Multiscale Inference for Nonparametric Time Trends

Marina Khismatullina[1] Michael Vogt[2]

University of Bonn University of Bonn

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

Key words: Multiscale statistics; nonparametric regression; time series errors; shape constraints; strong approximations; anti-concentration bounds.

AMS 2010 subject classifications: 62E20; 62G10; 62G20; 62M10.

1 Introduction

When several time series $\mathcal{Y}_i = \{Y_{it} : 1 \le t \le T\}$ are observed for $1 \le i \le n$, we similarly model each time series $\mathcal{Y}_i$ by the equation

$$Y_{it} = m_i\Big(\frac{t}{T}\Big) + \alpha_i + \varepsilon_{it} \tag{1.1}$$

for $1 \le t \le T$, where m_i is a nonparametric time trend, α_i is a (random or deterministic) intercept and ε_{it} are time series errors with $\mathbb{E}[\varepsilon_{it}] = 0$ for all t.

An important question in many applications is whether the time trends m_i are the same for all i. When some of the trends are different, there may still be groups of time series with the same trend. In this case, it is often of interest to estimate the unknown groups from the data. In addition, when two trends m_i and m_j are not the same, it may also be relevant to know in which time regions they differ from each other. In Section 3, we construct statistical methods to approach these questions. In particular, we develop a test of the hypothesis that all time trends in model (1.1) are the same, that is, $m_1 = m_2 = \ldots = m_n$. Similar as before, our method does not only allow to test

[1]Address: Bonn Graduate School of Economics, University of Bonn, 53113 Bonn, Germany. Email: `marina.k@uni-bonn.de`.

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

whether the null hypothesis is violated. It also allows to detect, with a given statistical confidence, which time trends are different and in which time regions they differ from each other. Based on our test method, we further construct an algorithm which clusters the observed time series into groups with the same trend.
For the application we analyse temperature time series measured at 25 different weather stations located in Great Britain. We in particular apply our procedure from Section 3 to test whether the different time series have the same trend.

2 The model

The model setting is as follows. We observe time series $\mathcal{Y}_i = \{Y_{it} : 1 \le t \le T\}$ of length T for $1 \le i \le n$. Each time series $\mathcal{Y}_i$ satisfies the model equation

$$Y_{it} = \beta' X_{it} + m_i\Big(\frac{t}{T}\Big) + \alpha_i + \varepsilon_{it} \tag{2.1}$$

for $1 \le t \le T$, where β is a $d \times 1$ vector of unknown parameters, X_{it} is a $d \times 1$ vector of individual covariates, m_i is an unknown nonparametric trend function defined on $[0,1]$, α_i is a (deterministic or random) intercept term and $\mathcal{E}_i = \{\varepsilon_{it} : 1 \le t \le T\}$ is a zero-mean stationary error process. For identification, we normalize the functions m_i such that $\int_0^1 m_i(u)du = 0$ for all $1 \le i \le n$. The term α_i can also be regarded as an additional error component. In the econometrics literature, it is commonly called a fixed effect error term. It can be interpreted as capturing unobserved characteristics of the time series $\mathcal{Y}_i$ which remain constant over time. We allow the error terms α_i to be dependent across i in an arbitrary way. Hence, by including them in model equation (2.1), we allow the n time series $\mathcal{Y}_i$ in our sample to be correlated with each other. Whereas the terms α_i may be correlated, the error processes $\mathcal{E}_i$ are assumed to be independent across i. In addition, each process $\mathcal{E}_i$ is supposed to satisfy the conditions **??**–**??**. Finally note that throughout the paper, we restrict attention to the case where the number of time series n in model (3.1) is fixed. Extending our theoretical results to the case where n slowly grows with the sample size T is a possible topic for further research.

3 Testing for equality of time trends

In this section, we adapt the multiscale method developed in Section **??** to test the hypothesis that the trend functions in model (3.1) are all the same. More formally, we test the null hypothesis $H_0 : m_1 = m_2 = \ldots = m_n$ in model (2.1). As we will see, the proposed multiscale method does not only allow to test whether the null hypothesis is violated. It also provides information on where violations occur. More specifically, it allows to identify, with a pre-specified confidence, (i) trend functions which are different from each other and (ii) time intervals where these trend functions differ.

3.1 Construction of the test statistic in the absence of exogenous regressors

As a starting point we will first provide a test for common trends in the absence of exogenous covariates X_{it}. So we begin with a model

$$Y_{it} = m_i\Big(\frac{t}{T}\Big) + \alpha_i + \varepsilon_{it}, \tag{3.1}$$

where ε_{it} are zero-mean error terms and α_i are (random or deterministic) intercepts. Defining $Y_{it}^\circ = Y_{it} - \alpha_i$, this equation can be rewritten as $Y_{it}^\circ = m_i(\frac{t}{T}) + \varepsilon_{it}$, which is a standard nonparametric regression equation. The variables Y_{it}° are not observed, but they can be approximated by $\widehat{Y}_{it} = Y_{it} - \widehat{\alpha}_i$, where $\widehat{\alpha}_i = T^{-1}\sum_{t=1}^T Y_{it}$ is an estimator of the intercept α_i. By construction, $\widehat{\alpha}_i - \alpha_i = T^{-1}\sum_{t=1}^T \varepsilon_{it} + T^{-1}\sum_{t=1}^T m_i(\frac{t}{T}) = O_p(T^{-1/2}) + T^{-1}\sum_{t=1}^T m_i(\frac{t}{T})$. Hence, $\widehat{\alpha}_i$ is a reasonable estimator of α_i if $T^{-1}\sum_{t=1}^T m_i(\frac{t}{T})$ converges to zero as $T \to \infty$. To ensure this, we suppose throughout the section that the functions m_i are Lipschitz continuous, that is, $|m_i(v) - m_i(w)| \le L|v-w|$ for all $v, w \in [0,1]$ and some constant $L < \infty$. Since $\int_0^1 m_i(u)du = 0$ by normalization, this implies that $T^{-1}\sum_{t=1}^T m_i(\frac{t}{T}) = O(T^{-1})$. We further let $\widehat{\sigma}_i^2$ be an estimator of the long-run error variance $\sigma_i^2 = \sum_{\ell=-\infty}^{\infty} \mathrm{Cov}(\varepsilon_{i0}, \varepsilon_{i\ell})$ which is computed from the constructed sample $\{\widehat{Y}_{it} : 1 \le t \le T\}$. We thus regard $\widehat{\sigma}_i^2 = \widehat{\sigma}_i^2(\widehat{Y}_{i1}, \dots, \widehat{Y}_{iT})$ as a function of the variables $\widehat{Y}_{it}$ for $1 \le t \le T$. Throughout the section, we assume that $\widehat{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$ with $\rho_T = o(1/\log T)$. Details on how to construct estimators of σ_i^2 are deferred to Section **??**.

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

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

where the kernel weights $w_{t,T}(u,h)$ are defined as in (**??**). The kernel average $\widehat{\psi}_{ij,T}(u,h)$ can be regarded as measuring the distance between the two trend curves m_i and m_j on the interval $[u-h, u+h]$. Similar as in Section **??**, we aggregate the kernel averages $\widehat{\psi}_{ij,T}(u,h)$ for all $(u,h) \in \mathcal{G}_T$ by the multiscale statistic

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

where $\lambda(h) = \sqrt{2\log\{1/(2h)\}}$ and the set $\mathcal{G}_T$ has been introduced in Section **??**. The statistic $\widehat{\Psi}_{ij,T}$ can be interpreted as a distance measure between the two curves m_i and m_j. We finally define the multiscale statistic for testing the null hypothesis $H_0 : m_1 = m_2 = \ldots = m_n$ as

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

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

3.2 The test procedure

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

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

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

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

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

3.3 Theoretical properties of the test

To start with, we introduce the auxiliary statistic

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

where

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

and $\widehat{\phi}_{ij,T}(u,h) = \sum_{t=1}^T w_{t,T}(u,h) \{ (\varepsilon_{it} - \bar{\varepsilon}_i) - (\varepsilon_{jt} - \bar{\varepsilon}_j) \}$ with $\bar{\varepsilon}_i = \bar{\varepsilon}_{i,T} = T^{-1} \sum_{t=1}^T \varepsilon_{it}$. Our first theoretical result characterizes the asymptotic behaviour of the statistic $\widehat{\Phi}_{n,T}$ and parallels Theorem **??** from Section **??**.

Theorem 3.1. *Suppose that the error processes $\mathcal{E}_i = \{\varepsilon_{it} : 1 \le t \le T\}$ are independent across i and satisfy* **??**–**??** *for each i. Moreover, let* **??**–**??** *be fulfilled and assume that*

$\widehat{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$ *with* $\rho_T = o(1/\log T)$ *for each* i*. Then*

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

Theorem 3.1 is the main stepping stone to derive the theoretical properties of our multiscale test. It can be proven by slightly modifying the arguments for Theorem **??**. The details are provided in the Supplementary Material. The following proposition characterizes the behaviour of our multiscale test under the null hypothesis and under local alternatives.

Proposition 3.2. *Let the conditions of Theorem 3.1 be satisfied.*

(a) Under the null hypothesis $H_0 : m_1 = m_2 = \ldots = m_n$*, it holds that*

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

(b) Let $m_i = m_{i,T}$ *be a Lipschitz continuous function with* $\int_0^1 m_{i,T}(w)dw = 0$ *for any* i*. In particular, suppose that* $|m_{i,T}(v) - m_{i,T}(w)| \le L|v-w|$ *for all* $v, w \in [0,1]$ *and some fixed constant* $L < \infty$ *which does not depend on* T*. Moreover, assume that for some pair of indices* i *and* j*, the functions* $m_{i,T}$ *and* $m_{j,T}$ *have the following property: There exists* $(u,h) \in \mathcal{G}_T$ *with* $[u-h, u+h] \subseteq [0,1]$ *such that* $m_{i,T}(w) - m_{j,T}(w) \ge c_T\sqrt{\log T/(Th)}$ *for all* $w \in [u-h, u+h]$ *or* $m_{j,T}(w) - m_{i,T}(w) \ge c_T\sqrt{\log T/(Th)}$ *for all* $w \in [u-h, u+h]$*, where* $\{c_T\}$ *is any sequence of positive numbers with* $c_T \to \infty$*. Then*

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

Part (a) of Proposition 3.2 is a direct consequence of Theorem 3.1. The proof of part (b) is very similar to that of Proposition **??** and thus omitted.

3.4 Construction of the test statistic in the presence of exogenous regressors

We now extend the model (3.1) to include the exogenous regressors:

$$Y_{it} = \beta' X_{it} + m_i\Big(\frac{t}{T}\Big) + \alpha_i + \varepsilon_{it} \tag{3.2}$$

It is obvious that if β is known, the problem of testing for the common time trend would be reduced to the one discussed before. That is, we would test $H_0 : m_1 = m_2 = \ldots = m_n$ in the model

$$\begin{aligned} Y_{it} - \beta' X_{it} &=: V_{it} \\ &= m_i\Big(\frac{t}{T}\Big) + \alpha_i + \varepsilon_{it} \end{aligned}$$

replacing Y_{it} by V_{it} in the construction of test statistic. However, β is not known, so given an estimator $\widehat{\beta}$ we then consider

$$\widehat{V}_{it} =: Y_{it} - \widehat{\beta}' X_{it} = (\widehat{\beta} - \beta)' X_{it} + m_i\Big(\frac{t}{T}\Big) + \alpha_i + \varepsilon_{it}.$$

4 Clustering

4.1 Clustering of time trends

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

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

where g_ℓ are group-specific trend functions. Hence, the time series which belong to the group G_ℓ all have the same time trend g_ℓ. Throughout the section, we suppose that the group-specific trend functions g_ℓ have the following properties: For each ℓ, $g_\ell = g_{\ell,T}$ is a Lipschitz continuous function with $\int_0^1 g_{\ell,T}(w)dw = 0$. In particular, it holds that $|g_{\ell,T}(v) - g_{\ell,T}(w)| \le L|v - w|$ for all $v, w \in [0, 1]$ and some constant $L < \infty$ that does not depend on T. Moreover, for any $\ell \neq \ell'$, the trends $g_{\ell,T}$ and $g_{\ell',T}$ are assumed to differ in the following sense: There exists $(u, h) \in \mathcal{G}_T$ with $[u - h, u + h] \subseteq [0, 1]$ such that $g_{\ell,T}(w) - g_{\ell',T}(w) \ge c_T\sqrt{\log T/(Th)}$ for all $w \in [u - h, u + h]$ or $g_{\ell',T}(w) - g_{\ell,T}(w) \ge c_T\sqrt{\log T/(Th)}$ for all $w \in [u - h, u + h]$, where $0 < c_T \to \infty$.
In many applications, it is natural to suppose that there is a group structure in the data. In this case, a particular interest lies in estimating the unknown groups from the data at hand. In what follows, we combine our multiscale methods with a clustering algorithm to achieve this. More specifically, we use the multiscale statistics $\widehat{\Psi}_{ij,T}$ as distance measures which are fed into a hierarchical clustering algorithm. To describe the algorithm, we first need to introduce the notion of a dissimilarity measure: Let $S \subseteq \{1, \ldots, n\}$ and $S' \subseteq \{1, \ldots, n\}$ be two sets of time series from our sample. We define a dissimilarity measure between S and S' by setting

$$\widehat{\Delta}(S, S') = \max_{\substack{i \in S, \\ j \in S'}} \widehat{\Psi}_{ij,T}. \tag{4.1}$$

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

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

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

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

and merge them into a new cluster.

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

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

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

where we write $\widehat{\Delta}(S) = \widehat{\Delta}(S, S)$ for short and $q_{n,T}(\alpha)$ is the $(1-\alpha)$-quantile of $\Phi_{n,T}$ defined in Section 3.2.

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

Proposition 4.1. *Let the conditions of Theorem 3.1 be satisfied. Then*

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

and

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

This result allows us to make statistical confidence statements about the estimated clusters $\{\widehat{G}_1, \ldots, \widehat{G}_{\widehat{N}}\}$ and their number $\widehat{N}$. In particular, we can claim with asymptotic confidence $\ge 1-\alpha$ that the estimated group structure is identical to the true group structure. Note that it is possible to let the significance level α depend on the sample size T in Proposition 4.1. In particular, we can allow $\alpha = \alpha_T$ to converge slowly to zero as $T \to \infty$, in which case we obtain that $\mathbb{P}(\{\widehat{G}_1, \ldots, \widehat{G}_{\widehat{N}}\} = \{G_1, \ldots, G_N\}) \to 1$ and $\mathbb{P}(\widehat{N} = N) \to 1$. The proof of Proposition 4.1 can be found in the Supplementary Material.

Our multiscale methods do not only allow us to compute estimators of the unknown groups $G_1, \ldots, G_N$. They also provide information on the locations where two group-specific trend functions g_ℓ and $g_{\ell'}$ differ from each other. To turn this claim into a mathematically precise statement, we need to introduce some notation. First of all, note that the indexing of the estimators $\widehat{G}_1, \ldots, \widehat{G}_{\widehat{N}}$ is completely arbitrary. We could, for example, change the indexing according to the rule $\ell \mapsto \widehat{N} - \ell + 1$. In what follows, we suppose that the estimated groups are indexed such that $P(\widehat{G}_\ell = G_\ell \text{ for all } \ell) \ge (1-\alpha) + o(1)$. Theorem 4.1 implies that this is possible without loss of generality. Keeping this convention in mind, we define the sets

$$\mathcal{A}_{n,T}(\ell, \ell') = \Big\{(u,h) \in \mathcal{G}_T : \Big|\frac{\widehat{\psi}_{ij,T}(u,h)}{(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2)^{1/2}}\Big| > q_{n,T}(\alpha) + \lambda(h) \text{ for some } i \in \widehat{G}_\ell, j \in \widehat{G}_{\ell'}\Big\}$$

and

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

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

$$E_{n,T}(\ell, \ell') = \Big\{\forall I_{u,h} \in \Pi_{n,T}(\ell, \ell') : g_\ell(v) \ne g_{\ell'}(v) \text{ for some } v \in I_{u,h} = [u-h, u+h]\Big\}$$

be the event that the group-specific time trends g_ℓ and $g_{\ell'}$ differ on all intervals $I_{u,h} \in$

$\Pi_{n,T}(\ell, \ell')$. With this notation at hand, we can make the following formal statement whose proof is given in the Supplementary Material.

Proposition 4.2. *Under the conditions of Theorem 3.1, the event*

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

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

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

The statement of Proposition 4.2 remains to hold true when the sets of intervals $\Pi_{n,T}(\ell, \ell')$ are replaced by the corresponding sets of minimal intervals. According to Proposition 4.2, the sets $\Pi_{n,T}(\ell, \ell')$ allow us to locate, with a pre-specified confidence, time regions where the group-specific trend functions g_ℓ and $g_{\ell'}$ differ from each other. In particular, we can claim with asymptotic confidence $\ge 1-\alpha$ that the trend functions g_ℓ and $g_{\ell'}$ differ on all intervals $I_{u,h} \in \Pi_{n,T}(\ell, \ell')$.

5 Simulations

We next turn to the test methods from Section 3. The simulation design extends the setup from above. We generate data from the model $Y_{it} = m_i(\frac{t}{T}) + \varepsilon_{it}$, where the number of time series i is set to $n = 15$ and we consider different time series lengths T. For each i, the errors ε_{it} are drawn from the AR(1) model $\varepsilon_{it} = a\varepsilon_{i,t-1} + \eta_{it}$, where as before $a = 0.267$ and the innovations η_{it} are i.i.d. normally distributed with mean 0 and variance 0.35. To generate data under the null $H_0 : m_1 = \ldots = m_n$, we let $m_i = 0$ for all i without loss of generality. To produce data under the alternative, we define $m_1(u) = \beta\,(u - 0.5)$ with $\beta = 0.75, 1, 1.25$ and set $m_i = 0$ for all $i \neq 1$. Hence, all trend functions are the same except for m_1 which is an increasing linear function. We here use a linear function rather than a broken line as the normalization constraint $\int_0^1 m_1(u)du = 0$ is directly satisfied in this case.
The test is implemented analogously as before. We in particular use an Epanechnikov kernel, we define the grid $\mathcal{G}_T$ as in (**??**) and we set the two tuning parameters L_1 and L_2 to $\lfloor \sqrt{T} \rfloor$ and $\lfloor 2\sqrt{T} \rfloor$, respectively. In order to compute the critical values of the test, we simulate 1000 values of the statistic $\Phi_{n,T}$ defined in Section 3.2 and compute their empirical $(1 - \alpha)$ quantile $q_{n,T}(\alpha)$. Note that the statistic $\Phi_{n,T}$ depends on the estimators $\widehat{\sigma}_i^2$ of the long-run error variances σ_i^2. This implies that for each simulated sample, we have to recompute the empirical quantile $q_{n,T}(\alpha)$ and thus the critical value of the test. This is of course computationally extremely expensive. In order to circumvent this issue, we make the additional assumption that the long-run

Table 0: Size of the multiscale test from Section 3 for $n = 15$ time series, different sample sizes T and nominal sizes α.

	nominal size α		
T	0.01	0.05	0.1
250	0.018	0.049	0.079
350	0.019	0.069	0.120
500	0.020	0.049	0.086
1000	0.011	0.045	0.089

Table 1: Power of the multiscale test from Section 3 for $n = 15$ time series, different sample sizes T and nominal sizes α. Each panel corresponds to a different slope parameter β.

(a) $\beta = 0.75$

	nominal size α		
T	0.01	0.05	0.1
250	0.354	0.557	0.687
350	0.505	0.753	0.850
500	0.859	0.946	0.964
1000	0.997	1.000	1.000

(b) $\beta = 1.00$

	nominal size α		
T	0.01	0.05	0.1
250	0.758	0.895	0.946
350	0.902	0.976	0.986
500	0.997	0.999	0.999
1000	1.000	1.000	1.000

(c) $\beta = 1.25$

	nominal size α		
T	0.01	0.05	0.1
250	0.961	0.990	0.997
350	0.997	1.000	1.000
500	1.000	1.000	1.000
1000	1.000	1.000	1.000

error variance is known to be the same across i, that is, $\sigma_i^2 = \sigma^2$ for all i. Under this assumption, we can estimate σ^2 by $\widehat{\sigma}^2 = (\sum_{i=1}^n \widehat{\sigma}_i^2)/n$, and the Gaussian statistic $\Phi_{n,T}$ simplifies to $\Phi_{n,T} = \max_{1 \le i < j \le n} \Phi_{ij,T}$ with $\Phi_{ij,T} = \max_{(u,h) \in \mathcal{G}_T} \{|\phi_{ij,T}(u,h)| - \lambda(h)\}$ and $\phi_{ij,T}(u,h) = \sum_{t=1}^T w_{t,T}(u,h)\{(Z_{it} - \bar{Z}_i) - (Z_{jt} - \bar{Z}_j)\}$. This statistic does not depend on the estimators $\widehat{\sigma}_i^2$ anymore. We thus do not need to recompute the critical values for each simulated sample, which decreases the running time significantly.

The simulation results are reported in Tables 0 and 1. The entries of the tables are computed in the same way as those in Tables **??** and **??**. Inspecting Table 0, the actual size of the test can be seen to approximate the nominal target α quite well even for small values of T. Moreover, as can be seen from Table 1, the test also has reasonable power against the alternatives considered. For the smallest slope $\beta = 0.75$ and the smallest sample size $T = 250$, the power is only moderate, reflecting the fact that the alternative is not very far away from the null. However, as we increase the slope β and the sample size T, the power quickly increases.

We finally investigate the finite sample performance of the clustering algorithm from Section 4.1. To do so, we partition the $n = 15$ time series into $N = 3$ groups, each containing 5 time series. Specifically, we set $G_1 = \{1, \ldots, 5\}$, $G_2 = \{6, \ldots, 10\}$ and $G_3 = \{11, \ldots, 15\}$. Moreover, we define the group-specific trend functions g_1, g_2 and g_3 by $g_1(u) = 0$, $g_2(u) = 1 \cdot (u - 0.5)$ and $g_3(u) = (-1) \cdot (u - 0.5)$. In order to compute our estimators of the groups G_1, G_2, G_3 and their number $N = 3$, we use the same implementation as before followed by the clustering procedure from Section 4.1. The estimation results are reported in Table 2. The entries in Table 2a are computed as the

Table 2: Clustering results for different sample sizes T and nominal sizes α.

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

	nominal size α		
T	0.01	0.05	0.1
250	0.711	0.911	0.944
350	0.946	0.979	0.966
500	0.990	0.978	0.969
1000	0.998	0.987	0.972

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

	nominal size α		
T	0.01	0.05	0.1
250	0.581	0.747	0.776
350	0.894	0.931	0.921
500	0.984	0.974	0.966
1000	0.998	0.987	0.972

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 2*b* give the empirical probabilities with which the estimated group structure $\{\widehat{G}_1, \ldots, \widehat{G}_{\widehat{N}}\}$ equals the true one $\{G_1, G_2, G_3\}$.

The simulation results nicely illustrate the theoretical properties of our clustering algorithm. According to Proposition 4.1, the probability that $\widehat{N} = N$ and $\{\widehat{G}_1, \ldots, \widehat{G}_{\widehat{N}}\} = \{G_1, G_2, G_3\}$ should be at least $(1-\alpha)$ asymptotically. For the sample sizes $T = 500$ and $T = 1000$, the empirical probabilities reported in Table 2 can indeed be seen to exceed the value $(1 - \alpha)$ as predicted by Proposition 4.1. Only for $T = 500$ and $\alpha = 0.01$, the empirical probability is slightly below $(1-\alpha)$. For the smaller sample sizes $T = 250$ and $T = 350$, in contrast, some of the empirical probabilities are much smaller than $(1-\alpha)$. This reflects the asymptotic nature of Proposition 4.1 and is not very suprising. It simply mirrors the fact that for small sample sizes, the effective noise level in the simulated data is quite high. Even though some of the empirical probabilities for $T = 250$ and $T = 350$ are clearly below the target $(1 - \alpha)$, they are still quite substantial. Hence, even for these small sample sizes, our estimates $\widehat{N}$ and $\{\widehat{G}_1, \ldots, \widehat{G}_{\widehat{N}}\}$ are equal to the true values in a large number of simulations.

6 Applications

6.1 Analysis of UK weather station data

To illustrate our test method from Section 3, we examine a dataset of monthly mean temperatures from 34 different UK weather stations. The data are publicly available on the webpage of the UK Met Office. We use a subset of 25 stations for which data are available over the time span from 1986 to 2017. We thus observe a time series $\mathcal{Y}_i = \{Y_{it} : 1 \le t \le T\}$ of length $T = 386$ for each station $i \in \{1, \ldots, 25\}$. The time

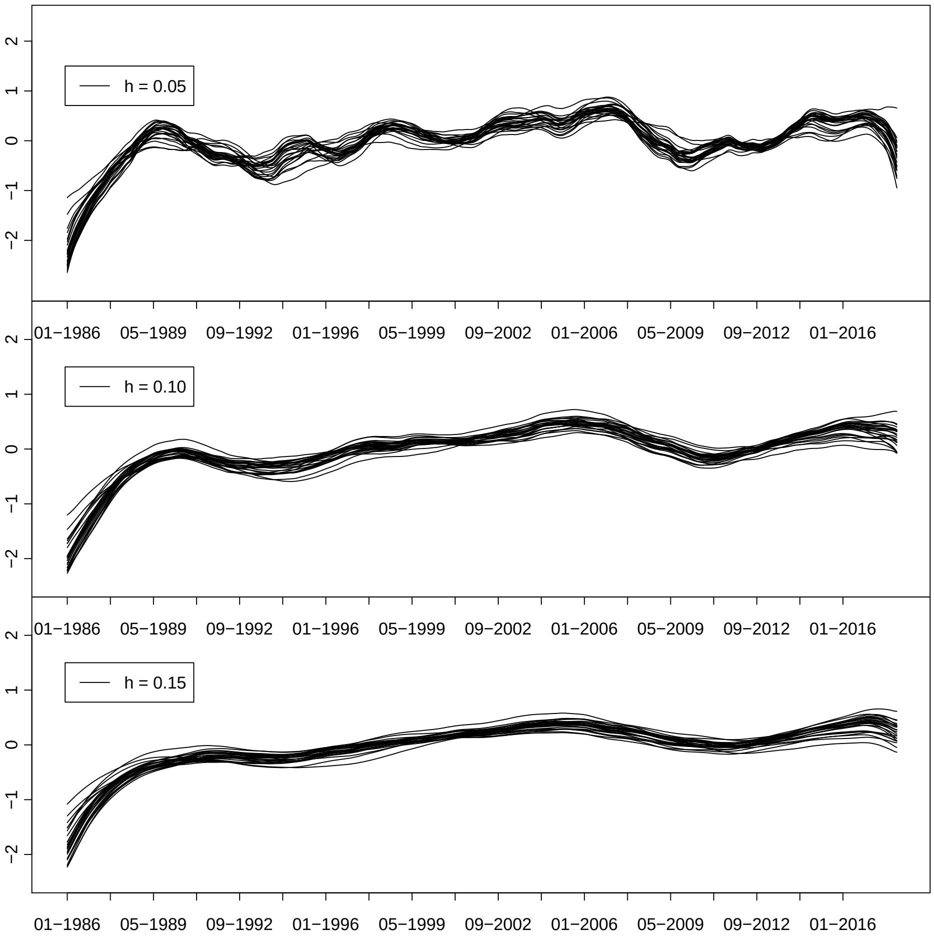

Figure 1: Local linear kernel estimates of the $n = 25$ time trends from the application of Section 6.1. Each panel shows the estimates for a different bandwidth h.

series $\mathcal{Y}_i$ is assumed to follow the model

$$Y_{it} = \alpha_i(t) + m_i\Big(\frac{t}{T}\Big) + \varepsilon_{it}, \tag{6.1}$$

where m_i is an unknown nonparametric time trend and $\alpha_i(t)$ is a month-specific intercept which captures the seasonality pattern in the data. We suppose that $\alpha_i(t) = \alpha_i(t + 12\ell)$ for any integer ℓ, that is, we have a different intercept $\alpha_i(k)$ for each month $k = 1, \ldots, 12$. The test method and the underlying theory from Section 3 can be easily adapted to model (6.1), which is a slight extension of model (**??**). The details are provided below. As in Section **??**, the error process $\mathcal{E}_i = \{\varepsilon_{it} : 1 \le t \le T\}$ is assumed to have the AR(1) structure $\varepsilon_{it} = a_i \varepsilon_{i,t-1} + \eta_{it}$ for each i, where η_{it} are i.i.d. innovations with mean zero.

We aim to test whether the time trend m_i is the same at each of the 25 weather stations. In other words, we want to test the null hypothesis $H_0 : m_1 = \ldots = m_n$ with $n = 25$ in model (6.1). To do so, we apply the multiscale test from Section 3 with two minor modifications: (i) We define $\widehat{Y}_{it} = Y_{it} - \widehat{\alpha}_i(t)$, where $\widehat{\alpha}_i(t)$ is an estimator of $\alpha_i(t)$.

In particular, we set $\widehat{\alpha}_i(t) = \widehat{\alpha}_i(k)$ for any $t = k + 12\ell$ with $1 \le k \le 12$ and some $\ell \in \mathbb{Z}$, where $\widehat{\alpha}_i(k) = T_k^{-1} \sum_{t=1}^T 1_k(t) Y_{it}$ with $1_k(t) = 1(t = k + \lfloor (t-1)/12 \rfloor \cdot 12)$ and $T_k = \sum_{t=1}^T 1_k(t)$ for $1 \le k \le 12$. (ii) We define the Gaussian statistic $\Phi_{n,T}$ as in Section 3.2 with $\phi_{ij,T}(u,h) = \sum_{t=1}^T w_{t,T}(u,h)\{\widehat{\sigma}_i(Z_{it} - \bar{Z}_i(t)) - \widehat{\sigma}_j(Z_{jt} - \bar{Z}_j(t))\}$, where $\bar{Z}_i(t) = \sum_{k=1}^{12} 1_k(t)\{T_k^{-1} \sum_{s=1}^T 1_k(s) Z_{is}\}$. Apart from these two modifications, the multiscale test is constructed exactly as described in Section 3. We implement the test in the same way as in the simulations of Section 5.

We are now ready to apply the test procedure to the data. Figure 1 depicts local linear estimates of the trend functions m_i for the $n = 25$ different stations. Each panel corresponds to a different bandwidth h. As can be seen, for a given bandwidth h, the fits look very similar to each other. Visual inspection thus suggests that there are no strong differences between the time trends m_i. Our test confirms this impression. It does not reject the null hypothesis at the most common levels $\alpha = 0.01, 0.05, 0.1$. Hence, the test does not provide any evidence for a violation of the null.

References

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