

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

where we have used that $n$ is fixed and $|\mathcal{G}_T|=O(T^\theta)$ for some large but fixed constant $\theta$ by Assumption (C10). Plugging this into (C.13) yields $|\Phi_{n,T}^{\diamond}-\Phi_{n,T}|=o_p(\rho_T\sqrt{\log T})$, which completes the proof of (C.10). □

**Proof of Proposition B.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 B.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 $q \in \mathbb{R}$:

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

Consequently,

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

where the last line is by Lemma B.10. This immediately implies Proposition B.8. □

**Proof of Proposition B.9**. Straightforward calculations yield that

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

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

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

and

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

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

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

Hence,

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

where

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

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

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

We examine these three terms separately.

We first prove that

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

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

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

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

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

Since for any constant $c>0$,

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

and $\mathbb{P}(\max_{i,j,(u,h)} |\widetilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h)| > c\sqrt{\log T}/2) = o(1)$ by (C.17), we get that

$$\begin{aligned}
&\mathbb{P}\Big(\max_{i,j,(u,h)} \Big|\widetilde{\phi}_{ij,T}(u,h)\Big| \le c\sqrt{\log T}\Big)\\
&\qquad \ge \mathbb{P}\left(\max_{i,j,(u,h)} |\phi_{ij,T}(u,h)| \le \frac{c\sqrt{\log T}}{2}\right) - o(1).
\end{aligned} \tag{C.19}$$

Moreover, since $\max_{i,j,(u,h)} |\phi_{ij,T}(u,h)| = O_p(\sqrt{\log T})$ as already proven in (C.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 (C.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 (C.18), which in turn yields (C.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

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

Note that

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

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 (C.20), we can normalize the kernel weights as follows:

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

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

$$\begin{aligned}\mathbb{P}\bigg(\Big|\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} \sqrt{D_{T,u,h}}\cdot w_{t,T}(u,h)\boldsymbol{X}_{it}\Big| \geq x\bigg) \\ \leq C_1 \frac{\big(\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} |\sqrt{D_{T,u,h}}\cdot w_{t,T}(u,h)|^{q'}\big)\|\boldsymbol{X}_{i\cdot}\|_{q',\alpha}^{q'}}{x^{q'}} \\ + C_2\exp\left(-\frac{C_3x^2}{D_{T,u,h}\|\boldsymbol{X}_{i\cdot}\|_{2,\alpha}^2}\right)\end{aligned} \tag{C.21}$$

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 B.1 and $\|\boldsymbol{X}_{i\cdot}\|_{q',\alpha}^{q'} < \infty$ by Assumption (C6). From (C.21),

it follows that for any $\delta > 0$,

$$
\begin{aligned}
&\mathbb{P}\left( \max_{(u,h)\in\mathcal{G}_T} \Big| \sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} w_{t,T}(u,h)\boldsymbol{X}_{it} \Big| \geq \delta T^{1/q} \right) \\
&\leq \sum_{(u,h)\in\mathcal{G}_T} \mathbb{P}\left( \Big| \sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} w_{t,T}(u,h)\boldsymbol{X}_{it} \Big| \geq \delta T^{1/q} \right) \\
&= \sum_{(u,h)\in\mathcal{G}_T} \mathbb{P}\left( \Big| \sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} \sqrt{D_{T,u,h}} \cdot w_{t,T}(u,h)\boldsymbol{X}_{it} \Big| \geq \delta \sqrt{D_{T,u,h}} T^{1/q} \right) \\
&\leq \sum_{(u,h)\in\mathcal{G}_T} \left[ C_1 \frac{(\sqrt{D_{T,u,h}})^{q'} \big( \sum |w_{t,T}(u,h)|^{q'} \big) \|\boldsymbol{X}_{i\cdot}\|_{q',\alpha}^{q'}}{\big(\delta\sqrt{D_{T,u,h}} T^{1/q}\big)^{q'}} + C_2 \exp\left( -\frac{C_3 \big(\delta\sqrt{D_{T,u,h}} T^{1/q}\big)^2}{D_{T,u,h}\|\boldsymbol{X}_{i\cdot}\|_{2,\alpha}^2} \right) \right] \\
&= \sum_{(u,h)\in\mathcal{G}_T} \left[ C_1 \frac{\big( \sum |w_{t,T}(u,h)|^{q'} \big) \|\boldsymbol{X}_{i\cdot}\|_{q',\alpha}^{q'}}{\delta^{q'} T^{q'/q}} + C_2 \exp\left( -\frac{C_3\delta^2 T^{2/q}}{\|\boldsymbol{X}_{i\cdot}\|_{2,\alpha}^2} \right) \right] \\
&\leq C_1 \frac{T^\theta \|\boldsymbol{X}_{i\cdot}\|_{q',\alpha}^{q'}}{\delta^{q'} T^{q'/q}} \max_{(u,h)\in\mathcal{G}_T} \left( \sum\nolimits_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} |w_{t,T}(u,h)|^{q'} \right) + C_2 T^\theta \exp\left( -\frac{C_3\delta^2 T^{2/q}}{\|\boldsymbol{X}_{i\cdot}\|_{2,\alpha}^2} \right) \\
&= C\frac{T^{\theta - q'/q}}{\delta^{q'}} + CT^\theta \exp\left(-CT^{2/q}\delta^2\right), \qquad \text{(C.22)}
\end{aligned}
$$

where the constant $C$ depends neither on $T$ nor on $\delta$. 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 (C.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 (C.22) converges to 0 as $T \to \infty$ for

any fixed $\delta > 0$. Consequently, we obtain that

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

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 B.5), we arrive at the bound

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

We finally turn to $C_{n,T}$. Straightforward calculations yield that $|\sum_{t=1}^{T} w_{t,T}(u,h)| \le C\sqrt{Th_{\max}} = o(\sqrt{T})$. Moreover, $\bar{\boldsymbol{X}}_i = O_p(1/\sqrt{T})$ by Lemma B.3 and $\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i = O_p(1/\sqrt{T})$ by Lemma B.5. This immediately yields that

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

To summarize, we have shown that

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

This immediately implies the desired result. □

## Proof of Proposition 4.4

We first show that

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

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

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

Hence,

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

for $\delta > 0$ small enough. This contradicts the definition of the quantile $q_{n,T}(\alpha)$ according to which $q_{n,T}(\alpha) = \inf_{q\in\mathbb{R}}\{\mathbb{P}(\Phi_{n,T} \le q) \ge 1 - \alpha\}$. We thus arrive at (C.24).

Proposition 4.4 is a simple consequence of Theorem 4.1 and equation (C.24). Specifically, we obtain that under $H_0$,

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

## Proof of Proposition 4.5

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

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

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

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

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

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

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

We first derive a lower bound on the term $\widehat{\psi}^B_{ij,T}(u_0, h_0)$. Since the kernel $K$ is symmetric

and $u_0 = t/T$ for some $t$, it holds that $S_{T,1}(u_0, h_0) = 0$ and thus,

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

Together with (C.26), this implies that

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

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 $\ell$,

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

where the constant $C$ does not depend on $u$, $h$ and $T$. With the help of (C.28), we obtain that for any $(u, h) \in \mathcal{G}_T$ with $[u-h, u+h] \subseteq [0, 1]$,

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

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 (C.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 (C.27) allows us to infer that

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

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 B.9 shows that

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

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

that

$$\begin{aligned}
&\max_{1\le i<j\le n}\max_{(u,h)\in\mathcal{G}_T}\left|\widehat{\psi}^A_{ij,T}(u,h)\right| \\
&= \max_{1\le i<j\le n}\max_{(u,h)\in\mathcal{G}_T}\left|\widehat{\phi}_{ij,T}(u,h) + (\bar{m}_{j,T}-\bar{m}_{i,T})\sum_{t=1}^{T} w_{t,T}(u,h)\right| = O_p(\sqrt{\log T}). \qquad \text{(C.31)}
\end{aligned}$$

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

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

Since $q_{n,T}(\alpha) = O(\sqrt{\log T})$ for any fixed $\alpha \in (0,1)$ and $c_T \to \infty$, (C.32) immediately implies that $\mathbb{P}(\widehat{\Psi}_{n,T} \le q_{n,T}(\alpha)) = o(1)$.

## Proof of Proposition 4.6

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

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

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

## Proof of Proposition 5.1

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

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

Now consider the event

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

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

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

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 4.4, it is easy to see that

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

Moreover, the same arguments as for Proposition 4.5 show that

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

Taken together, these two statements imply that

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

In what follows, we show that on the event $B_{n,T}$, (i) $\{\widehat{G}_1^{[n-N]}, \ldots, \widehat{G}_N^{[n-N]}\} = \{G_1, \ldots, G_N\}$ and (ii) $\widehat{N} = N$. From (i), (ii) and (C.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 (C.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 $\ell$, $\widehat{G}_\ell^{[r]} \subseteq G_k$ for some $k$. Because of (C.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 $\{\widehat{G}_1^{[n-N]}, \ldots, \widehat{G}_N^{[n-N]}\} = \{G_1, \ldots, G_N\}$. □

**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 $\ell$. 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 $B_{n,T}$,

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

that is, $\widehat{N} = N$. □

## Proof of Proposition 5.2

We consider the event

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

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

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

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

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

and

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

Moreover, we have that

$$D_{n,T} \subseteq \bigcap_{1\leq \ell<\ell'\leq \widehat{N}} E^{[\ell,\ell']}_{n,T}(\alpha), \tag{C.37}$$

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

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

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

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

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