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. 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 an application 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

The analysis of time trends is an important aspect of many time series applications. In this paper, we develop new methods to analyse nonparametric time trends. We consider two different model settings, depending on whether a single or multiple time series are observed. When the observations come from a single time series $\{Y_t : 1 \le t \le T\}$, we consider the model

$$Y_t = m\Big(\frac{t}{T}\Big) + \varepsilon_t \tag{1.1}$$

for $1 \le t \le T$, where $m : [0,1] \to \mathbb{R}$ is an unknown nonparametric trend function and the error terms ε_t form a time series process with $\mathbb{E}[\varepsilon_t] = 0$ for all t. As usual in nonparametric regression, we let the trend functions in (1.1) and (**??**) depend on rescaled time t/T. A detailed description of models (1.1) and (**??**) is provided in Section 2.

Let us first have a closer look at the situation where a single time series is observed. In this case, practitioners are interested in questions such as the following: Does the observed time series have a trend at all? If so, which are the time regions where there is a strong trend? Is the trend decreasing or increasing in these regions? As an example, consider the time series plotted in Figure 1 which shows the yearly mean temperature

[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`.

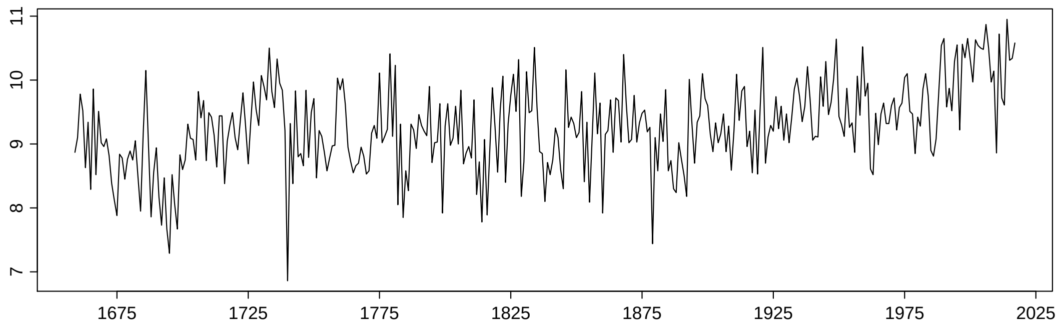

Figure 1: Yearly mean temperature in Central England from 1659 to 2017 measured in °C.

in Central England from 1659 to 2017. Climatologists are very much interested in analysing the trending behaviour of temperature time series like this; see e.g. Benner (1999) and Rahmstorf et al. (2017). Among other things, they would like to know whether there is an upward trend in the Central England mean temperature towards the end of the sample as visual inspection might suggest. In Section 4, we develop a statistical procedure to approach questions like this. Specifically, we construct a method to test the null hypothesis that there is no time trend in the data. Importantly, the proposed method does not only allow to test whether the null hypothesis of no trend is violated. It also allows to identify, with a pre-specified statistical confidence, time regions where there is some upward or downward movement in the trend. As regards the temperature time series in Figure 1, we can for example claim, with a statistical confidence of approximately 95%, that there is some upward movement of the trend in the time period from about 1880 onwards. This is one of the results obtained from a detailed analysis of the time series as conducted in Section 7.

We develop our methods and theory step by step in Sections 3–**??**. In Section 3, we introduce our methods in the context of a simple baseline case. We in particular discuss the problem of testing the simple hypothesis $H_0 : m = 0$ in model (1.1). In Sections 4 and **??**, we adapt the methods and the theory to the test problems we are actually interested in. To construct our methods, we build on ideas from statistical multiscale testing as developed in Chaudhuri and Marron (1999, 2000), Hall and Heckman (2000) and Dümbgen and Spokoiny (2001) among others. Our test procedure for the simple hypothesis $H_0 : m = 0$ in model (1.1) can be outlined as follows: In a first step, we set up statistics $\widehat{s}_T(u, h)$ to test the hypothesis $H_0(u, h)$ that $m = 0$ on the interval $[u-h, u+h]$. In a second step, we aggregate these statistics $\widehat{s}_T(u, h)$ for a wide range of intervals $[u-h, u+h]$. We thus construct a multiscale statistic which allows to test the hypothesis $H_0(u, h)$ simultaneously for many intervals $[u-h, u+h]$. A simple approach to aggregate the statistics $\widehat{s}_T(u, h)$ is to take their supremum $\sup_{(u,h)\in\mathcal{G}_T} |\widehat{s}_T(u, h)|$, where $\mathcal{G}_T$ denotes the set of points (u, h) which are taken into account. As shown in the seminal work of Dümbgen and Spokoiny (2001), this approach is suboptimal in some sense. Following their lead, we define our multiscale statistic by $\widehat{\Psi}_T = \sup_{(u,h)\in\mathcal{G}_T}\{|\widehat{s}_T(u, h)| - \lambda(h)\}$, where $\lambda(h)$ are additive correction terms. The idea behind this additively corrected supremum statistic is discussed in detail in Section 3.

In recent years, multiscale approaches have been developed for a variety of test problems. The general idea of all these approaches is to simultaneously consider a family of test statistics for a wide range of locations u and scales or bandwidths h. This idea has been put to work in different ways, thus resulting in different multiscale test approaches. In the regression context, Chaudhuri and Marron (1999, 2000) have developed the so-called SiZer method. Hall and Heckman (2000) have constructed a multiscale test on monotonicity of a regression function. As already discussed above, Dümbgen and Spokoiny (2001) have developed a multiscale approach which works with additively corrected supremum statistics. They derive theoretical results in the context of a continuous Gaussian white noise model. Their theory can be extended in a fairly straightforward way to a nonparametric regression model $Y_t = m(t/T) + \varepsilon_t$ with i.i.d. Subgaussian errors ε_t. However, it is far from trivial to extend their theory to our setting, that is, to a nonparametric regression model $Y_t = m(t/T) + \varepsilon_t$ with a general weakly dependent error process $\{\varepsilon_t\}$. To derive our theoretical results, we come up with a proof strategy which is quite different from that of Dümbgen and Spokoiny (2001). This strategy is of interest in itself and may be applied to other multiscale test problems. It combines strong approximation results for dependent processes as developed in Berkes et al. (2014) with anti-concentration bounds for Gaussian random vectors as derived in Chernozhukov et al. (2015). The details are described in Section 3.

Whereas a number of multiscale tests for independent data have been developed in recent years, multiscale tests for dependent data are much rarer. Most notably, there are some extensions of the SiZer approach to a time series context. Park et al. (2004) and Rondonotti et al. (2007) have introduced dependent SiZer methods which can be regarded as an alternative to our multiscale test developed in Section 4. However, whereas their analysis is mainly methodological, we back up our multiscale test by a complete asymptotic theory which characterizes its size and power properties. Park et al. (2009) extend the SiZer approach to the problem of comparing multiple time trends. They provide some theory for the special case that the number of observed time series is $n = 2$. In contrast to this, the theory for our multiscale method developed in Section **??** is valid for any given number of time series n.

Besides the SiZer methods just discussed, there are several non-multiscale approaches to analyse nonparametric time trends. Tests on the presence and parametric form of a time trend have been developed in Dette (1999) and Zhang and Wu (2011) among others. Tests to compare nonparametric time trends have been constructed for example in Degras et al. (2012) and Chen and Wu (2018), whereas Vogelsang and Franses (2005) and Lyubchich and Gel (2016) consider tests for comparing parametric time trends. One of the advantages of our multiscale approach is that it is very informative: It does not only allow to test whether the null hypothesis is violated. It also gives information on where violations occur. It thus provides valuable additional information for practitioners.

We complement the theoretical analysis of the paper by a simulation study and two application examples in Sections 6 and 7. The simulation study investigates the finite sample properties of the test methods and the clustering algorithm from Sections 4 and **??**. In the first application example, we examine the temperature time series from Figure 1 with the help of the methods developed in Section 4.

2 The model

We now describe the two model settings in detail which were briefly outlined in the Introduction. The model for the test problems considered in Sections 3 and 4 is as follows: We observe a single time series $\{Y_t : 1 \leq t \leq T\}$ of length T which satisfies the model equation

$$Y_t = m\Big(\frac{t}{T}\Big) + \varepsilon_t \tag{2.1}$$

for $1 \leq t \leq T$. Here, m is an unknown nonparametric trend function defined on $[0, 1]$ and $\{\varepsilon_t : 1 \leq t \leq T\}$ is a zero-mean stationary error process. For simplicity, we restrict attention to equidistant design points $x_t = t/T$. However, our methods and theory can also be carried over to non-equidistant designs. The stationary error process $\{\varepsilon_t\}$ is assumed to have the following properties:

(C1) The variables ε_t allow for the representation $\varepsilon_t = G(\ldots, \eta_{t-1}, \eta_t)$, where η_t are i.i.d. random variables and G is a measurable function.

(C2) It holds that $\|\varepsilon_t\|_q < \infty$ for some $q > 4$, where $\|\varepsilon_t\|_q = (\mathbb{E}|\varepsilon_t|^q)^{1/q}$.

Following Wu (2005), we impose conditions on the dependence structure of the error process $\{\varepsilon_t\}$ in terms of the physical dependence measure $d_{t,q} = \|\varepsilon_t - \varepsilon_t'\|_q$, where $\varepsilon_t' = G(\ldots, \eta_{-1}, \eta_0', \eta_1, \ldots, \eta_{t-1}, \eta_t, \eta_{t+1}, \ldots)$ with $\{\eta_t'\}$ being an i.i.d. copy of $\{\eta_t\}$. In particular, we assume the following:

(C3) Define $\Theta_{t,q} = \sum_{|s| \geq t} d_{s,q}$ for $t \geq 0$. It holds that

$$\Theta_{t,q} = O\big(t^{-\tau_q}(\log t)^{-A}\big),$$

where $A > \frac{2}{3}(1/q + 1 + \tau_q)$ and $\tau_q = \{q^2 - 4 + (q-2)\sqrt{q^2 + 20q + 4}\}/8q$.

The conditions (C1)–(C3) are fulfilled by a wide range of stationary processes $\{\varepsilon_t\}$. As a first example, consider linear processes of the form $\varepsilon_t = \sum_{i=0}^{\infty} c_i \eta_{t-i}$ with $\|\varepsilon_t\|_q < \infty$, where c_i are absolutely summable coefficients and η_t are i.i.d. innovations with $\mathbb{E}[\eta_t] = 0$ and $\|\eta_t\|_q < \infty$. Trivially, (C1) and (C2) are fulfilled in this case. Moreover, if $|c_i| = O(\rho^i)$ for some $\rho \in (0, 1)$, then (C3) is easily seen to be satisfied as well. As a special case, consider an ARMA process $\{\varepsilon_t\}$ of the form $\varepsilon_t + \sum_{i=1}^p a_i \varepsilon_{t-i} = \eta_t + \sum_{j=1}^r b_j \eta_{t-j}$ with $\|\varepsilon_t\|_q < \infty$, where $a_1, \ldots, a_p$ and $b_1, \ldots, b_r$ are real-valued parameters. As before, we let η_t be i.i.d. innovations with $\mathbb{E}[\eta_t] = 0$ and $\|\eta_t\|_q < \infty$. Moreover, as usual, we suppose that the complex polynomials $A(z) = 1 + \sum_{j=1}^p a_j z^j$ and $B(z) = 1 + \sum_{j=1}^r b_j z^j$ do not have any roots in common. If $A(z)$ does not have any roots inside the unit disc, then the ARMA process $\{\varepsilon_t\}$ is stationary and causal. Specifically, it has the representation $\varepsilon_t = \sum_{i=0}^\infty c_i \eta_{t-i}$ with $|c_i| = O(\rho^i)$ for some $\rho \in (0,1)$, implying that (C1)–(C3) are fulfilled. The results in Wu and Shao (2004) show that condition (C3) (as well as the other two conditions) is not only fulfilled for linear time series processes but also for a variety of non-linear processes.

3 The multiscale method

In this section, we introduce our multiscale test method and the underlying theory for the simple hypothesis $H_0 : m = 0$ in model (2.1). As we will see in Sections 4 and **??**, both the method and the theory for this simple case can be easily adapted to more interesting test problems, in particular to the test problems discussed in the Introduction.

3.1 Construction of the test statistic

To construct a multiscale test statistic for the hypothesis $H_0 : m = 0$ in model (2.1), we consider the kernel averages

$$\widehat{\psi}_T(u,h) = \sum_{t=1}^T w_{t,T}(u,h) Y_t,$$

where $w_{t,T}(u,h)$ is a kernel weight with $u \in [0,1]$ and the bandwidth parameter h. In order to avoid boundary issues, we work with a local linear weighting scheme. We in particular set

$$w_{t,T}(u,h) = \frac{\Lambda_{t,T}(u,h)}{\{\sum_{t=1}^T \Lambda_{t,T}^2(u,h)\}^{1/2}}, \tag{3.1}$$

where

$$\Lambda_{t,T}(u,h) = K\Big(\frac{\frac{t}{T}-u}{h}\Big)\Big[S_{T,2}(u,h) - S_{T,1}(u,h)\Big(\frac{\frac{t}{T}-u}{h}\Big)\Big],$$

$S_{T,\ell}(u,h) = (Th)^{-1}\sum_{t=1}^T K(\frac{\frac{t}{T}-u}{h})(\frac{\frac{t}{T}-u}{h})^\ell$ for $\ell = 0,1,2$ and K is a kernel function with the following properties:

(C4) The kernel K is non-negative, symmetric about zero and integrates to one. Moreover, it has compact support $[-1,1]$ and is Lipschitz continuous, that is, $|K(v) - K(w)| \le C|v-w|$ for any $v, w \in \mathbb{R}$ and some constant $C > 0$.

Alternatively to the local linear weights defined in (3.1), we could also work with local constant weights which are defined analogously with $\Lambda_{t,T}(u,h) = K(\frac{\frac{t}{T}-u}{h})$. We however prefer to use local linear weights as these have superior theoretical properties at the boundary.

The kernel average $\widehat{\psi}_T(u,h)$ is a local average of the observations $Y_1,\ldots,Y_T$ which gives positive weight only to data points Y_t with $t/T \in [u-h, u+h]$. Hence, only observations Y_t with t/T close to the location u are taken into account, the amount of localization being determined by the bandwidth h. With the weights defined in (3.1), the kernel average $\widehat{\psi}_T(u,h)$ is nothing else than a rescaled local linear estimator of $m(u)$ with bandwidth h. The weights are chosen such that in the case of independent error terms ε_t, $\text{Var}(\widehat{\psi}_T(u,h)) = \sigma^2$ for any location u and bandwidth h, where $\sigma^2 = \text{Var}(\varepsilon_t)$. In the more general case that the error terms satisfy the weak dependence conditions from Section 2, it holds that $\text{Var}(\widehat{\psi}_T(u,h)) = \sigma^2 + o(1)$ for any location u and any bandwidth h with $h \to 0$ and $Th \to \infty$, where $\sigma^2 = \sum_{\ell=-\infty}^{\infty} \text{Cov}(\varepsilon_0, \varepsilon_\ell)$ is the long-run variance of the error terms. Hence, the statistics $\widehat{\psi}_T(u,h)$ have approximately the same variance across u and h for sufficiently large sample sizes T. In what follows, we consider normalized versions $\widehat{\psi}_T(u,h)/\widehat{\sigma}$ of the kernel averages $\widehat{\psi}_T(u,h)$, where $\widehat{\sigma}^2$ is an estimator of the long-run error variance σ^2. The problem of estimating σ^2 is discussed in detail in Section 5. There, we construct estimators $\widehat{\sigma}^2$ with the property that $\widehat{\sigma}^2 = \sigma^2 + O_p(1/\sqrt{T})$ under appropriate conditions. For the time being, we suppose that $\widehat{\sigma}^2$ is an estimator with reasonable theoretical properties. We in particular assume that $\widehat{\sigma}^2 = \sigma^2 + o_p(\rho_T)$ with $\rho_T = o(1/\log T)$. The convergence rate ρ_T is thus allowed to be much slower than $1/\sqrt{T}$.

Our multiscale statistic combines the kernel averages $\widehat{\psi}_T(u,h)$ for a wide range of different locations u and bandwidths or scales h. Specifically, it is defined as

$$\widehat{\Psi}_T = \max_{(u,h)\in\mathcal{G}_T} \Big\{ \Big| \frac{\widehat{\psi}_T(u,h)}{\widehat{\sigma}} \Big| - \lambda(h) \Big\}, \tag{3.2}$$

where $\lambda(h) = \sqrt{2\log\{1/(2h)\}}$ and $\mathcal{G}_T$ is the set of points (u,h) that are taken into consideration. The details on the set $\mathcal{G}_T$ are discussed below. As can be seen, the statistic $\widehat{\Psi}_T$ does not simply aggregate the individual statistics $\widehat{\psi}_T(u,h)/\widehat{\sigma}$ by taking the supremum over all points $(u,h) \in \mathcal{G}_T$ as in more traditional multiscale approaches. We rather follow the approach pioneered by Dümbgen and Spokoiny (2001) and subtract the additive correction term $\lambda(h)$ from the statistics $\widehat{\psi}_T(u,h)/\widehat{\sigma}$ that correspond to the bandwidth level h. To see the heuristic idea behind the additive correction $\lambda(h)$, consider for a moment the uncorrected statistic

$$\widehat{\Psi}_{T,\text{uncorrected}} = \max_{(u,h)\in\mathcal{G}_T} \Big| \frac{\widehat{\psi}_T(u,h)}{\widehat{\sigma}} \Big|$$

and suppose that the null hypothesis $H_0 : m = 0$ holds true. For simplicity, assume that the errors ε_t are i.i.d. normally distributed and neglect the estimation error in $\widehat{\sigma}$, that is, set $\widehat{\sigma} = \sigma$. Moreover, suppose that the set $\mathcal{G}_T$ only consists of the points $(u_k, h_\ell) = ((2k-1)h_\ell, h_\ell)$ with $k = 1, \ldots, \lfloor 1/2h_\ell \rfloor$ and $\ell = 1, \ldots, L$. In this case, we can write

$$\widehat{\Psi}_{T,\text{uncorrected}} = \max_{1 \le \ell \le L} \max_{1 \le k \le \lfloor 1/2h_\ell \rfloor} \left| \frac{\widehat{\psi}_T(u_k, h_\ell)}{\sigma} \right|.$$

Under our simplifying assumptions, the statistics $\widehat{\psi}_T(u_k, h_\ell)/\sigma$ with $k = 1, \ldots, \lfloor 1/2h_\ell \rfloor$ are independent and standard normal for any given bandwidth h_ℓ. Since the maximum over $\lfloor 1/2h \rfloor$ independent standard normal random variables is $\lambda(h) + o_p(1)$ as $h \to 0$, we obtain that $\max_k \widehat{\psi}_T(u_k, h_\ell)/\sigma$ is approximately of size $\lambda(h_\ell)$ for small bandwidths h_ℓ. As $\lambda(h) \to \infty$ for $h \to 0$, this implies that $\max_k \widehat{\psi}_T(u_k, h_\ell)/\sigma$ tends to be much larger in size for small than for large bandwidths h_ℓ. As a result, the stochastic behaviour of the uncorrected statistic $\widehat{\Psi}_{T,\text{uncorrected}}$ tends to be dominated by the statistics $\widehat{\psi}_T(u_k, h_\ell)$ corresponding to small bandwidths h_ℓ. The additively corrected statistic $\widehat{\Psi}_T$, in contrast, puts the statistics $\widehat{\psi}_T(u_k, h_\ell)$ corresponding to different bandwidths h_ℓ on a more equal footing, thus counteracting the dominance of small bandwidth values.

The multiscale statistic $\widehat{\Psi}_T$ simultaneously takes into account all locations u and bandwidths h with $(u, h) \in \mathcal{G}_T$. Throughout the paper, we suppose that $\mathcal{G}_T$ is some subset of $\mathcal{G}_T^{\text{full}} = \{(u, h) : u = t/T \text{ for some } 1 \le t \le T \text{ and } h \in [h_{\min}, h_{\max}]\}$, where $h_{\min}$ and $h_{\max}$ denote some minimal and maximal bandwidth value, respectively. For our theory to work, we require the following conditions to hold:

(C5) $|\mathcal{G}_T| = O(T^\theta)$ for some arbitrarily large but fixed constant $\theta > 0$, where $|\mathcal{G}_T|$ denotes the cardinality of $\mathcal{G}_T$.

(C6) $h_{\min} \gg T^{-(1-\frac{2}{q})} \log T$, that is, $h_{\min}/\{T^{-(1-\frac{2}{q})} \log T\} \to \infty$ with $q > 4$ defined in (C2) and $h_{\max} = o(1)$.

According to (C5), the number of points (u, h) in $\mathcal{G}_T$ should not grow faster than T^θ for some arbitrarily large but fixed $\theta > 0$. This is a fairly weak restriction as it allows the set $\mathcal{G}_T$ to be extremely large as compared to the sample size T. For example, we may work with the set

$$\begin{aligned} \mathcal{G}_T = \big\{ (u, h) : u = t/T \text{ for some } 1 \le t \le T \text{ and } h \in [h_{\min}, h_{\max}] \\ \text{with } h = t/T \text{ for some } 1 \le t \le T \big\}, \end{aligned}$$

which contains more than enough points (u, h) for most practical applications. Condition (C6) imposes some restrictions on the minimal and maximal bandwidths $h_{\min}$ and $h_{\max}$. These conditions are fairly weak, allowing us to choose the bandwidth window $[h_{\min}, h_{\max}]$ extremely large. In particular, we can choose the minimal bandwidth $h_{\min}$

to be of the order $T^{-1/2}$ for any $q > 4$, which means that we can let $h_{\min}$ converge to 0 very quickly. Moreover, the maximal bandwidth $h_{\max}$ is allowed to converge to 0 arbitrarily slowly, which implies that we can pick it very large.

3.2 The test procedure

In order to formulate a test for the hypothesis $H_0 : m = 0$, we still need to specify a critical value. To do so, we define the statistic

$$\Phi_T = \max_{(u,h)\in\mathcal{G}_T} \Big\{ \Big| \frac{\phi_T(u,h)}{\sigma} \Big| - \lambda(h) \Big\}, \tag{3.3}$$

where $\phi_T(u,h) = \sum_{t=1}^T w_{t,T}(u,h)\,\sigma Z_t$ and Z_t are independent standard normal random variables. The statistic Φ_T can be regarded as a Gaussian version of the test statistic $\widehat{\Psi}_T$ under the null hypothesis H_0. Let $q_T(\alpha)$ be the $(1-\alpha)$-quantile of Φ_T. Importantly, the quantile $q_T(\alpha)$ can be computed by Monte Carlo simulations and can thus be regarded as known. Our multiscale test of the hypothesis $H_0 : m = 0$ is now defined as follows: For a given significance level $\alpha \in (0,1)$, we reject H_0 if $\widehat{\Psi}_T > q_T(\alpha)$.

3.3 Theoretical properties of the test

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

$$\widehat{\Phi}_T = \max_{(u,h)\in\mathcal{G}_T} \Big\{ \Big| \frac{\widehat{\phi}_T(u,h)}{\widehat{\sigma}} \Big| - \lambda(h) \Big\} \tag{3.4}$$

with $\widehat{\phi}_T(u,h) = \widehat{\psi}_T(u,h) - \mathbb{E}[\widehat{\psi}_T(u,h)] = \sum_{t=1}^T w_{t,T}(u,h)\varepsilon_t$. According to the following theorem, the (known) quantile $q_T(\alpha)$ of Φ_T defined in Section 3.2 can be used as a proxy for the $(1-\alpha)$-quantile of the statistic $\widehat{\Phi}_T$.

Theorem 3.1. *Let (C1)–(C6) be fulfilled and assume that $\widehat{\sigma}^2 = \sigma^2 + o_p(\rho_T)$ with $\rho_T = o(1/\log T)$. Then*

$$\mathbb{P}\big(\widehat{\Phi}_T \le q_T(\alpha)\big) = (1-\alpha) + o(1).$$

A full proof of Theorem 3.1 is given in the Appendix. We here shortly outline the proof strategy, which splits up into two main steps. In the first, we replace the statistic $\widehat{\Phi}_T$ for each $T \ge 1$ by a statistic $\widetilde{\Phi}_T$ with the same distribution as $\widehat{\Phi}_T$ and the property that

$$\big|\widetilde{\Phi}_T - \Phi_T\big| = o_p(\delta_T), \tag{3.5}$$

where $\delta_T = o(1)$ and the Gaussian statistic Φ_T is defined in Section 3.2. We thus replace the statistic $\widehat{\Phi}_T$ by an identically distributed version which is close to a Gaussian statistic whose distribution is known. To do so, we make use of strong approximation theory for dependent processes as derived in Berkes et al. (2014). In the second step, we show that

$$\sup_{x \in \mathbb{R}} \big| \mathbb{P}(\widetilde{\Phi}_T \leq x) - \mathbb{P}(\Phi_T \leq x) \big| = o(1), \tag{3.6}$$

which immediately implies the statement of Theorem 3.1. Importantly, the convergence result (3.5) is not sufficient for establishing (3.6). Put differently, the fact that $\widetilde{\Phi}_T$ can be approximated by Φ_T in the sense that $\widetilde{\Phi}_T - \Phi_T = o_p(\delta_T)$ does not imply that the distribution of $\widetilde{\Phi}_T$ is close to that of Φ_T in the sense of (3.6). For (3.6) to hold, we additionally require the distribution of Φ_T to have some sort of continuity property. Specifically, we prove that

$$\sup_{x \in \mathbb{R}} \mathbb{P}\big(|\Phi_T - x| \leq \delta_T \big) = o(1), \tag{3.7}$$

which says that Φ_T does not concentrate too strongly in small regions of the form $[x - \delta_T, x + \delta_T]$. The main tool for verifying (3.7) are anti-concentration results for Gaussian random vectors as derived in Chernozhukov et al. (2015). The claim (3.6) can be proven by combining (3.5) and (3.7), which in turn yields Theorem 3.1.

With the help of Theorem 3.1, we can investigate the theoretical properties of our multiscale test. The first result is an immediate consequence of Theorem 3.1. It says that the test has the correct (asymptotic) size.

Proposition 3.2. *Let the conditions of Theorem 3.1 be satisfied. Under the null hypothesis $H_0 : m = 0$, it holds that*

$$\mathbb{P}\big(\widehat{\Psi}_T \leq q_T(\alpha) \big) = (1 - \alpha) + o(1).$$

The second result characterizes the power of the multiscale test against local alternatives. To formulate it, we consider any sequence of functions $m = m_T$ with the following property: There exists $(u, h) \in \mathcal{G}_T$ with $[u - h, u + h] \subseteq [0, 1]$ such that

$$m_T(w) \geq c_T \sqrt{\frac{\log T}{Th}} \quad \text{for all } w \in [u - h, u + h], \tag{3.8}$$

where $\{c_T\}$ is any sequence of positive numbers with $c_T \to \infty$. Alternatively to (3.8), we may also assume that $-m_T(w) \geq c_T \sqrt{\log T/(Th)}$ for all $w \in [u - h, u + h]$. According to the following result, our test has asymptotic power 1 against local alternatives of the form (3.8).

Proposition 3.3. *Let the conditions of Theorem 3.1 be satisfied and consider any sequence of functions m_T with the property* (3.8)*. Then*

$$\mathbb{P}\big(\widehat{\Psi}_T \leq q_T(\alpha) \big) = o(1).$$

The proof of Proposition 3.3 can be found in the Appendix. To formulate the next result, we define

$$\Pi_T = \big\{ I_{u,h} = [u-h, u+h] : (u,h) \in \mathcal{A}_T \big\}$$

with

$$\mathcal{A}_T = \Big\{ (u,h) \in \mathcal{G}_T : \Big| \frac{\widehat{\psi}_T(u,h)}{\widehat{\sigma}} \Big| - \lambda(h) > q_T(\alpha) \Big\}.$$

Π_T is the collection of intervals $I_{u,h} = [u-h, u+h]$ for which the (corrected) test statistic $|\widehat{\psi}_T(u,h)/\widehat{\sigma}| - \lambda(h)$ lies above the critical value $q_T(\alpha)$. With this notation at hand, we consider the event

$$E_T = \Big\{ \forall I_{u,h} \in \Pi_T : m(v) \neq 0 \text{ for some } v \in I_{u,h} = [u-h, u+h] \Big\}.$$

This is the event that the null hypothesis is violated on all intervals $I_{u,h}$ for which the (corrected) test statistic $|\widehat{\psi}_T(u,h)/\widehat{\sigma}| - \lambda(h)$ is above the critical value $q_T(\alpha)$. We can make the following formal statement about the event E_T whose proof is given in the Appendix.

Proposition 3.4. *Under the conditions of Theorem 3.1, it holds that*

$$\mathbb{P}\big(E_T\big) \geq (1-\alpha) + o(1).$$

According to Proposition 3.4, our test procedure allows to make uniform confidence statements of the following form: With (asymptotic) probability $\geq (1-\alpha)$, the null hypothesis $H_0 : m = 0$ is violated on all intervals $I_{u,h} \in \Pi_T$. Hence, our multiscale test does not only allow to check whether the null hypothesis is violated. It also allows to identify regions where violations occur with a pre-specified level of confidence.

The statement of Proposition 3.4 suggests to graphically present the results of our multiscale test by plotting the intervals $I_{u,h} \in \Pi_T$, that is, by plotting the intervals where (with asymptotic confidence $\geq 1-\alpha$) our test detects a violation of the null hypothesis. The drawback of this graphical presentation is that the number of intervals in Π_T is often quite large. To obtain a better graphical summary of the results, we replace Π_T by a subset $\Pi_T^{\min}$ which is constructed as follows: As in Dümbgen (2002), we call an interval $I_{u,h} \in \Pi_T$ minimal if there is no other interval $I_{u',h'} \in \Pi_T$ with $I_{u',h'} \subset I_{u,h}$. Let $\Pi_T^{\min}$ be the set of all minimal intervals in Π_T and define the event

$$E_T^{\min} = \Big\{ \forall I_{u,h} \in \Pi_T^{\min} : m(v) \neq 0 \text{ for some } v \in I_{u,h} = [u-h, u+h] \Big\}.$$

It is easily seen that $E_T = E_T^{\min}$. Hence, by Proposition 3.4, it holds that

$$\mathbb{P}\big(E_T^{\min}\big) \geq (1-\alpha) + o(1).$$

This suggests to plot the minimal intervals in $\Pi_T^{\min}$ rather than the whole collection of intervals Π_T as a graphical summary of the test results. We in particular use this way of presenting the test results in our application examples of Section 7.

4 Testing for the presence of a time trend

In what follows, we construct a multiscale test for the null hypothesis that the trend function m in model (2.1) is constant. To achieve this, we adapt the methodology developed in Section 3. Importantly, the resulting multiscale procedure does not only allow to test whether the null hypothesis is violated. As we will see, it also allows to identify, with a pre-specified statistical confidence, time regions where violations occur. Put differently, it allows to identify, with a given confidence, intervals $I_{u,h} = [u-h, u+h]$ where m is not constant over time. It thus provides information on where the time trend is increasing/decreasing, which is important knowledge in many applications.

4.1 Construction of the test statistic

Throughout the section, we suppose that the trend m is continuously differentiable. The null hypothesis that m is constant can be formulated as $H_0 : m' = 0$, where m' denotes the first derivative of m. To construct a test statistic for the hypothesis H_0, we proceed analogously as in Section 3.1. To start with, we introduce the kernel averages

$$\widehat{\psi}'_T(u,h) = \sum_{t=1}^{T} w'_{t,T}(u,h) Y_t,$$

where the kernel weights $w'_{t,T}(u,h)$ are given by

$$w'_{t,T}(u,h) = \frac{\Lambda'_{t,T}(u,h)}{\{\sum_{t=1}^{T} \Lambda'_{t,T}(u,h)^2\}^{1/2}} \tag{4.1}$$

with

$$\Lambda'_{t,T}(u,h) = K\Big(\frac{\frac{t}{T}-u}{h}\Big)\Big[S_{T,0}(u,h)\Big(\frac{\frac{t}{T}-u}{h}\Big) - S_{T,1}(u,h)\Big].$$

Here, $S_{T,\ell}(u,h)$ is defined as in Section 3.1 and K is a kernel function which satisfies (C4). The kernel average $\widehat{\psi}'_T(u,h)$ is a rescaled version of the local linear estimator of the derivative $m'(u)$ with bandwidth h. Alternatively to the local linear weights defined in (4.1), we could employ the weights $w'_{t,T}(u,h) = K'(\frac{u-\frac{t}{T}}{h})/\{\sum_{t=1}^{T} K'(\frac{u-\frac{t}{T}}{h})^2\}^{1/2}$, where the kernel function K is assumed to be differentiable and K' is its derivative. To avoid boundary problems, we however work with the local linear weights from (4.1) throughout the paper. Our multiscale statistic is defined as

$$\widehat{\Psi}'_T = \max_{(u,h)\in\mathcal{G}_T} \Big\{\Big|\frac{\widehat{\psi}'_T(u,h)}{\widehat{\sigma}}\Big| - \lambda(h)\Big\},$$

where $\lambda(h) = \sqrt{2 \log\{1/(2h)\}}$ and the set $\mathcal{G}_T$ has been introduced in Section 3.1. As can be seen, the statistic $\widehat{\Psi}'_T$ is very similar to that from Section 3. Only the kernel averages $\widehat{\psi}'_T(u,h)$ have a somewhat different form.

4.2 The test procedure

As in Section 3.2, we define a Gaussian version Φ'_T of the test statistic $\widehat{\Psi}'_T$ under the null hypothesis H_0 by

$$\Phi'_T = \max_{(u,h) \in \mathcal{G}_T} \Big\{ \Big| \frac{\phi'_T(u,h)}{\sigma} \Big| - \lambda(h) \Big\},$$

where $\phi'_T(u,h) = \sum_{t=1}^T w'_{t,T}(u,h)\, \sigma Z_t$ and Z_t are independent standard normal random variables. Denoting the $(1-\alpha)$-quantile of Φ'_T by $q'_T(\alpha)$, our multiscale test of the hypothesis H_0: $m' = 0$ is defined as follows: For a given significance level $\alpha \in (0,1)$, we reject H_0 if $\widehat{\Psi}'_T > q'_T(\alpha)$.

4.3 Theoretical properties of the test

The theoretical analysis parallels that of Section 3.3. We first investigate the theoretical properties of the auxiliary statistic

$$\widehat{\Phi}'_T = \max_{(u,h) \in \mathcal{G}_T} \Big\{ \Big| \frac{\widehat{\phi}'_T(u,h)}{\widehat{\sigma}} \Big| - \lambda(h) \Big\},$$

where $\widehat{\phi}'_T(u,h) = \sum_{t=1}^T w'_{t,T}(u,h) \varepsilon_t$. The following result adapts Theorem 3.1 to our current test problem.

Theorem 4.1. *Let (C1)–(C6) be fulfilled and assume that* $\widehat{\sigma}^2 = \sigma^2 + o_p(\rho_T)$ *with* $\rho_T = o(1/\log T)$. *Then*

$$\mathbb{P}\big(\widehat{\Phi}'_T \le q'_T(\alpha)\big) = (1-\alpha) + o(1).$$

The proof of Theorem 4.1 is essentially the same as that of Theorem 3.1 and thus omitted. With the help of Theorem 4.1, we can derive the following theoretical properties of our multiscale test.

Proposition 4.2. *Let the conditions of Theorem 4.1 be satisfied.*

(a) Under the null hypothesis H_0, it holds that

$$\mathbb{P}\big(\widehat{\Psi}'_T \le q'_T(\alpha)\big) = (1-\alpha) + o(1).$$

(b) Consider any sequence of functions $m = m_T$ with the following property: There exists $(u,h) \in \mathcal{G}_T$ with $[u-h, u+h] \subseteq [0,1]$ such that $m'_T(w) \ge c_T \sqrt{\log T/(Th^3)}$ for all $w \in [u-h, u+h]$ or $-m'_T(w) \ge c_T \sqrt{\log T/(Th^3)}$ for all $w \in [u-h, u+h]$, where $\{c_T\}$ is any sequence of positive numbers with $c_T \to \infty$. Then

$$\mathbb{P}\big(\widehat{\Psi}'_T \le q'_T(\alpha)\big) = o(1).$$

Part (a) of Proposition 4.2 is a simple consequence of Theorem 4.1. Part (b) can be proven by similar arguments as Proposition 3.3. The details are given in the Supplementary Material. Taken together, the two parts of Proposition 4.2 show that our multiscale test has the correct (asymptotic) size and that it is able to detect certain local alternatives with probability tending to 1. We next consider the events

$$\begin{aligned} E_T^+ &= \Big\{\forall I_{u,h} \in \Pi_T^+ : m'(v) > 0 \text{ for some } v \in I_{u,h} = [u-h, u+h]\Big\} \\ E_T^- &= \Big\{\forall I_{u,h} \in \Pi_T^- : m'(v) < 0 \text{ for some } v \in I_{u,h} = [u-h, u+h]\Big\}, \end{aligned}$$

where the sets Π_T^+ and Π_T^- are given by

$$\begin{aligned} \Pi_T^+ &= \big\{I_{u,h} = [u-h, u+h] : (u,h) \in \mathcal{A}_T^+ \text{ and } I_{u,h} \subseteq [0,1]\big\} \\ \Pi_T^- &= \big\{I_{u,h} = [u-h, u+h] : (u,h) \in \mathcal{A}_T^- \text{ and } I_{u,h} \subseteq [0,1]\big\} \end{aligned}$$

with

$$\begin{aligned} \mathcal{A}_T^+ &= \Big\{(u,h) \in \mathcal{G}_T : \frac{\widehat{\psi}'_T(u,h)}{\widehat{\sigma}} > q'_T(\alpha) + \lambda(h)\Big\} \\ \mathcal{A}_T^- &= \Big\{(u,h) \in \mathcal{G}_T : -\frac{\widehat{\psi}'_T(u,h)}{\widehat{\sigma}} > q'_T(\alpha) + \lambda(h)\Big\}. \end{aligned}$$

E_T^+ is the event that for each interval $I_{u,h} \in \Pi_T^+$, there is a subset $J_{u,h} \subseteq I_{u,h}$ with m being an increasing function on $J_{u,h}$. An analogous description applies to the event E_T^-. The following result shows that the events E_T^+ and E_T^- occur with asymptotic probability $\ge 1-\alpha$.

Proposition 4.3. *Under the conditions of Theorem 4.1, it holds that*

$$\mathbb{P}\big(E_T^+\big) \geq (1-\alpha) + o(1)$$
$$\mathbb{P}\big(E_T^-\big) \geq (1-\alpha) + o(1).$$

The proof of Proposition 4.3 parallels that of Proposition 3.4. The details are provided in the Supplementary Material. The statement of Proposition 4.3 can be summarized as follows: With asymptotic probability $\geq 1-\alpha$, there is a subset $J_{u,h} \subseteq I_{u,h}$ for each interval $I_{u,h} \in \Pi_T^+$ such that m is an increasing function on $J_{u,h}$. Put differently, with asymptotic probability $\geq 1-\alpha$, the trend m is increasing on some part of the interval $I_{u,h}$ for any $I_{u,h} \in \Pi_T^+$. An analogous statement holds for the intervals in the set Π_T^-. Our multiscale procedure thus allows to identify, with a pre-specified confidence, time regions where there is an increase/decrease in the time trend m.

We close the section with some additional remarks on Proposition 4.3: (i) The statement of Proposition 4.3 remains to hold true when we replace the sets Π_T^+ and Π_T^- by the corresponding sets of minimal intervals. (ii) In the sets Π_T^+ and Π_T^-, we only take into account intervals $I_{u,h} = [u-h, u+h]$ which are subsets of $[0,1]$. We thus exclude points $(u,h) \in \mathcal{A}_T^+$ and $(u,h) \in \mathcal{A}_T^-$ which lie at the boundary, that is, for which $I_{u,h} \nsubseteq [0,1]$. The reason is as follows: Let $(u,h) \in \mathcal{A}_T^+$ with $I_{u,h} \nsubseteq [0,1]$. Our technical arguments allow us to say, with asymptotic confidence $\geq 1-\alpha$, that $m'(v) \neq 0$ for some $v \in I_{u,h}$. However, we cannot say whether $m'(v) > 0$ or $m'(v) < 0$, that is, we cannot make confidence statements about the sign. Roughly speaking, the problem is that the local linear weights $w'_{t,T}(u,h)$ behave quite differently at boundary points (u,h) with $I_{u,h} \nsubseteq [0,1]$. If we are only interested in whether there is some movement in the trend on an interval $I_{u,h}$ but we do not care whether it is an upward or downward movement, we may also consider the event $E_T^\pm = \{\forall I_{u,h} \in \Pi_T^\pm : m'(v) \neq 0 \text{ for some } v \in I_{u,h}\}$, where the set $\Pi_T^\pm = \{I_{u,h} : (u,h) \in \mathcal{A}_T^+ \cup \mathcal{A}_T^-\}$ contains all intervals $I_{u,h}$ with $(u,h) \in \mathcal{A}_T^+ \cup \mathcal{A}_T^-$, in particular those with $I_{u,h} \nsubseteq [0,1]$. With the help of the technical arguments for Proposition 4.3, it follows that $\mathbb{P}\big(E_T^\pm\big) \geq (1-\alpha) + o(1)$.

5 Estimation of the long-run error variance

We now discuss how to estimate the long-run error variance $\sigma^2 = \sum_{\ell=-\infty}^{\infty} \gamma(\ell)$ with $\gamma(\ell) = \mathrm{Cov}(\varepsilon_0, \varepsilon_\ell)$ in model (2.1). The same methods can be applied in the context of model (**??**). A number of different methods have been established in the literature to estimate the long-run error variance σ^2 in the trend model (2.1) under various assumptions on the error terms. In what follows, we give a brief overview of estimation methods which are suitable for our purposes. We in particular focus attention on difference-based methods as these have the following advantage: They do not involve a nonparametric estimator of the function m and thus do not require to specify a smoothing parameter for the estimation of m.

In principle, it is possible to construct an estimator of σ^2 under the general conditions on the error process laid out in Section 2 (or at least under somewhat stronger versions of these conditions). However, as is well-known, it is quite involved to estimate the long-run variance of a time series process under general conditions, the resulting estimators often tending to be quite imprecise. From a practical point of view, one might thus prefer to impose some time series model on the error terms and to estimate σ^2 under the restrictions of this model. Of course, this will create some bias due to misspecification. However, as long as the model gives a reasonable approximation to the true error process, this bias may very well be less severe than the error stemming from the instable behaviour of a general estimator of σ^2. In what follows, we consider an autoregressive (AR) model for the error terms since this error model is widely used in practice and is also appropriate for our applications in Section 7.

5.1 Independent error terms

Before we discuss the case of autoregressive error terms, we introduce the idea of difference-based methods for estimating σ^2 in the simple case of i.i.d. errors ε_t. In this case, σ^2 is identical to the variance of the random variables ε_t, that is, $\sigma^2 = \mathrm{Var}(\varepsilon_t)$. Let $D_\ell Y_t = Y_t - Y_{t-\ell}$ denote the difference between Y_t and $Y_{t-\ell}$ and suppose that m is sufficiently smooth. In particular, assume that m is Lipschitz continuous on $[0,1]$, that is, $|m(u) - m(v)| \le C|u-v|$ for all $u, v \in [0,1]$ and some constant $C < \infty$. Under these conditions, it holds that $|m(\frac{t}{T}) - m(\frac{t-\ell}{T})| \le C\ell/T$, which implies that $D_\ell Y_t = D_\ell \varepsilon_t + O(\ell/T)$ uniformly over t. Hence, the observed differences $D_\ell Y_t$ approximate the unobserved differences of the error terms $D_\ell \varepsilon_t$. This together with the fact that $\mathbb{E}[\{D_\ell \varepsilon_t\}^2]/2 = \sigma^2$ suggests to estimate σ^2 by $\widehat{\sigma}^2 = (T-\ell)^{-1} \sum_{t=\ell+1}^{T} \{D_\ell Y_t\}^2/2$, where most commonly $\ell = 1$. As can be easily verified, the estimator $\widehat{\sigma}^2$ has the property that $\widehat{\sigma}^2 = \sigma^2 + O_p(T^{-1/2})$.

5.2 Autoregressive error terms

The differencing approach presented above can be extended to more complicated error structures. For the case of k-dependent error terms, estimators for σ^2 have been proposed by Müller and Stadtmüller (1988), Herrmann et al. (1992) and Tecuapetla-Gómez and Munk (2017) among others. We here focus attention on the case of autoregressive error terms. Specifically, we suppose that $\{\varepsilon_t\}$ is an AR(p) process of the form $\varepsilon_t = \sum_{j=1}^{p} a_j \varepsilon_{t-j} + \eta_t$, where $a_1, \ldots, a_p$ are unknown parameters and η_t are i.i.d. innovations with $\mathbb{E}[\eta_t] = 0$ and $\mathbb{E}[\eta_t^2] = \sigma_\eta^2$. Throughout the discussion, we assume that $\{\varepsilon_t\}$ is a stationary and causal AR(p) process of known order p with finite fourth moment $\mathbb{E}[\varepsilon_t^4] < \infty$. A difference-based method to estimate the long-run variance σ^2 of the

AR(p) error process $\{\varepsilon_t\}$ in model (2.1) has been developed in Hall and Van Keilegom (2003). Their estimator $\widehat{\sigma}^2$ is constructed in the following three steps:

Step 1. We first set up an estimator of the autocovariance $\gamma(\ell) = \mathrm{Cov}(\varepsilon_t, \varepsilon_{t+\ell})$ for a given lag ℓ. As in the case of independent errors, it holds that $D_\ell Y_t = D_\ell \varepsilon_t + O(\ell/T)$ uniformly over t provided that m is Lipschitz. This together with the fact that $\mathbb{E}[\{D_\ell \varepsilon_t\}^2]/2 = \gamma(0) - \gamma(\ell)$ motivates to estimate $\gamma(0)$ by $\widehat{\gamma}(0) = \frac{1}{L_2 - L_1 + 1} \sum_{r=L_1}^{L_2} \frac{1}{2(T-r)} \sum_{t=r+1}^{T} \{D_r Y_t\}^2$, where $L_1 \le L_2$ are tuning parameters which are discussed in more detail below. Moreover, an estimator of $\gamma(\ell)$ for $1 \le \ell \le p$ is given by $\widehat{\gamma}(\ell) = \widehat{\gamma}(0) - \frac{1}{2(T-\ell)} \sum_{t=\ell+1}^{T} \{D_\ell Y_t\}^2$. As $\gamma(\ell) = \gamma(-\ell)$, we finally set $\widehat{\gamma}(-\ell) = \widehat{\gamma}(\ell)$ for $1 \le \ell \le p$.

Step 2. We next estimate the AR coefficients $(a_1, \dots, a_p)^\top$ by the Yule-Walker estimators $(\widehat{a}_1, \dots, \widehat{a}_p)^\top = \widehat{\Gamma}^{-1}(\widehat{\gamma}(1), \dots, \widehat{\gamma}(p))^\top$, where $\widehat{\Gamma} = \{\widehat{\gamma}(|k - \ell|)\}_{1 \le k, \ell \le p}$.

Step 3. Let $\widehat{d}_0 = 1$ and define the parameters $\widehat{d}_1, \widehat{d}_2, \dots$ by the equation $1 + \sum_{\ell=1}^{\infty} \widehat{d}_\ell z^\ell = (1 - \sum_{j=1}^{p} \widehat{a}_j z^j)^{-1}$. In the AR(1) case $\varepsilon_t = a \varepsilon_{t-1} + \eta_t$ with $|a| < 1$, for instance, it holds that $\sum_{\ell=0}^{\infty} \widehat{a}^\ell z^\ell = (1 - \widehat{a} z)^{-1}$ and thus $\widehat{d}_\ell = \widehat{a}^\ell$ for $\ell \ge 1$. The variance $\sigma_\eta^2 = \mathbb{E}[\eta_t^2]$ of the innovations can be estimated by $\widehat{\sigma}_\eta^2 = \widehat{\gamma}(0)/(\sum_{\ell=0}^{\infty} \widehat{d}_\ell^2)$. With this notation at hand, we define

$$\widehat{\sigma}^2 = \widehat{\sigma}_\eta^2 \Big(1 - \sum_{j=1}^{p} \widehat{a}_j\Big)^{-2}$$

to be our estimator of the long-run error variance σ^2.

The estimator $\widehat{\sigma}^2$ depends on the two tuning parameters L_1 and L_2 which are required to compute $\widehat{\gamma}(0)$. To better understand the role of these tuning parameters, let us have a closer look at the estimator $\widehat{\gamma}(0)$. As $\mathbb{E}[\{D_\ell Y_t\}^2]/2 = \mathbb{E}[\{D_\ell \varepsilon_t\}^2]/2 + O(\{\ell/T\}^2) = \gamma(0) - \gamma(\ell) + O(\{\ell/T\}^2)$, it can be easily shown that

$$\mathbb{E}\big[\widehat{\gamma}(0)\big] = \gamma(0) - \frac{1}{L_2 - L_1 + 1} \sum_{r=L_1}^{L_2} \gamma(r) + O\Big(\Big\{\frac{L_2}{T}\Big\}^2\Big).$$

The two bias terms $\sum_{r=L_1}^{L_2} \gamma(r)/(L_2 - L_1 + 1)$ and $O(\{L_2/T\}^2)$ can be asymptotically neglected if we choose the tuning parameters L_1 and L_2 appropriately. Since $\{\varepsilon_t\}$ is an AR(p) process, the autocovariances $\gamma(r)$ decay exponentially fast to zero as $r \to \infty$. Hence, the bias term $\sum_{r=L_1}^{L_2} \gamma(r)/(L_2 - L_1 + 1)$ is asymptotically negligible if L_1 grows sufficiently fast with the sample size T. Due to the exponential decay of the autocovariances, it in particular suffices to assume that $L_1/\log T \to \infty$. For the second bias term $O(\{L_2/T\}^2)$ to be asymptotically negligible, we need to assume that L_2 grows more slowly than the sample size T. In practice, L_1 should be chosen so large that the autocovariances $\gamma(\ell)$ with $\ell \ge L_1$ can be expected to be close to zero, ensuring that the bias term $\sum_{r=L_1}^{L_2} \gamma(r)/(L_2 - L_1 + 1)$ is sufficiently small. The choice of L_2 can be expected to be less important in practice than that of L_1 as long as we do not pick L_2 too close to the sample size T. As pointed out in Hall and Van Keilegom (2003), it can be shown that $\widehat{\sigma}^2 = \sigma^2 + O_p(T^{-1/2})$ provided that $L_1/\log T \to \infty$ and $L_2 = O(T^{1/2})$.

6 Simulations

To assess the finite sample performance of the methods from Sections 4 and **??**, we conduct a number of simulations. We first investigate the test procedure from Section 4. The simulation design is set up to mimic the situation in the application example of Section 7.1: We generate data from the model $Y_t = m(\frac{t}{T}) + \varepsilon_t$ for different time series lengths T. The errors ε_t are drawn from the AR(1) process $\varepsilon_t = a\varepsilon_{t-1} + \eta_t$, where η_t are independent and normally distributed with mean 0 and variance σ_η^2. We set $a = 0.267$ and $\sigma_\eta^2 = 0.35$, thus matching the estimated values obtained in the application of Section 7.1. To simulate data under the null $H_0 : m' = 0$, we let m be a constant function. In particular, we set $m = 0$ without loss of generality. To generate data under the alternative, we consider the trend functions $m(u) = \beta\,(u - 0.6)\,1(0.6 \le u \le 1)$ with $\beta = 1.25, 1.875, 2.5$. These functions are broken lines with a kink at $u = 0.6$ and different slopes β. The slope parameter β corresponds to a trend with the value $m(1) = 0.4\,\beta$ at the right endpoint $u = 1$. We thus consider broken lines with the values $m(1) = 0.5, 0.75, 1.0$. Inspecting the middle panel of Figure 2, the broken line with the slope $\beta = 2.5$ can be seen to resemble the local linear trend estimates in the real-data example of Section 7.1 the most (where we neglect the nonlinearities of the local linear fits at the beginning of the observation period). The broken lines with the smaller slopes $\beta = 1.25$ and $\beta = 1.875$ are closer to the null making it harder for our test to detect these alternatives.

To implement our test, we choose K to be an Epanechnikov kernel and define the set $\mathcal{G}_T$ of location-scale points (u, h) as

$$\begin{aligned} \mathcal{G}_T = \big\{(u,h) : u = 5k/T \text{ for some } 1 \le k \le T/5 \text{ and} \\ h = (3 + 5\ell)/T \text{ for some } 0 \le \ell \le T/20\big\}. \end{aligned} \tag{6.1}$$

We thus take into account all rescaled time points $u \in [0,1]$ on an equidistant grid with step length $5/T$. For the bandwidth $h = (3+5\ell)/T$ and any $u \in [h, 1-h]$, the local linear weights $w'_{t,T}(u,h)$ are non-zero for exactly $5 + 10\ell$ observations. Hence, the bandwidths h in $\mathcal{G}_T$ correspond to effective sample sizes of $5, 15, 25, \dots$ up to approximately $T/2$ data points. We estimate the long-run error variance σ^2 by the procedure from Section 5.2, setting the tuning parameters L_1 and L_2 to $\lfloor\sqrt{T}\rfloor$ and $\lfloor 2\sqrt{T}\rfloor$, respectively. To compute the critical values of the test, we simulate 1000 values of the statistic Φ'_T defined in Section 4.2 and compute their empirical $(1-\alpha)$ quantile $q'_T(\alpha)$.

Tables 1 and 2 report the simulation results for the sample sizes $T = 250, 350, 500, 1000$

Table 1: Size of the multiscale test from Section 4 for different sample sizes T and nominal sizes α.

	nominal size α		
T	0.01	0.05	0.1
250	0.004	0.039	0.092
350	0.012	0.051	0.069
500	0.006	0.047	0.094
1000	0.014	0.058	0.105

Table 2: Power of the multiscale test from Section 4 for different sample sizes T and nominal sizes α. Each panel corresponds to a different slope parameter β.

(a) $\beta = 1.25$

	nominal size α		
T	0.01	0.05	0.1
250	0.085	0.252	0.341
350	0.236	0.396	0.470
500	0.315	0.577	0.669
1000	0.763	0.900	0.936

(b) $\beta = 1.875$

	nominal size α		
T	0.01	0.05	0.1
250	0.318	0.621	0.714
350	0.648	0.796	0.865
500	0.793	0.943	0.967
1000	0.997	1.000	1.000

(c) $\beta = 2.5$

	nominal size α		
T	0.01	0.05	0.1
250	0.693	0.898	0.937
350	0.929	0.981	0.990
500	0.986	1.000	1.000
1000	1.000	1.000	1.000

and the significance levels $\alpha = 0.01, 0.05, 0.10$. The sample size $T = 350$ is approximately equal to the time series length 359 in the real-data example of Section 7.1. To produce our simulation results, we generate $S = 1000$ samples for each time series length T and carry out the multiscale test for each simulated sample. The entries of Tables 1 and 2 are computed as the number of simulations in which the test rejects divided by the total number of simulations. As can be seen from Table 1, the actual size of the test is fairly close to the nominal target α even for small values of T. Hence, the test has approximately the correct size. Inspecting Table 2, one can further see that the test has reasonable power properties. For the smallest value $\beta = 1.25$, the deviation from the null is quite small, making it hard for the test to detect the alternative. As a consequence, the power is only moderate for $T = 250$ and $T = 350$. When we move further away from the null by increasing the slope parameter β, the power of the test quickly increases. It can also be seen to rapidly get larger as the sample size grows. For the slope $\beta = 2.5$ and the sample size $T = 350$, which are the values that resemble the real-life data in Section 7.1 the most, the power of the test is above 92.9% for all significance levels α considered and thus comes quite close to 1.

7 Applications

In what follows, we illustrate the multiscale methods from Sections 4 and **??** by two real-data examples. In the first example, we apply the test method from Section 4 to a long time series of temperature data from Central England. In the second, we analyse

Table 3: Size of the multiscale test from Section **??** 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 4: Power of the multiscale test from Section **??** 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

a sample of temperature time series from 34 different weather stations in Great Britain with the help of the methods from Section **??**.

7.1 Analysis of Central England temperature data

The analysis of time trends in long temperature records is an important task in climatology. Information on the shape of the trend is needed in order to better understand long-term climate variability. The Central England temperature record is the longest instrumental temperature time series in the world. It is a valuable asset for analysing climate variability over the last few hundred years. The data is publicly available on the webpage of the UK Met Office. A detailed description of the data can be found in Parker et al. (1992). For our analysis, we use the dataset of yearly mean temperatures which consists of $T = 359$ observations covering the years from 1659 to 2017. We assume that the data follow the nonparametric trend model

$$Y_t = m\Big(\frac{t}{T}\Big) + \varepsilon_t,$$

where m is the unknown time trend of interest. The error process $\{\varepsilon_t\}$ is supposed to have the AR(1) structure $\varepsilon_t = a\varepsilon_{t-1} + \eta_t$, where η_t are i.i.d. innovations with mean 0 and variance σ_η^2. As pointed out in Mudelsee (2010) among others, this is the most widely used error model for discrete climate time series. We estimate the parameters a and σ_η^2 as described in Section 5.2 which yields the estimates $\widehat{a} \approx 0.267$ and $\widehat{\sigma}_\eta^2 \approx 0.35$. With the help of our multiscale method from Section 4, we test the null hypothesis

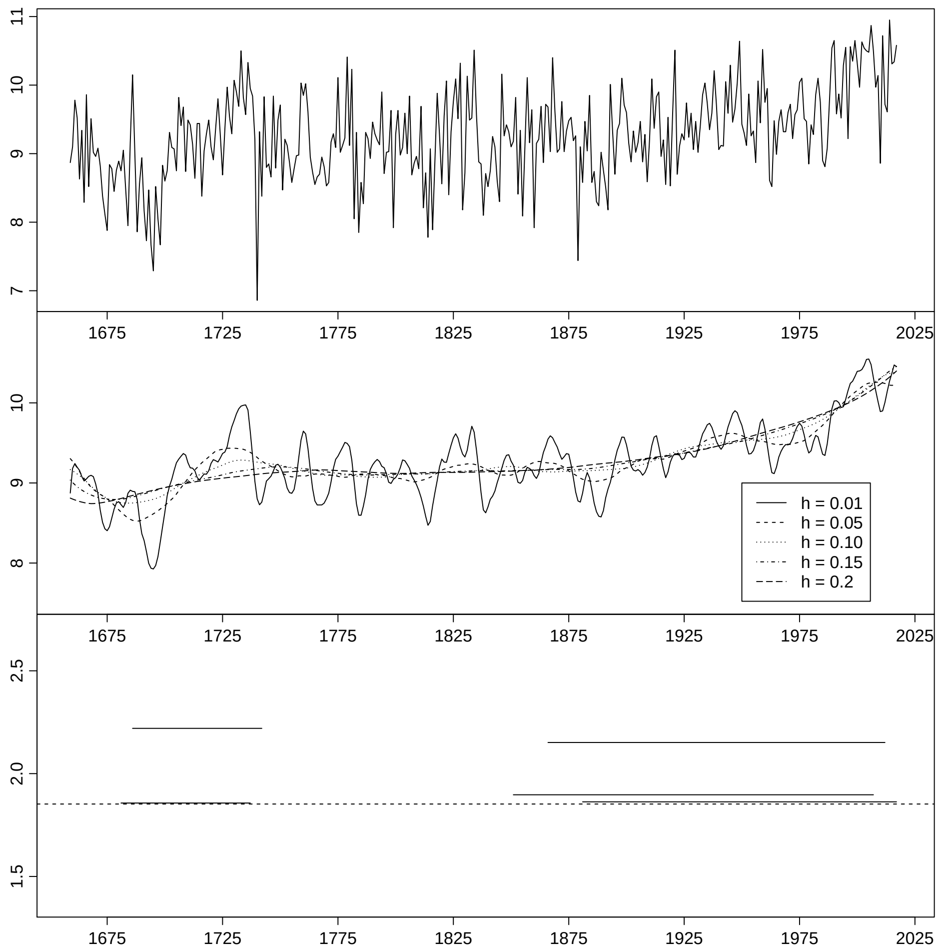

Figure 2: Summary of the application results for Section 7.1. The upper panel shows the Central England mean temperature time series. The middle panel depicts local linear kernel estimates of the time trend for a number of different bandwidths h. The lower panel presents the minimal intervals in the set Π_T^+ produced by the multiscale test. These are $[1681, 1737]$, $[1686, 1742]$, $[1851, 2007]$, $[1866, 2012]$ and $[1881, 2017]$.

$H_0 : m' = 0$, that is, the hypothesis that m is constant. To do so, we set the significance level to $\alpha = 0.05$ and implement the test in exactly the same way as in the simulations of Section 6. The results are presented in Figure 2. The upper panel shows the raw temperature time series, whereas the middle panel depicts local linear kernel estimates of the trend m for different bandwidths h. As one can see, the shape of the estimated time trend strongly differs with the chosen bandwidth. When the bandwidth is small, there are many local increases and decreases in the estimated trend. When the bandwidth is large, most of these local variations get smoothed out. Hence, by themselves, the nonparametric fits do not give much information on whether the trend m is increasing or decreasing in certain time regions.

Our multiscale test provides this kind of information, which is summarized in the lower panel of Figure 2. The plot depicts the minimal intervals contained in the set Π_T^+ which is defined in Section 4.3. The set of intervals Π_T^- is empty in the present case. The height at which a minimal interval $I_{u,h} = [u-h, u+h] \in \Pi_t^+$ is plotted indicates the value of the corresponding (additively corrected) test statistic $\widehat{\psi}'_T(u,h)/\widehat{\sigma} - \lambda(h)$. The dashed line specifies the critical value $q'_T(\alpha)$, where $\alpha = 0.05$ as already mentioned above. According to Proposition 4.3, we can make the following simultaneous confidence statement about the collection of minimal intervals in Π_T^+. We can claim, with confidence of about 95%, that the trend function m has some increase on each minimal interval. More specifically, we can claim with this confidence that there has been some upward movement in the trend both in the period from around 1680 to 1740 and in the period from about 1880 onwards. Hence, our test in particular provides evidence that there has been some warming trend in the period over approximately the last 140 years. On the other hand, as the set Π_T^- is empty, there is no evidence of any downward movement of the trend.

Appendix

In what follows, we prove the theoretical results from Section 3. The proofs of the results from Sections 4 and **??** are deferred to the Supplementary Material. Throughout the Appendix, we use the following notation: The symbol C denotes a universal real constant which may take a different value on each occurrence. For $a, b \in \mathbb{R}$, we write $a_+ = \max\{0, a\}$ and $a \vee b = \max\{a, b\}$. For any set A, the symbol $|A|$ denotes the cardinality of A. The notation $X \overset{\mathcal{D}}{=} Y$ means that the two random variables X and Y have the same distribution. Finally, $f_0(\cdot)$ and $F_0(\cdot)$ denote the density and distribution function of the standard normal distribution, respectively.

Auxiliary results using strong approximation theory

The main purpose of this section is to prove that there is a version of the multiscale statistic $\widehat{\Phi}_T$ defined in (3.4) which is close to a Gaussian statistic whose distribution is known. More specifically, we prove the following result.

Proposition A.1. *Under the conditions of Theorem 3.1, there exist statistics $\widetilde{\Phi}_T$ for $T = 1, 2, \ldots$ with the following two properties: (i) $\widetilde{\Phi}_T$ has the same distribution as $\widehat{\Phi}_T$ for any T, and (ii)*

$$\big|\widetilde{\Phi}_T - \Phi_T\big| = o_p\Big(\frac{T^{1/q}}{\sqrt{Th_{\min}}} + \rho_T \sqrt{\log T}\Big),$$

where Φ_T is a Gaussian statistic as defined in (3.3)*.*

Proof of Proposition A.1. For the proof, we draw on strong approximation theory for stationary processes $\{\varepsilon_t\}$ that fulfill the conditions (C1)–(C3). By Theorem 2.1 and Corollary 2.1 in Berkes et al. (2014), the following strong approximation result holds true: On a richer probability space, there exist a standard Brownian motion $\mathbb{B}$ and a sequence $\{\widetilde{\varepsilon}_t : t \in \mathbb{N}\}$ such that $[\widetilde{\varepsilon}_1, \ldots, \widetilde{\varepsilon}_T] \stackrel{\mathcal{D}}{=} [\varepsilon_1, \ldots, \varepsilon_T]$ for each T and

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

where $\sigma^2 = \sum_{k \in \mathbb{Z}} \mathrm{Cov}(\varepsilon_0, \varepsilon_k)$ denotes the long-run error variance. To apply this result, we define

$$\widetilde{\Phi}_T = \max_{(u,h) \in \mathcal{G}_T} \Big\{ \Big| \frac{\widetilde{\phi}_T(u,h)}{\widetilde{\sigma}} \Big| - \lambda(h) \Big\},$$

where $\widetilde{\phi}_T(u,h) = \sum_{t=1}^T w_{t,T}(u,h) \widetilde{\varepsilon}_t$ and $\widetilde{\sigma}^2$ is the same estimator as $\widehat{\sigma}^2$ with $Y_t = m(t/T) + \varepsilon_t$ replaced by $\widetilde{Y}_t = m(t/T) + \widetilde{\varepsilon}_t$ for $1 \le t \le T$. In addition, we let

$$\Phi_T = \max_{(u,h) \in \mathcal{G}_T} \Big\{ \Big| \frac{\phi_T(u,h)}{\sigma} \Big| - \lambda(h) \Big\}$$
$$\Phi_T^{\diamond} = \max_{(u,h) \in \mathcal{G}_T} \Big\{ \Big| \frac{\phi_T(u,h)}{\widetilde{\sigma}} \Big| - \lambda(h) \Big\}$$

with $\phi_T(u,h) = \sum_{t=1}^T w_{t,T}(u,h) \sigma Z_t$ and $Z_t = \mathbb{B}(t) - \mathbb{B}(t-1)$. With this notation, we can write

$$\big| \widetilde{\Phi}_T - \Phi_T \big| \le \big| \widetilde{\Phi}_T - \Phi_T^{\diamond} \big| + \big| \Phi_T^{\diamond} - \Phi_T \big| = \big| \widetilde{\Phi}_T - \Phi_T^{\diamond} \big| + o_p\big(\rho_T \sqrt{\log T}\big), \tag{A.2}$$

where the last equality follows by taking into account that $\phi_T(u,h) \sim N(0, \sigma^2)$ for all $(u,h) \in \mathcal{G}_T$, $|\mathcal{G}_T| = O(T^\theta)$ for some large but fixed constant θ and $\widetilde{\sigma}^2 = \sigma^2 + o_p(\rho_T)$. Straightforward calculations yield that

$$\big| \widetilde{\Phi}_T - \Phi_T^{\diamond} \big| \le \widetilde{\sigma}^{-1} \max_{(u,h) \in \mathcal{G}_T} \big| \widetilde{\phi}_T(u,h) - \phi_T(u,h) \big|.$$

Using summation by parts, we further obtain that

$$\begin{aligned} \big| \widetilde{\phi}_T(u,h) - \phi_T(u,h) \big| &\le W_T(u,h) \max_{1 \le t \le T} \Big| \sum_{s=1}^t \widetilde{\varepsilon}_s - \sigma \sum_{s=1}^t \big\{ \mathbb{B}(s) - \mathbb{B}(s-1) \big\} \Big| \\ &= W_T(u,h) \max_{1 \le t \le T} \Big| \sum_{s=1}^t \widetilde{\varepsilon}_s - \sigma \mathbb{B}(t) \Big|, \end{aligned}$$

where

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

Standard arguments show that $\max_{(u,h) \in \mathcal{G}_T} W_T(u,h) = O(1/\sqrt{T h_{\min}})$. Applying the strong approximation result (A.1), we can thus infer that

$$\begin{aligned}\big|\widetilde{\Phi}_T - \Phi_T^\diamond\big| &\le \widetilde{\sigma}^{-1} \max_{(u,h)\in\mathcal{G}_T} \big|\widetilde{\phi}_T(u,h) - \phi_T(u,h)\big| \\ &\le \widetilde{\sigma}^{-1} \max_{(u,h)\in\mathcal{G}_T} W_T(u,h) \max_{1\le t\le T} \Big|\sum_{s=1}^t \widetilde{\varepsilon}_s - \sigma\mathbb{B}(t)\Big| = o_p\Big(\frac{T^{1/q}}{\sqrt{Th_{\min}}}\Big). \qquad (A.3)\end{aligned}$$

Plugging (A.3) into (A.2) completes the proof. □

Auxiliary results using anti-concentration bounds

In this section, we establish some properties of the Gaussian statistic Φ_T defined in (3.3). We in particular show that Φ_T does not concentrate too strongly in small regions of the form $[x - \delta_T, x + \delta_T]$ with δ_T converging to zero.

Proposition A.2. *Under the conditions of Theorem 3.1, it holds that*

$$\sup_{x\in\mathbb{R}} \mathbb{P}\Big(|\Phi_T - x| \le \delta_T\Big) = o(1),$$

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

Proof of Proposition A.2. The main technical tool for proving Proposition A.2 are anti-concentration bounds for Gaussian random vectors. The following proposition slightly generalizes anti-concentration results derived in Chernozhukov et al. (2015), in particular Theorem 3 therein.

Proposition A.3. *Let* $(X_1, \ldots, X_p)^\top$ *be a Gaussian random vector in* $\mathbb{R}^p$ *with* $\mathbb{E}[X_j] = \mu_j$ *and* $\mathrm{Var}(X_j) = \sigma_j^2 > 0$ *for* $1 \le j \le p$. *Define* $\overline{\mu} = \max_{1\le j\le p}|\mu_j|$ *together with* $\underline{\sigma} = \min_{1\le j\le p}\sigma_j$ *and* $\overline{\sigma} = \max_{1\le j\le p}\sigma_j$. *Moreover, set* $a_p = \mathbb{E}[\max_{1\le j\le p}(X_j - \mu_j)/\sigma_j]$ *and* $b_p = \mathbb{E}[\max_{1\le j\le p}(X_j - \mu_j)]$. *For every* $\delta > 0$, *it holds that*

$$\sup_{x\in\mathbb{R}} \mathbb{P}\Big(\big|\max_{1\le j\le p} X_j - x\big| \le \delta\Big) \le C\delta\big\{\overline{\mu} + a_p + b_p + \sqrt{1 \vee \log(\underline{\sigma}/\delta)}\big\},$$

where $C > 0$ *depends only on* $\underline{\sigma}$ *and* $\overline{\sigma}$.

The proof of Proposition A.3 is provided in the Supplementary Material. To apply Proposition A.3 to our setting at hand, we introduce the following notation: We write $x = (u,h)$ along with $\mathcal{G}_T = \{x : x \in \mathcal{G}_T\} = \{x_1, \ldots, x_p\}$, where $p := |\mathcal{G}_T| \le O(T^\theta)$ for some large but fixed $\theta > 0$ by our assumptions. Moreover, for $j = 1, \ldots, p$, we set

$$\begin{aligned} X_{2j-1} &= \frac{\phi_T(x_{j1}, x_{j2})}{\sigma} - \lambda(x_{j2}) \\ X_{2j} &= -\frac{\phi_T(x_{j1}, x_{j2})}{\sigma} - \lambda(x_{j2}) \end{aligned}$$

with $x_j = (x_{j1}, x_{j2})$. This notation allows us to write

$$\Phi_T = \max_{1 \le j \le 2p} X_j,$$

where $(X_1, \ldots, X_{2p})^\top$ is a Gaussian random vector with the following properties: (i) $\mu_j := \mathbb{E}[X_j] = -\lambda(x_{j2})$ and thus $\overline{\mu} = \max_{1 \le j \le 2p} |\mu_j| \le C\sqrt{\log T}$, and (ii) $\sigma_j^2 := \mathrm{Var}(X_j) = 1$ for all j. Since $\sigma_j = 1$ for all j, it holds that $a_{2p} = b_{2p}$. Moreover, as the variables $(X_j - \mu_j)/\sigma_j$ are standard normal, we have that $a_{2p} = b_{2p} \le \sqrt{2\log(2p)} \le C\sqrt{\log T}$. With this notation at hand, we can apply Proposition A.3 to obtain that

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

with $\delta_T = T^{1/q}/\sqrt{T h_{\min}} + \rho_T \sqrt{\log T}$, which is the statement of Proposition A.2. $\square$

Proof of Theorem 3.1

To prove Theorem 3.1, we make use of the two auxiliary results derived above. By Proposition A.1, there exist statistics $\widetilde{\Phi}_T$ for $T = 1, 2, \ldots$ which are distributed as $\widehat{\Phi}_T$ for any $T \ge 1$ and which have the property that

$$\big|\widetilde{\Phi}_T - \Phi_T\big| = o_p\Big(\frac{T^{1/q}}{\sqrt{T h_{\min}}} + \rho_T \sqrt{\log T} \Big), \tag{A.4}$$

where Φ_T is a Gaussian statistic as defined in (3.3). The approximation result (A.4) allows us to replace the multiscale statistic $\widehat{\Phi}_T$ by an identically distributed version $\widetilde{\Phi}_T$ which is close to the Gaussian statistic Φ_T. In the next step, we show that

$$\sup_{x \in \mathbb{R}} \big| \mathbb{P}(\widetilde{\Phi}_T \le x) - \mathbb{P}(\Phi_T \le x) \big| = o(1), \tag{A.5}$$

which immediately implies the statement of Theorem 3.1. For the proof of (A.5), we use the following simple lemma:

Lemma A.4. *Let V_T and W_T be real-valued random variables for $T = 1, 2, \ldots$ such that $V_T - W_T = o_p(\delta_T)$ with some $\delta_T = o(1)$. If*

$$\sup_{x \in \mathbb{R}} \mathbb{P}(|V_T - x| \le \delta_T) = o(1), \tag{A.6}$$

then

$$\sup_{x \in \mathbb{R}} \big| \mathbb{P}(V_T \le x) - \mathbb{P}(W_T \le x) \big| = o(1). \tag{A.7}$$

The statement of Lemma A.4 can be summarized as follows: If W_T can be approximated by V_T in the sense that $V_T - W_T = o_p(\delta_T)$ and if V_T does not concentrate too strongly in small regions of the form $[x-\delta_T, x+\delta_T]$ as assumed in (A.6), then the distribution of W_T can be approximated by that of V_T in the sense of (A.7).

Proof of Lemma A.4. It holds that

$$\begin{aligned}
&\big|\mathbb{P}(V_T \le x) - \mathbb{P}(W_T \le x)\big| \\
&= \big|\mathbb{E}\big[1(V_T \le x) - 1(W_T \le x)\big]\big| \\
&\le \big|\mathbb{E}\big[\big\{1(V_T \le x) - 1(W_T \le x)\big\}1(|V_T - W_T| \le \delta_T)\big]\big| + \big|\mathbb{E}\big[1(|V_T - W_T| > \delta_T)\big]\big| \\
&\le \mathbb{E}\big[1(|V_T - x| \le \delta_T, |V_T - W_T| \le \delta_T)\big] + o(1) \\
&\le \mathbb{P}(|V_T - x| \le \delta_T) + o(1). \qquad \square
\end{aligned}$$

We now apply this lemma with $V_T = \Phi_T$, $W_T = \widetilde{\Phi}_T$ and $\delta_T = T^{1/q}/\sqrt{Th_{\min}} + \rho_T\sqrt{\log T}$: From (A.4), we already know that $\widetilde{\Phi}_T - \Phi_T = o_p(\delta_T)$. Moreover, by Proposition A.2, it holds that

$$\sup_{x\in\mathbb{R}} \mathbb{P}\Big(|\Phi_T - x| \le \delta_T\Big) = o(1). \tag{A.8}$$

Hence, the conditions of Lemma A.4 are satisfied. Applying the lemma, we obtain (A.5), which completes the proof of Theorem 3.1.

Proof of Proposition 3.3

Write $\widehat{\psi}_T(u,h) = \widehat{\psi}_T^A(u,h) + \widehat{\psi}_T^B(u,h)$ with $\widehat{\psi}_T^A(u,h) = \sum_{t=1}^T w_{t,T}(u,h)\varepsilon_t$ and $\widehat{\psi}_T^B(u,h) = \sum_{t=1}^T w_{t,T}(u,h)m_T(\frac{t}{T})$. By assumption, there exists $(u_0,h_0) \in \mathcal{G}_T$ with $[u_0 - h_0, u_0 + h_0] \subseteq [0,1]$ such that $m_T(w) \ge c_T\sqrt{\log T/(Th_0)}$ for all $w \in [u_0 - h_0, u_0 + h_0]$. 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

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

Together with the assumption that $m_T(w) \ge c_T\sqrt{\log T/(Th_0)}$ for all $w \in [u_0 - h_0, u_0 + h_0]$, this implies that

$$\widehat{\psi}_T^B(u_0,h_0) \ge c_T\sqrt{\frac{\log T}{Th_0}} \sum_{t=1}^T w_{t,T}(u_0,h_0). \tag{A.9}$$

Standard calculations exploiting the Lipschitz continuity of the kernel K show that for any $(u,h) \in \mathcal{G}_T$ and any given natural number ℓ,

$$\Big|\frac{1}{Th}\sum_{t=1}^T K\Big(\frac{\frac{t}{T} - u}{h}\Big)\Big(\frac{\frac{t}{T} - u}{h}\Big)^\ell - \int_0^1 \frac{1}{h}K\Big(\frac{w-u}{h}\Big)\Big(\frac{w-u}{h}\Big)^\ell dw\Big| \le \frac{C}{Th}, \tag{A.10}$$

where the constant C does not depend on u, h and T. With the help of (A.10), we obtain that for any $(u,h) \in \mathcal{G}_T$ with $[u-h, u+h] \subseteq [0,1]$,

$$\Big| \sum_{t=1}^{T} w_{t,T}(u,h) - \frac{\sqrt{Th}}{\kappa} \Big| \le \frac{C}{\sqrt{Th}}, \tag{A.11}$$

where $\kappa = (\int K^2(\varphi) d\varphi)^{1/2}$ and the constant C does once again not depend on u, h and T. From (A.11), it follows that $\sum_{t=1}^{T} w_{t,T}(u,h) \ge \sqrt{Th}/(2\kappa)$ for sufficiently large T and any $(u,h) \in \mathcal{G}_T$ with $[u-h, u+h] \subseteq [0,1]$. This together with (A.9) allows us to infer that

$$\widehat{\psi}_T^B(u_0, h_0) \ge \frac{c_T \sqrt{\log T}}{2\kappa} \tag{A.12}$$

for sufficiently large T. Moreover, arguments very similar to those for the proof of Proposition A.1 yield that

$$\max_{(u,h) \in \mathcal{G}_T} |\widehat{\psi}_T^A(u,h)| = O_p(\sqrt{\log T}). \tag{A.13}$$

With the help of (A.12), (A.13) and the fact that $\lambda(h) \le \lambda(h_{\min}) \le C\sqrt{\log T}$, we finally arrive at

$$\begin{aligned}
\widehat{\Psi}_T &\ge \max_{(u,h) \in \mathcal{G}_T} \frac{|\widehat{\psi}_T^B(u,h)|}{\widehat{\sigma}} - \max_{(u,h) \in \mathcal{G}_T} \Big\{ \frac{|\widehat{\psi}_T^A(u,h)|}{\widehat{\sigma}} + \lambda(h) \Big\} \\
&= \max_{(u,h) \in \mathcal{G}_T} \frac{|\widehat{\psi}_T^B(u,h)|}{\widehat{\sigma}} + O_p(\sqrt{\log T}) \\
&\ge \frac{c_T \sqrt{\log T}}{2\kappa \widehat{\sigma}} + O_p(\sqrt{\log T}). \qquad\qquad \text{(A.14)}
\end{aligned}$$

Since $q_T(\alpha) = O(\sqrt{\log T})$ for any fixed $\alpha \in (0,1)$, (A.14) immediately implies that $\mathbb{P}(\widehat{\Psi}_T \le q_T(\alpha)) = o(1)$.

Proof of Proposition 3.4

The statement of Proposition 3.4 is a consequence of the following observation: For all $(u,h) \in \mathcal{G}_T$ with

$$\Big| \frac{\widehat{\psi}_T(u,h) - \mathbb{E}\widehat{\psi}_T(u,h)}{\widehat{\sigma}} \Big| - \lambda(h) \le q_T(\alpha) \quad \text{and} \quad \Big| \frac{\widehat{\psi}_T(u,h)}{\widehat{\sigma}} \Big| - \lambda(h) > q_T(\alpha),$$

it holds that $\mathbb{E}[\widehat{\psi}_T(u,h)] \ne 0$, which in turn implies that $m(v) \ne 0$ for some $v \in I_{u,h}$. From this observation, we can infer the following: On the event

$$\big\{ \widehat{\Phi}_T \le q_T(\alpha) \big\} = \Big\{ \max_{(u,h) \in \mathcal{G}_T} \Big(\Big| \frac{\widehat{\psi}_T(u,h) - \mathbb{E}\widehat{\psi}_T(u,h)}{\widehat{\sigma}} \Big| - \lambda(h) \Big) \le q_T(\alpha) \Big\},$$

it holds that for all $(u, h) \in \mathcal{A}_T$, $m(v) \neq 0$ for some $v \in I_{u,h}$. Hence, we obtain that

$$\big\{\widehat{\Phi}_T \leq q_T(\alpha)\big\} \subseteq E_T.$$

As a result, we arrive at

$$\mathbb{P}(E_T) \geq \mathbb{P}\big(\widehat{\Phi}_T \leq q_T(\alpha)\big) = (1-\alpha) + o(1),$$

where the last equality holds by Theorem 3.1.

References

BENNER, T. C. (1999). Central england temperatures: long-term variability and teleconnections. *International Journal of Climatology*, **19** 391–403.

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

CHAUDHURI, P. and MARRON, J. S. (1999). SiZer for the exploration of structures in curves. *Journal of the American Statistical Association*, **94** 807–823.

CHAUDHURI, P. and MARRON, J. S. (2000). Scale space view of curve estimation. *Annals of Statistics*, **28** 408–428.

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

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

DEGRAS, D., XU, Z., ZHANG, T. and WU, W. B. (2012). Testing for parallelism among trends in multiple time series. *IEEE Transactions on Signal Processing*, **60** 1087–1097.

DETTE, H. (1999). A consistent test for the functional form of a regression based on a difference of variance estimators. *Annals of Statistics*, **27** 1012–1040.

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

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

HALL, P. and HECKMAN, N. E. (2000). Testing for monotonicity of a regression mean by calibrating for linear functions. *Annals of Statistics*, **28** 20–39.

HALL, P. and VAN KEILEGOM, I. (2003). Using difference-based methods for inference in nonparametric regression with time series errors. *Journal of the Royal Statistical Society: Series B*, **65** 443–456.

HERRMANN, E., GASSER, T. and KNEIP, A. (1992). Choice of bandwidth for kernel regression when residuals are correlated. *Biometrika*, **79** 783–795.

LYUBCHICH, V. and GEL, Y. R. (2016). A local factor nonparametric test for trend synchronism in multiple time series. *Journal of Multivariate Analysis*, **150** 91–104.

MUDELSEE, M. (2010). *Climate time series analysis: classical statistical and bootstrap methods.* New York, Springer.

MÜLLER, H.-G. and STADTMÜLLER, U. (1988). Detecting dependencies in smooth regression models. *Biometrika*, **75** 639–650.

PARK, C., MARRON, J. S. and RONDONOTTI, V. (2004). Dependent SiZer: goodness-of-fit tests for time series models. *Journal of Applied Statistics*, **31** 999–1017.

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

Parker, D. E., Legg, T. P. and Folland, C. K. (1992). A new daily central england temperature series, 1772-1991. *International Journal of Climatology*, **12** 317–342.

Rahmstorf, S., Foster, G. and Cahill, N. (2017). Global temperature evolution: recent trends and some pitfalls. *Environmental Research Letters*, **12**.

Rondonotti, V., Marron, J. S. and Park, C. (2007). SiZer for time series: a new approach to the analysis of trends. *Electronic Journal of Statistics*, **1** 268–289.

Tecuapetla-Gómez, I. and Munk, A. (2017). Autocovariance estimation in regression with a discontinuous signal and m-dependent errors: a difference-based approach. *Scandinavian Journal of Statistics*, **44** 346–368.

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

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

Wu, W. B. and Shao, X. (2004). Limit theorems for iterated random functions. *Journal of Applied Probability* 425–436.

Zhang, T. and Wu, W. B. (2011). Testing parametric assumptions of trends of a nonstationary time series. *Biometrika*, **98** 599–614.