Three Essays in Nonparametric Econometrics

Inauguraldissertation
zur Erlangung des Grades eines Doktors
der Wirtschafts- und Gesellschaftswissenschaften
durch die
Rechts- und Staatswissenschaftliche Fakultät
der Rheinischen Friedrich-Wilhelms-Universität
Bonn

vorgelegt von

Marina Khismatullina

aus Ufa, Russland

Bonn 2021

To my Mom

Dekan:	Prof. Dr. Jürgen von Hagen
Erstreferent:	Prof. Dr. Michael Vogt
Zweitreferent:	Prof. Dr. Alois Kneip

Tag der mündlichen Prüfung: 20. August 2021

Acknowledgements

It was the best of times, it was the worst of times...
"A tale of two cities", Charles Dickens

The last six years has been both wonderful and incredibly difficult, but I feel lucky that on this journey I have been surrounded by amazing people and here I would like to express my gratitudes to them. First and foremost, I would like to thank my first supervisor and my co-author, Michael Vogt, for his total support and tremendous help. He has taught me almost everything that I know now, and he was incredibly patient during the not-so-steep parts of my learning curve. I am grateful for his guidance and I could not have wished for a better supervisor. I would also like to thank my second supervisor, Alois Kneip, for his encouragement and outstanding expertise. Working together with him, I have grown a lot both as a researcher and as a teacher.

I would like to thank Lena Janys for being not only my third supervisor, but a friend as well. Our regular meetings always motivated and inspired me, and I am glad that we have had this tradition.

I would like to thank Joachim Freyberger and Nina Bobkova for their tremendous help during the most stressful time of my PhD, the job market. I have battled the job market mostly due to their excellent advice (even if sometimes it was not what I wanted to hear).

I thank BGSE, the Institute of Finance and Statistics (IFS) and Deutsche Forschungsgemeinschaft (DFG) for their generous financial and administrative support, as well as Almut Lunkenheimer for her invaluable help with the everyday university life. I would like to thank my colleagues at the IFS: Dominik Liebl, Christopher Walsh, Jörg Stoye and Sven Otto, for always regarding me as an equal and making me feel at home.

I am thankful to my friends in academia, including but not limited to Mikhail Ananyev, Kai Arvai, Lisa Dähne, Laura Ehrmantraut, Andreas Klümper, Renske Stans and Nadezhda Zhuravleva, for making this time fun! I laughed a lot together with you, thank you, guys! I am also grateful to my friends outside academia, Inera Julien-Vauzelle and Maria Melnikova, for tolerating my seldom phone calls and still being my friends.

I would have loved to thank my Mom who always believed that I would make a fine professor. She was a fine professor herself, and she always was and will be the role model for me in academia. It is a pity that she did not see my first paper published or my defence, but I know that she is proud of me. I miss her terribly.

Last but definitely not least, I would like to thank my husband, Aleksandr (Sasha) Dylev, who was by my side at every possible step of this long journey: from supporting my decision to start an academic career in another country through all the bureaucratic

obstacles and the ups and downs of the job market to months of long hours of finishing this thesis. I have always known that I can count on him, but these years proved it without even a shadow of the doubt. Sasha, I am happy to have you as my partner-in-crime in this next chapter of my life.

Contents

List of Figures

List of Tables

Introduction

This thesis consists of three self-contained essays in econometrics and statistics. In these essays, I am interested in the nonparametric regression models and in developing new methods for testing various qualitative hypotheses about the trend functions in these models. Each of the three chapters proposes a novel multiscale testing procedure that is used either for investigating the properties of one time series (Chapter 1), or for comparison of the regression curves between multiple time series (Chapters 2 and 3). The underlying idea of any multiscale test is to consider a number of test statistics (each corresponding to a different set of values of some tuning parameters) simultaneously rather than to perform a separate test for each single test statistics, which leads to a well-known multiple testing problem. All of the proposed tests account for this problem by picking appropriate critical values, and the main methodological contributions of the current thesis are the theoretical results that these test all have (asymptotically) correct size and good power properties. Even though there are many similarities between the chapters, the research questions are quite distinct. In Chapter 1, the method is designed to determine whether the trend in one time series is decreasing or increasing, whereas in Chapters 2 and 3 the testing procedures were designed for comparison of multiple time trends and locating the differences. Moreover, the difference between Chapters 2 and 3 lies in the models under consideration: Chapter 2 deals with epidemic time trends in a simple nonparametric regression and places certain restrictions on the error terms in the observed times series, whereas Chapter 3 considers a very general model that allows for including covariates and fixed effects. The first two chapters are also completed by extensive simulation studies and the applications to the real-life data: temperature time series in Chapter 1 and the data on the new cases of COVID-19 in Chapter 2.

Chapter 1 is based on a joint work together with Michael Vogt which is published in the *Journal of Royal Statistical Society: Series B* (Khismatullina and Vogt, 2020). Chapter 2 is based on a joint work together with Michael Vogt which is forthcoming at the *Journal of Econometrics* (Khismatullina and Vogt, 2021). Chapter 3 is based on yet another joint work together with Michael Vogt. In the following, I will provide a short summary of each chapter.

CHAPTER 1: In this paper, we develop new multiscale methods to test qualitative hypotheses about the function m in the nonparametric regression model $Y_{t,T} = m(t/T) + \varepsilon_t$ with time series errors ε_t. In time series applications, m represents a nonparametric time trend. Practitioners are often interested in whether the trend m has certain shape properties. For example, they would like to know whether m is constant or whether it is increasing/decreasing in certain time intervals. Our multiscale methods allow to test for

such shape properties of the trend m. In order to perform the methods, we require an estimator of the long-run error variance $\sigma^2 = \sum_{\ell=-\infty}^{\infty} \mathrm{Cov}(\varepsilon_0, \varepsilon_\ell)$. We propose a new difference-based estimator of σ^2 for the case that $\{\varepsilon_t\}$ belongs to the class of AR(∞) processes. In the technical part of the paper, we derive asymptotic theory for the proposed multiscale test and the estimator of the long-run error variance. The theory is complemented by a simulation study and an empirical application to climate data.

CHAPTER 2: The COVID-19 pandemic is still one of the most pressing issues at present. A question which is particularly important for governments and policy makers is the following: Does the virus spread in the same way in different countries? Or are there significant differences in the development of the epidemic? Identifying differences between countries may help, for instance, to better understand which government policies have been more effective in containing the virus than others. In this paper, we devise new inference methods that allow to detect differences in the development of the COVID-19 epidemic across countries in a statistically rigorous way. We derive asymptotic theory for the proposed methods and we complement the theory by an extensive simulation study. In our empirical study, we use the methods to compare the outbreak patterns of the epidemic in a number of European countries.

CHAPTER 3: We develop multiscale methods to test qualitative hypotheses about nonparametric time trends in the presence of covariates. In many applications, practitioners are interested whether the observed time series all have the same time trend. Moreover, when there is evidence that this is not the case, one of the major related statistical problems is to determine which of the trends are different and whether we can group the time series with the similar trends together. In addition, when two trends are not the same, it may also be relevant to know in which time regions they differ from each other. We design multiscale tests to formally approach these questions. We derive asymptotic theory for the proposed tests and show that the proposed test has (asymptotically) the correct size and has asymptotic power of one against a certain class of local alternatives.

Chapter 1

Multiscale Inference and Long-Run Variance Estimation in Nonparametric Regression with Time Series Errors

Joint with Michael Vogt

1.1 Introduction

The analysis of time trends is an important aspect of many time series applications. In a wide range of situations, practitioners are particularly interested in certain shape properties of the trend. They raise questions such as the following: Does the observed time series have a trend at all? If so, is the trend increasing/decreasing in certain time intervals? Can one identify the intervals of increase/decrease? As an example, consider the time series plotted in Figure 1.1 which shows the yearly mean temperature in Central England from 1659 to 2017. Climatologists are very much interested in learning about 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 this paper, we develop new methods to test for certain shape properties of a nonparametric time trend. We in particular construct a multiscale test which allows to identify

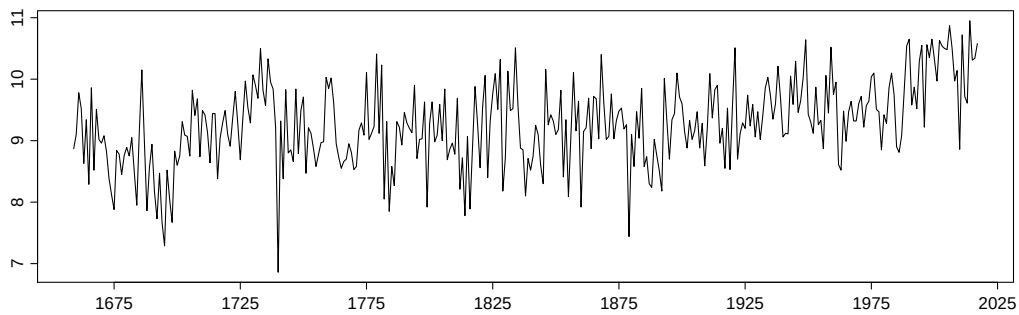


Figure 1.1: Yearly mean temperature in Central England from 1659 to 2017 measured in °C.

local increases/decreases of the trend function. We develop our test in the context of the following model setting: We observe a time series $\{Y_{t,T} : 1 \leq t \leq T\}$ of the form

$$Y_{t,T} = m\Big(\frac{t}{T}\Big) + \varepsilon_t \tag{1.1.1}$$

for $1 \leq t \leq T$, where $m : [0,1] \to \mathbb{R}$ is an unknown nonparametric regression function and the error terms ε_t form a stationary time series process with $\mathbb{E}[\varepsilon_t] = 0$. In a time series context, the design points t/T represent the time points of observation and m is a nonparametric time trend. As usual in nonparametric regression, we let the function m depend on rescaled time t/T rather than on real time t. A detailed description of model (1.1.1) is provided in Section 1.2.

Our multiscale test is developed step by step in Section 1.3. Roughly speaking, the procedure can be outlined as follows: Let $H_0(u,h)$ be the hypothesis that m is constant in the time window $[u-h, u+h] \subseteq [0,1]$, where u is the midpoint and $2h$ the size of the window. In a first step, we set up a test statistic $\widehat{\varphi}_T(u,h)$ for the hypothesis $H_0(u,h)$. In a second step, we aggregate the statistics $\widehat{\varphi}_T(u,h)$ for a large number of different time windows $[u-h, u+h]$. We thereby construct a multiscale statistic which allows to test the hypothesis $H_0(u,h)$ simultaneously for many time windows $[u-h, u+h]$. In the technical part of the paper, we derive the theoretical properties of the resulting multiscale test. To do so, we come up with a proof strategy which combines strong approximation results for dependent processes with anti-concentration bounds for Gaussian random vectors. This strategy is of interest in itself and may be applied to other multiscale test problems for dependent data. As shown by our theoretical analysis, our multiscale test is a rigorous level-α-test of the overall null hypothesis H_0 that $H_0(u,h)$ is simultaneously fulfilled for all time windows $[u-h, u+h]$ under consideration. Moreover, for a given significance level $\alpha \in (0,1)$, the test allows to make simultaneous confidence statements of the following form: We can claim, with statistical confidence $1-\alpha$, that there is an increase/decrease in the trend m on all time windows $[u-h, u+h]$ for which the hypothesis $H_0(u,h)$ is rejected. Hence, the test allows to identify, with a pre-specified statistical confidence, time intervals where the trend m is increasing/decreasing.

For independent data, multiscale tests have been developed in a variety of different contexts in recent years. In the regression context, Chaudhuri and Marron (1999, 2000) introduced the so-called SiZer method which has been extended in various directions; see e.g. Hannig and Marron (2006) where a refined distribution theory for SiZer is derived. Hall and Heckman (2000) constructed a multiscale test on monotonicity of a regression function. Dümbgen and Spokoiny (2001) developed a multiscale approach which works with additively corrected supremum statistics and derived theoretical results in the context of a continuous Gaussian white noise model. Rank-based multiscale tests for nonparametric regression were proposed in Dümbgen (2002) and Rohde (2008). More recently, Proksch et al. (2018) have constructed multiscale tests for inverse regression models. In the context of density estimation, multiscale tests have been investigated in Dümbgen and Walther

(2008), Rufibach and Walther (2010), Schmidt-Hieber et al. (2013) and Eckle et al. (2017) among others.

Whereas a large 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 SiZer methods for dependent data which can be used to find local increases/decreases of a trend and which may thus be regarded as an alternative to our multiscale test. However, these SiZer methods are mainly designed for data exploration rather than for rigorous statistical inference. Our multiscale method, in contrast, is a rigorous level-α-test of the hypothesis H_0 which allows to make simultaneous confidence statements about the time intervals where the trend m is increasing/decreasing. Some theoretical results for dependent SiZer methods have been derived in Park et al. (2009a), but only under a quite severe restriction: Only time windows $[u-h, u+h]$ with window sizes or scales h are taken into account that remain bounded away from zero as the sample size T grows. Scales h that converge to zero as T increases are excluded. This effectively means that only large time windows $[u-h, u+h]$ are taken into consideration. Our theory, in contrast, allows to simultaneously consider scales h of fixed size and scales h that converge to zero at various different rates. We are thus able to take into account time windows of many different sizes. In Section 1.3.4, we compare our approach to SiZer methods for dependent data in more detail.

Our multiscale approach is also related to Wavelet-based methods: Similar to the latter, it takes into account different locations u and resolution levels or scales h simultaneously. However, while our multiscale approach is designed to test for local increases/decreases of a nonparametric trend, Wavelet methods are commonly used for other purposes. Among other things, they are employed for estimating/reconstructing nonparametric regression curves [see e.g. Donoho et al. (1995) or Von Sachs and MacGibbon (2000)] and for change point detection [see e.g. Cho and Fryzlewicz (2012)].

The test statistic of our multiscale method depends on the long-run error variance $\sigma^2 = \sum_{\ell=-\infty}^{\infty} \mathrm{Cov}(\varepsilon_0, \varepsilon_\ell)$, which is usually unknown in practice. To carry out our multiscale test, we thus require an estimator of σ^2. Indeed, such an estimator is required for virtually all inferential procedures in the context of model (1.1.1). Hence, the problem of estimating σ^2 in model (1.1.1) is of broader interest and has received a lot of attention in the literature; see Müller and Stadtmüller (1988), Herrmann et al. (1992) and Hall and Van Keilegom (2003) among many others. In Section 1.4, we introduce a new difference-based estimator of σ^2 for the case that $\{\varepsilon_t\}$ belongs to the class of AR(∞) processes. This estimator improves on existing methods in several respects.

The methodological and theoretical analysis of the paper is complemented by a simulation study in Section 1.5 and two empirical applications in Section 1.6. In the simulation study, we examine the finite sample properties of our multiscale test and compare it to the dependent SiZer methods introduced in Park et al. (2004) and Rondonotti et al. (2007). Moreover, we investigate the small sample performance of our estimator of σ^2 and compare

it to the estimator of Hall and Van Keilegom (2003). In Section 1.6, we use our methods to analyse the temperature data from Figure 1.1 as well as a sample of global temperature data.

1.2 The model

We now describe the model setting in detail which was briefly outlined in the Introduction. We observe a time series $\{Y_{t,T} : 1 \leq t \leq T\}$ of length T which satisfies the nonparametric regression equation

$$Y_{t,T} = m\Big(\frac{t}{T}\Big) + \varepsilon_t \tag{1.2.1}$$

for $1 \leq t \leq T$. Here, m is an unknown nonparametric 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 $\varepsilon_t = G(\ldots, \eta_{t-1}, \eta_t, \eta_{t+1}, \ldots)$, where η_t are i.i.d. random variables and $G : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ 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 $\Theta_{t,q} = O(t^{-\tau_q}(\log t)^{-A})$, 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.

1.3 The multiscale test

In this section, we introduce our multiscale method to test for local increases/decreases of the trend function m and analyse its theoretical properties. We assume throughout that m is continuously differentiable on $[0,1]$. The test problem under consideration can be formulated as follows: Let $H_0(u,h)$ be the hypothesis that m is constant on the interval $[u-h, u+h]$. Since m is continuously differentiable, $H_0(u,h)$ can be reformulated as

$$H_0(u,h) : m'(w) = 0 \text{ for all } w \in [u-h, u+h],$$

where m' is the first derivative of m. We want to test the hypothesis $H_0(u,h)$ not only for a single interval $[u-h, u+h]$ but simultaneously for many different intervals. The overall null hypothesis is thus given by

$$H_0 : \text{ The hypothesis } H_0(u,h) \text{ holds true for all } (u,h) \in \mathcal{G}_T,$$

where $\mathcal{G}_T$ is some large set of points (u,h). The details on the set $\mathcal{G}_T$ are discussed at the end of Section 1.3.1 below. Note that $\mathcal{G}_T$ in general depends on the sample size T, implying that the null hypothesis $H_0 = H_{0,T}$ depends on T as well. We thus consider a sequence of null hypotheses $\{H_{0,T} : T = 1, 2, \ldots\}$ as T increases. For simplicity of notation, we however suppress the dependence of H_0 on T. In Sections 1.3.1 and 1.3.2, we step by step construct the multiscale test of the hypothesis H_0. The theoretical properties of the test are analysed in Section 1.3.3.

1.3.1 Construction of the multiscale statistic

We first construct a test statistic for the hypothesis $H_0(u,h)$, where $[u-h, u+h]$ is a given interval. To do so, we consider the kernel average

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

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

$$w_{t,T}(u,h) = \frac{\Lambda_{t,T}(u,h)}{\{\sum_{t=1}^{T} \Lambda_{t,T}(u,h)^2\}^{1/2}}, \tag{1.3.1}$$

where

$$\Lambda_{t,T}(u,h) = K\Big(\frac{\frac{t}{T}-u}{h}\Big)\Big[S_{T,0}(u,h)\Big(\frac{\frac{t}{T}-u}{h}\Big) - S_{T,1}(u,h)\Big],$$

$S_{T,\ell}(u,h) = (Th)^{-1}\sum_{t=1}^{T} K(\frac{\frac{t}{T}-u}{h})(\frac{\frac{t}{T}-u}{h})^{\ell}$ 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. More-

over, it has compact support $[-1,1]$ and is Lipschitz continuous, that is, $|K(v) - K(w)| \leq C|v-w|$ for any $v, w \in \mathbb{R}$ and some constant $C > 0$.

The kernel average $\widehat{\psi}_T(u,h)$ is nothing else than a rescaled local linear estimator of the derivative $m'(u)$ with bandwidth h.[1]

A test statistic for the hypothesis $H_0(u,h)$ is given by the normalized kernel average $\widehat{\psi}_T(u,h)/\widehat{\sigma}$, where $\widehat{\sigma}^2$ is an estimator of the long-run variance $\sigma^2 = \sum_{\ell=-\infty}^{\infty} \mathrm{Cov}(\varepsilon_0, \varepsilon_\ell)$ of the error process $\{\varepsilon_t\}$. The problem of estimating σ^2 is discussed in detail in Section 1.4. For the time being, we suppose that $\widehat{\sigma}^2$ is an estimator with reasonable theoretical properties. Specifically, we assume that $\widehat{\sigma}^2 = \sigma^2 + o_p(\rho_T)$ with $\rho_T = o(1/\log T)$. This is a fairly weak condition which is in particular satisfied by the estimator of σ^2 analysed in Section 1.4. The kernel weights $w_{t,T}(u,h)$ are chosen such that in the case of independent errors ε_t, $\mathrm{Var}(\widehat{\psi}_T(u,h)) = \sigma^2$ for any location u and bandwidth h, where the long-run error variance σ^2 simplifies to $\sigma^2 = \mathrm{Var}(\varepsilon_t)$. In the more general case that the error terms satisfy the weak dependence conditions from Section 1.2, $\mathrm{Var}(\widehat{\psi}_T(u,h)) = \sigma^2 + o(1)$ for any u and h under consideration. Hence, for sufficiently large sample sizes T, the test statistic $\widehat{\psi}_T(u,h)/\widehat{\sigma}$ has approximately unit variance.

We now combine the test statistics $\widehat{\psi}_T(u,h)/\widehat{\sigma}$ for a wide range of different locations u and bandwidths or scales h. There are different ways to do so, leading to different types of multiscale statistics. Our multiscale statistic is defined as

$$\widehat{\Psi}_T = \max_{(u,h)\in\mathcal{G}_T} \Big\{ \Big| \frac{\widehat{\psi}_T(u,h)}{\widehat{\sigma}} \Big| - \lambda(h) \Big\}, \tag{1.3.2}$$

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 given 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 calibrate the statistics $\widehat{\psi}_T(u,h)/\widehat{\sigma}$ that correspond to the bandwidth h by subtracting the additive correction term $\lambda(h)$. This approach was pioneered by Dümbgen and Spokoiny (2001) and has been used in numerous other studies since then; see e.g. Dümbgen (2002), Rohde (2008), Dümbgen and Walther (2008), Rufibach and Walther (2010), Schmidt-Hieber et al. (2013) and Eckle et al. (2017).

To see the heuristic idea behind the additive correction $\lambda(h)$, consider for a moment the uncorrected statistic

$$\widehat{\Psi}_{T,\mathrm{uncorrected}} = \max_{(u,h)\in\mathcal{G}_T} \Big| \frac{\widehat{\psi}_T(u,h)}{\widehat{\sigma}} \Big| \tag{1.3.3}$$

and suppose that the hypothesis $H_0(u,h)$ is true for all $(u,h) \in \mathcal{G}_T$. For simplicity,

[1] Alternatively to the local linear weights defined in (1.3.1), we could also work with the weights $w_{t,T}(u,h) = K'(h^{-1}[u-t/T])/\{\sum_{t=1}^{T} K'(h^{-1}[u-t/T])^2\}^{1/2}$, where the kernel function K is assumed to be differentiable and K' is its derivative. We however prefer to use local linear weights as these have superior theoretical properties at the boundary.

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

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

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 $h_{\max} < 1/2$.

According to (C12), 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 compared to the sample size T. For example, we may work with the set

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

which contains more than enough points (u, h) for most practical applications. Condition (C13) 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. The lower bound on $h_{\min}$ depends on the parameter q defined in (C2) which specifies the number of existing moments for the error terms ε_t. As one can see, we can choose $h_{\min}$ to be of the order $T^{-1/2}$ for any $q > 4$. Hence, we can let $h_{\min}$ converge to 0 very quickly even if only the first few moments of the error terms ε_t exist. If all moments

exist (i.e. $q = \infty$), $h_{\min}$ may converge to 0 almost as quickly as $T^{-1} \log T$. Furthermore, the maximal bandwidth $h_{\max}$ is not even required to converge to 0, which implies that we can pick it very large.

Remark 1.3.1. *The above construction of the multiscale statistic can be easily adapted to hypotheses other than H_0. To do so, one simply needs to replace the kernel weights $w_{t,T}(u,h)$ defined in* (1.3.1) *by appropriate versions which are suited to test the hypothesis of interest. For example, if one wants to test for local convexity/concavity of m, one may define the kernel weights $w_{t,T}(u,h)$ such that the kernel average $\widehat{\psi}_T(u,h)$ is a (rescaled) estimator of the second derivative of m at the location u with bandwidth h.*

1.3.2 The test procedure

In order to formulate a test for the null hypothesis H_0, we still need to specify a critical value. To do so, we define the statistic

$$\Phi_T = \max_{(u,h)\in\mathcal{G}_T} \Big\{ \Big| \frac{\phi_T(u,h)}{\sigma} \Big| - \lambda(h) \Big\}, \tag{1.3.4}$$

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 is now defined as follows: For a given significance level $\alpha \in (0,1)$, we reject the overall null hypothesis H_0 if $\widehat{\Psi}_T > q_T(\alpha)$. In particular, for any $(u,h) \in \mathcal{G}_T$, we reject $H_0(u,h)$ if the (corrected) test statistic $|\widehat{\psi}_T(u,h)/\widehat{\sigma}| - \lambda(h)$ lies above the critical value $q_T(\alpha)$, that is, if $|\widehat{\psi}_T(u,h)/\widehat{\sigma}| > q_T(\alpha) + \lambda(h)$.

1.3.3 Theoretical properties of the test

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

$$\widehat{\Phi}_T = \max_{(u,h)\in\mathcal{G}_T} \Big\{ \Big| \frac{\widehat{\phi}_T(u,h)}{\widehat{\sigma}} \Big| - \lambda(h) \Big\} \tag{1.3.5}$$

with $\widehat{\phi}_T(u,h) = \widehat{\psi}_T(u,h) - \mathbb{E}[\widehat{\psi}_T(u,h)] = \sum_{t=1}^T w_{t,T}(u,h)\varepsilon_t$. The following result is central to the theoretical analysis of our multiscale test. According to it, the (known) quantile $q_T(\alpha)$ of the Gaussian statistic Φ_T defined in Section 1.3.2 can be used as a proxy for the $(1-\alpha)$-quantile of the multiscale statistic $\widehat{\Phi}_T$.

Theorem 1.3.1. *Let (C1)–(C13) 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 1.3.1 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 $T \geq 1$ by a statistic $\widetilde{\Phi}_T$ with the same distribution as $\widehat{\Phi}_T$ and the property that

$$\left|\widetilde{\Phi}_T - \Phi_T\right| = o_p(\delta_T), \tag{1.3.6}$$

where $\delta_T = o(1)$ and the Gaussian statistic Φ_T is defined in Section 1.3.2. 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

$$\sup_{x\in\mathbb{R}} \left|\mathbb{P}(\widetilde{\Phi}_T \leq x) - \mathbb{P}(\Phi_T \leq x)\right| = o(1), \tag{1.3.7}$$

which immediately implies the statement of Theorem 1.3.1. Importantly, the convergence result (1.3.6) is not sufficient for establishing (1.3.7). 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 (1.3.7). For (1.3.7) to hold, we additionally require the distribution of Φ_T to have some sort of continuity property. Specifically, we prove that

$$\sup_{x\in\mathbb{R}} \mathbb{P}\big(|\Phi_T - x| \leq \delta_T\big) = o(1), \tag{1.3.8}$$

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 (1.3.8) are anti-concentration results for Gaussian random vectors as derived in Chernozhukov et al. (2015). The claim (1.3.7) can be proven by using (1.3.6) together with (1.3.8), which in turn yields Theorem 1.3.1.

The main idea of our proof strategy is to combine strong approximation theory with anti-concentration bounds for Gaussian random vectors to show that the quantiles of the multiscale statistic $\widehat{\Phi}_T$ can be proxied by those of a Gaussian analogue. This strategy is quite general in nature and may be applied to other multiscale problems for dependent data. Strong approximation theory has also been used to investigate multiscale tests for independent data; see e.g. Schmidt-Hieber et al. (2013). However, it has not been combined with anti-concentration results to approximate the quantiles of the multiscale statistic. As an alternative to strong approximation theory, Eckle et al. (2017) and Proksch et al. (2018) have recently used Gaussian approximation results derived in Chernozhukov et al. (2014, 2017) to analyse multiscale tests for independent data. Even though it might be possible to adapt these techniques to the case of dependent data, this is not trivial at all as part of the technical arguments and the Gaussian approximation tools strongly rely on the assumption of independence.

We now investigate the theoretical properties of our multiscale test with the help of Theorem 1.3.1. The first result is an immediate consequence of Theorem 1.3.1. It says that the test has the correct (asymptotic) size.

Proposition 1.3.1. *Let the conditions of Theorem 1.3.1 be satisfied. Under the null hypothesis H_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

$$m_T'(w) \geq c_T \sqrt{\frac{\log T}{Th^3}} \quad \text{for all } w \in [u-h, u+h], \tag{1.3.9}$$

where $\{c_T\}$ is any sequence of positive numbers with $c_T \to \infty$. Alternatively to (1.3.9), we may also assume that $-m_T'(w) \geq c_T \sqrt{\log T/(Th^3)}$ for all $w \in [u-h, u+h]$.

Proposition 1.3.2. *Let the conditions of Theorem 1.3.1 be satisfied and consider any sequence of functions m_T with the property* (1.3.9)*. Then*

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

According to Proposition 1.3.2, our test has asymptotic power 1 against local alternatives of the form (1.3.9). The proof can be found in the Appendix.

The next result formally shows that we can make simultaneous confidence statements about the time intervals where the trend m is increasing/decreasing. To formulate it, we define

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

where

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

The object $\Pi_T^{\pm}$ can be interpreted as follows: Our multiscale test rejects the null hypothesis $H_0(u,h)$ if $|\widehat{\psi}_T(u,h)/\widehat{\sigma}| > q_T(\alpha) + \lambda(h)$. Put differently, it rejects $H_0(u,h)$ for all $(u,h) \in \mathcal{A}_T^{\pm}$. Hence, $\Pi_T^{\pm}$ is the collection of time intervals $I_{u,h} = [u-h, u+h]$ for which our test rejects $H_0(u,h)$. The objects Π_T^{+} and Π_T^{-} can be interpreted analogously: If $\widehat{\psi}_T(u,h)/\widehat{\sigma} > q_T(\alpha) + \lambda(h)$, that is, if $(u,h) \in \mathcal{A}_T^{+}$, then our test rejects $H_0(u,h)$ and indicates an increase in the trend m on the interval $I_{u,h}$, taking into account the positive

sign of the statistic $\widehat{\psi}_T(u,h)/\widehat{\sigma}$. Hence, Π_T^+ is the collection of time intervals $I_{u,h}$ for which our test indicates an increase in the trend m. Likewise, Π_T^- is the collection of intervals for which the test indicates a decrease. Note that $\Pi_T^\pm$ (as well as Π_T^+ and Π_T^-) is a random collection of intervals: Whether our test rejects $H_0(u,h)$ for some (u,h) depends on the realization of the random vector $(Y_{1,T},\dots,Y_{T,T})$. Hence, whether an interval $I_{u,h}$ belongs to $\Pi_T^\pm$ depends on this realization as well. Having defined the objects $\Pi_T^\pm$, Π_T^+ and Π_T^-, we now consider the events

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

$E_T^\pm$ (E_T^+, E_T^-) is the event that the function m is non-constant (increasing, decreasing) on all intervals $I_{u,h} \in \Pi_T^\pm$ (Π_T^+, Π_T^-). More precisely, $E_T^\pm$ (E_T^+, E_T^-) is the event that for each interval $I_{u,h} \in \Pi_T^\pm$ (Π_T^+, Π_T^-), there is a subset $J_{u,h} \subseteq I_{u,h}$ with m being a non-constant (increasing, decreasing) function on $J_{u,h}$. We can make the following formal statement about the events $E_T^\pm$, E_T^+ and E_T^-, whose proof is given in the Appendix.

Proposition 1.3.3. *Let the conditions of Theorem 1.3.1 be fulfilled. Then for $\ell \in \{\pm,+,-\}$, it holds that*

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

According to Proposition 1.3.3, we can make simultaneous confidence statements of the following form: With (asymptotic) probability $\geq (1-\alpha)$, the trend function m is non-constant (increasing, decreasing) on each interval $I_{u,h} \in \Pi_T^\pm$ (Π_T^+, Π_T^-). Hence, our multiscale procedure allows to identify, with a pre-specified confidence, time intervals where there is an increase/decrease in the trend m.

Remark 1.3.2. *Unlike $\Pi_T^\pm$, the sets Π_T^+ and Π_T^- only contain 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 $I_{u,h} \not\subseteq [0,1]$. The reason is as follows: Let $(u,h) \in \mathcal{A}_T^+$ with $I_{u,h} \not\subseteq [0,1]$. 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. Crudely speaking, the problem is that the local linear weights $w_{t,T}(u,h)$ behave quite differently at boundary points (u,h) with $I_{u,h} \not\subseteq [0,1]$. As a consequence, we can include boundary points (u,h) in $\Pi_T^\pm$ but not in Π_T^+ and Π_T^-.*

Remark 1.3.3. *The statement of Proposition 1.3.3 suggests to graphically present the results of our multiscale test by plotting the intervals $I_{u,h} \in \Pi_T^\ell$ for $\ell \in \{\pm,+,-\}$, 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^ℓ is often quite large. To obtain a better graphical summary of the*

results, we replace Π_T^ℓ *by a subset* $\Pi_T^{\ell,\min}$ *which is constructed as follows: As in Dümbgen (2002), we call an interval* $I_{u,h} \in \Pi_T^\ell$ *minimal if there is no other interval* $I_{u',h'} \in \Pi_T^\ell$ *with* $I_{u',h'} \subset I_{u,h}$*. Let* $\Pi_T^{\ell,\min}$ *be the set of all minimal intervals in* Π_T^ℓ *for* $\ell \in \{\pm, +, -\}$ *and define the events*

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

It is easily seen that $E_T^\ell = E_T^{\ell,\min}$ *for* $\ell \in \{\pm, +, -\}$*. Hence, by Proposition 1.3.3, it holds that*

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

for $\ell \in \{\pm, +, -\}$*. This suggests to plot the minimal intervals in* $\Pi_T^{\ell,\min}$ *rather than the whole collection of intervals* Π_T^ℓ *as a graphical summary of the test results. We in particular use this way of presenting the test results in our application in Section 1.6.*

Proposition 1.3.3 allows to make confidence statements for a fixed significance level $\alpha \in (0,1)$. In some situations, one may be interested in letting $\alpha = \alpha_T \in (0,1)$ tend to zero as $T \to \infty$. This situation is considered in the following corollary to Proposition 1.3.3, whose proof can be found in the Appendix.

Corollary 1.3.1. *Let the conditions of Theorem 1.3.1 be fulfilled and let* $\alpha = \alpha_T \in (0,1)$ *go to zero as* $T \to \infty$*. Then* $\mathbb{P}(E_T^\ell) \to 1$ *for* $\ell \in \{\pm, +, -\}$*.*

Corollary 1.3.1 can be interpreted as a consistency result: If we let the significance level $\alpha = \alpha_T$ go to zero, then the event $E_T^\pm$ (E_T^+, E_T^-) occurs with probability tending to 1, that is, the trend m is non-constant (increasing, decreasing) on each interval $I_{u,h} \in \Pi_T^\pm$ (Π_T^+, Π_T^-) with probability tending to 1.

1.3.4 Comparison to SiZer methods

As already mentioned in the Introduction, some SiZer methods for dependent data have been introduced in Park et al. (2004) and Rondonotti et al. (2007), which we refer to as dependent SiZer for short. Informally speaking, both our approach and dependent SiZer are methods to test for local increases/decreases of a nonparametric trend function m. The formal problem is to test the hypothesis $H_0(u,h)$ simultaneously for all $(u,h) \in \mathcal{G}_T$, where in this section, we let $\mathcal{G}_T = U_T \times H_T$ with U_T being the set of locations and H_T the set of bandwidths or scales. In what follows, we compare our approach to dependent SiZer and point out the most important differences.

Dependent SiZer is based on the statistics $\widehat{s}_T(u,h) = \widehat{m}'(u,h)/\widehat{\mathrm{sd}}(\widehat{m}'(u,h))$, where $\widehat{m}'(u,h)$ is a local linear kernel estimator of $m'(u)$ with bandwidth h and $\widehat{\mathrm{sd}}(\widehat{m}'(u,h))$ is an estimator of its standard deviation. The statistic $\widehat{s}_T(u,h)$ parallels the statistic

$\widehat{\psi}_T(u,h)/\widehat{\sigma}$ in our approach. In particular, both can be regarded as test statistics of the hypothesis $H_0(u,h)$. There are two versions of dependent SiZer:

(a) The global version aggregates the individual statistics $\widehat{s}_T(u,h)$ into the overall statistic $\widehat{S}_T = \max_{h\in H_T} \widehat{S}_T(h)$, where $\widehat{S}_T(h) = \max_{u\in U_T} |\widehat{s}_T(u,h)|$. The statistic $\widehat{S}_T$ is the counterpart to the multiscale statistic $\widehat{\Psi}_T$ in our approach.

(b) The row-wise version considers each scale $h \in H_T$ separately. In particular, for each bandwidth $h \in H_T$, a test is carried out based on the statistic $\widehat{S}_T(h)$. A row-wise analogue of our approach would be obtained by carrying out a test for each scale $h \in H_T$ separately based on the statistic $\widehat{\Psi}_T(h) = \max_{u\in U_T} |\widehat{\psi}_T(u,h)/\widehat{\sigma}|$.[2]

In practice, SiZer is commonly implemented in its row-wise form. The main reason is that it has more power than the global version by construction. However, this gain of power comes at a cost: Row-wise SiZer carries out a test *separately* for each scale $h \in H_T$, thus ignoring the simultaneous test problem across scales h. Hence, it is not a rigorous level-α-test of the null H_0. For this reason, we focus on global SiZer in the rest of this section.

Even though related, our methods and theory are markedly different from those of the SiZer approach:

(i) Theory for SiZer is derived under the assumption that $H_T \subseteq H$ for all T, where H is a compact subset of $(0,\infty)$. As already pointed out in Chaudhuri and Marron (2000) on p.420, this is a quite severe restriction: Only bandwidths h are taken into account that remain bounded away from zero as the sample size T increases. Bandwidths h that converge to zero are excluded. Our theory, in contrast, allows to simultaneously consider bandwidths h of fixed size and bandwidths h that converge to zero at different rates. To achieve this, we come up with a proof strategy which is very different from that in the SiZer literature: As proven in Chaudhuri and Marron (2000) for the i.i.d. case and in Park et al. (2009a) for the dependent data case, $\widehat{S}_T$ weakly converges to some limit S under the overall null hypothesis H_0. This is the central technical result on which the theoretical properties of SiZer are based. In contrast to this, our proof strategy (which combines strong approximation theory with anti-concentration bounds as outlined in Section 1.3.3) does not even require the statistic $\widehat{\Psi}_T$ to have a weak limit and is thus not restricted by the limitations of classic weak convergence theory.

(ii) There are different ways to combine the test statistics $\widehat{S}_T(h) = \max_{u\in U_T} |\widehat{s}_T(u,h)|$ for different scales $h \in H_T$. One way is to take their maximum, which leads to the SiZer statistic $\widehat{S}_T = \max_{h\in H_T} \widehat{S}_T(h)$. We could proceed analogously and consider the statistic $\widehat{\Psi}_{T,\text{uncorrected}} = \max_{h\in H_T} \widehat{\Psi}_T(h) = \max_{(u,h)\in U_T\times H_T} |\widehat{\psi}_T(u,h)/\widehat{\sigma}|$. However, as argued in Dümbgen and Spokoiny (2001) and as discussed in Section 1.3.1, this

[2]Note that we can drop the correction term $\lambda(h)$ in this case as it is a fixed constant if only a single bandwidth h is taken into account.

aggregation scheme is not optimal when the set H_T contains scales h of many different rates. Following the lead of Dümbgen and Spokoiny (2001), we consider the test statistic $\widehat{\Psi}_T = \max_{(u,h)\in U_T\times H_T}\{|\widehat{\psi}_T(u,h)/\widehat{\sigma}| - \lambda(h)\}$ with the additive correction terms $\lambda(h)$. Hence, even though related, our multiscale test statistic $\widehat{\Psi}_T$ differs from the SiZer statistic $\widehat{S}_T$ in important ways.

(iii) The main complication in carrying out both our multiscale test and SiZer is to determine the critical values, that is, the quantiles of the test statistics $\widehat{\Psi}_T$ and $\widehat{S}_T$ under H_0. In order to approximate the quantiles, we proceed quite differently than in the SiZer literature. The quantiles of the SiZer statistic $\widehat{S}_T$ can be approximated by those of the weak limit S. Usually, however, the quantiles of S cannot be determined analytically but have to be approximated themselves (e.g. by the bootstrap procedures of Chaudhuri and Marron (1999, 2000)). Alternatively, the quantiles of $\widehat{S}_T$ can be approximated by procedures based on extreme value theory (as proposed in Hannig and Marron (2006) and Park et al. (2009a)). In our approach, the quantiles of $\widehat{\Psi}_T$ under H_0 are approximated by those of a suitably constructed Gaussian analogue of $\widehat{\Psi}_T$. It is far from obvious that this Gaussian approximation is valid when the data are dependent. To see this, deep strong approximation theory for dependent data (as derived in Berkes et al. (2014)) is needed. It is important to note that our Gaussian approximation procedure is not the same as the bootstrap procedures proposed in Chaudhuri and Marron (1999, 2000). Both procedures can of course be regarded as resampling methods. However, the resampling is done in a quite different way in our case.

1.4 Estimation of the long-run error variance

In this section, we discuss how to estimate the long-run variance $\sigma^2 = \sum_{\ell=-\infty}^{\infty}\mathrm{Cov}(\varepsilon_0,\varepsilon_\ell)$ of the error terms in model (1.2.1). There are two broad classes of estimators: residual- and difference-based estimators. In residual-based approaches, σ^2 is estimated from the residuals $\widehat{\varepsilon}_t = Y_{t,T} - \widehat{m}_h(t/T)$, where $\widehat{m}_h$ is a nonparametric estimator of m with the bandwidth or smoothing parameter h. Difference-based methods proceed by estimating σ^2 from the ℓ-th differences $Y_{t,T} - Y_{t-\ell,T}$ of the observed time series $\{Y_{t,T}\}$ for certain orders ℓ. In what follows, we focus attention on difference-based methods as these do not involve a nonparametric estimator of the function m and thus do not require to specify a bandwidth h for the estimation of m.

So far, we have assumed that $\{\varepsilon_t\}$ is a general stationary error process which fulfills the weak dependence conditions (C3). Estimating the long-run error variance σ^2 in model (1.2.1) under general weak dependence conditions is a notoriously difficult problem. Estimators of σ^2 often tend to be quite imprecise. To circumvent this issue in practice, it may be beneficial to impose a time series model on the error process $\{\varepsilon_t\}$. Estimating σ^2 under the restrictions of such a model may of course create some misspecification bias.

However, as long as the model gives a reasonable approximation to the true error process, the produced estimates of σ^2 can be expected to be fairly reliable even though they are a bit biased.

Estimators of the long-run error variance σ^2 in model (1.2.1) have been developed for different kinds of error models. A number of authors have analysed the case of MA(m) or, more generally, m-dependent error terms. Difference-based estimators of σ^2 for this case were proposed in Müller and Stadtmüller (1988), Herrmann et al. (1992) and Tecuapetla-Gómez and Munk (2017) among others. Presumably the most widely used error model in practice is an AR(p) process. Residual-based methods to estimate σ^2 in model (1.2.1) with AR(p) errors can be found for example in Truong (1991), Shao and Yang (2011) and Qiu et al. (2013). A difference-based method was proposed in Hall and Van Keilegom (2003).

We consider the class of AR(∞) processes as an error model, which is a quite large and important subclass of linear time series processes. Formally speaking, we let $\{\varepsilon_t\}$ be a process of the form

$$\varepsilon_t = \sum_{j=1}^{\infty} a_j \varepsilon_{t-j} + \eta_t, \tag{1.4.1}$$

where $a_1, a_2, a_3, \ldots$ are unknown coefficients and η_t are i.i.d. with $\mathbb{E}[\eta_t] = 0$ and $\mathbb{E}[\eta_t^2] = \nu^2$. We assume that $A(z) := 1 - \sum_{j=1}^{\infty} a_j z^j \neq 0$ for all complex numbers $|z| \leq 1 + \delta$ with some small $\delta > 0$, which has the following implications: (i) $\{\varepsilon_t\}$ is stationary and causal. (ii) The coefficients a_j decay to zero exponentially fast, that is, $|a_j| \leq C\xi^j$ with some $C > 0$ and $\xi \in (0, 1)$. (iii) $\{\varepsilon_t\}$ has an MA(∞) representation of the form $\varepsilon_t = \sum_{k=0}^{\infty} c_k \eta_{t-k}$. The coefficients c_k can be computed iteratively from the equations

$$c_k - \sum_{j=1}^{k} a_j c_{k-j} = b_k \tag{1.4.2}$$

for $k = 0, 1, 2, \ldots$, where $b_0 = 1$ and $b_k = 0$ for $k > 0$. Moreover, they decay to zero exponentially fast, that is, $|c_k| \leq C\xi^k$ with some $C > 0$ and $\xi \in (0, 1)$. Notably, the error model (1.4.1) nests AR(p^*) processes of any finite order p^* as a special case: If $a_{p^*} \neq 0$ and $a_j = 0$ for all $j > p^*$, then $\{\varepsilon_t\}$ is an AR process of order p^*. In the sequel, we let $p^* \in \mathbb{N} \cup \{\infty\}$ denote the true AR order of $\{\varepsilon_t\}$ which may be finite or infinite. We can thus rewrite (1.4.1) as

$$\varepsilon_t = \sum_{j=1}^{p^*} a_j \varepsilon_{t-j} + \eta_t, \tag{1.4.3}$$

where the AR order p^* is treated as unknown.

We now construct a difference-based estimator of σ^2 for the case that $\{\varepsilon_t\}$ is an AR(p^*) process of the form (1.4.3). To do so, we will fit AR(p) type models to $\{\varepsilon_t\}$, where we distinguish between the following two cases:

(A) We do not know the precise AR order p^* but we know an upper bound p on it. In

this case, p is a fixed natural number with $p \geq p^*$.

(B) We neither know p^* nor an upper bound on it. In this case, we let $p = p_T \to \infty$ as $T \to \infty$, where formal conditions on the growth of $p = p_T$ are specified later on.

To simplify notation, we let $\Delta_\ell Z_t = Z_t - Z_{t-\ell}$ denote the ℓ-th differences of a general time series $\{Z_t\}$. Our estimation method relies on the following simple observation: If $\{\varepsilon_t\}$ is an AR(p^*) process of the form (1.4.3), then the time series $\{\Delta_q \varepsilon_t\}$ of the differences $\Delta_q \varepsilon_t = \varepsilon_t - \varepsilon_{t-q}$ is an ARMA(p^*, q) process of the form

$$\Delta_q \varepsilon_t - \sum_{j=1}^{p^*} a_j \Delta_q \varepsilon_{t-j} = \eta_t - \eta_{t-q}. \tag{1.4.4}$$

As m is Lipschitz, the differences $\Delta_q \varepsilon_t$ of the unobserved error process are close to the differences $\Delta_q Y_{t,T}$ of the observed time series in the sense that

$$\Delta_q Y_{t,T} = \big[\varepsilon_t - \varepsilon_{t-q}\big] + \Big[m\Big(\frac{t}{T}\Big) - m\Big(\frac{t-q}{T}\Big)\Big] = \Delta_q \varepsilon_t + O\Big(\frac{q}{T}\Big). \tag{1.4.5}$$

Taken together, (1.4.4) and (1.4.5) imply that the differenced time series $\{\Delta_q Y_{t,T}\}$ is approximately an ARMA(p^*, q) process of the form (1.4.4). It is precisely this point which is exploited by our estimation method.

We first describe our procedure to estimate the AR parameters a_j. For any $q \geq 1$, the ARMA(p^*, q) process $\{\Delta_q \varepsilon_t\}$ satisfies the Yule-Walker equations

$$\gamma_q(\ell) - \sum_{j=1}^{p^*} a_j \gamma_q(\ell - j) = \begin{cases} -\nu^2 c_{q-\ell} & \text{for } 1 \leq \ell < q+1 \\ 0 & \text{for } \ell \geq q+1, \end{cases} \tag{1.4.6}$$

where $\gamma_q(\ell) = \mathrm{Cov}(\Delta_q \varepsilon_t,\, \Delta_q \varepsilon_{t-\ell})$ and c_k are the coefficients from the MA(∞) expansion of $\{\varepsilon_t\}$. Combining the equations (1.4.6) for $\ell = 1, \ldots, p$, we get that

$$\boldsymbol{\Gamma}_q \boldsymbol{a} = \boldsymbol{\gamma}_q + \nu^2 \boldsymbol{c}_q - \boldsymbol{\rho}_q, \tag{1.4.7}$$

where $\boldsymbol{a} = (a_1, \ldots, a_p)^\top$, $\boldsymbol{\gamma}_q = (\gamma_q(1), \ldots, \gamma_q(p))^\top$ and $\boldsymbol{\Gamma}_q$ denotes the $p \times p$ covariance matrix $\boldsymbol{\Gamma}_q = (\gamma_q(i-j) : 1 \leq i, j \leq p)$. Moreover, $\boldsymbol{c}_q = (c_{q-1}, \ldots, c_{q-p})^\top$ and $\boldsymbol{\rho}_q = (\rho_q(1), \ldots, \rho_q(p))^\top$ with $\rho_q(\ell) = \sum_{j=p+1}^{p^*} a_j \gamma_q(\ell - j)$. Since the AR coefficients a_j as well as the MA coefficients c_k decay exponentially fast to zero, $\boldsymbol{\rho}_q \approx \boldsymbol{0}$ and $\boldsymbol{c}_q \approx \boldsymbol{0}$ for large values of q, implying that $\boldsymbol{\Gamma}_q \boldsymbol{a} \approx \boldsymbol{\gamma}_q$. This suggests to estimate $\boldsymbol{a}$ by

$$\widetilde{\boldsymbol{a}}_q = \widehat{\boldsymbol{\Gamma}}_q^{-1} \widehat{\boldsymbol{\gamma}}_q, \tag{1.4.8}$$

where $\widehat{\boldsymbol{\Gamma}}_q$ and $\widehat{\boldsymbol{\gamma}}_q$ are defined analogously as $\boldsymbol{\Gamma}_q$ and $\boldsymbol{\gamma}_q$ with $\gamma_q(\ell)$ replaced by the sample autocovariances $\widehat{\gamma}_q(\ell) = (T-q)^{-1} \sum_{t=q+\ell+1}^{T} \Delta_q Y_{t,T} \Delta_q Y_{t-\ell,T}$ and $q = q_T$ goes to infinity as $T \to \infty$. For our theory to work, we require that $q/p \to \infty$, that is, q needs to grow faster than p. Formal conditions on the growth of q are given later on.

The estimator $\widetilde{\boldsymbol{a}}_q$ depends on the tuning parameter q, that is, on the order of the differences $\Delta_q Y_{t,T}$. An appropriate choice of q needs to take care of the following two points: (i) q should be chosen large enough to ensure that the vector $\boldsymbol{c}_q = (c_{q-1}, \dots, c_{q-p})^\top$ is close to zero. As we have already seen, the constants c_k decay to zero exponentially fast and can be computed from the recursive equations (1.4.2) for given parameters $a_1, a_2, a_3, \dots$ In the special case of an AR(1) process, for example, one can readily calculate that $c_k \leq 0.0035$ for any $k \geq 20$ and any $|a_1| \leq 0.75$. Hence, if we have an AR(1) model for the errors ε_t and the error process is not too persistent, choosing $q \geq 20$ should make sure that $\boldsymbol{c}_q$ is close to zero. Generally speaking, the recursive equations (1.4.2) can be used to get some idea for which values of q the vector $\boldsymbol{c}_q$ can be expected to be approximately zero. (ii) q should not be chosen too large in order to ensure that the trend m is appropriately eliminated by taking q-th differences. As long as the trend m is not very strong, the two requirements (i) and (ii) can be fulfilled without much difficulty. For example, by choosing $q = 20$ in the AR(1) case just discussed, we do not only take care of (i) but also make sure that moderate trends m are differenced out appropriately.

When the trend m is very pronounced, in contrast, even moderate values of q may be too large to eliminate the trend appropriately. As a result, the estimator $\widetilde{\boldsymbol{a}}_q$ will have a strong bias. In order to reduce this bias, we refine our estimation procedure as follows: By solving the recursive equations (1.4.2) with $\boldsymbol{a}$ replaced by $\widetilde{\boldsymbol{a}}_q$, we can compute estimators $\widetilde{c}_k$ of the coefficients c_k and thus estimators $\widetilde{\boldsymbol{c}}_r$ of the vectors $\boldsymbol{c}_r$ for any $r \geq 1$. Moreover, the innovation variance ν^2 can be estimated by $\widetilde{\nu}^2 = (2T)^{-1} \sum_{t=p+2}^{T} \widetilde{r}_{t,T}^2$, where $\widetilde{r}_{t,T} = \Delta_1 Y_{t,T} - \sum_{j=1}^{p} \widetilde{a}_j \Delta_1 Y_{t-j,T}$ and $\widetilde{a}_j$ is the j-th entry of the vector $\widetilde{\boldsymbol{a}}_q$. Plugging the expressions $\widehat{\boldsymbol{\Gamma}}_r$, $\widehat{\boldsymbol{\gamma}}_r$, $\widetilde{\boldsymbol{c}}_r$ and $\widetilde{\nu}^2$ into (1.4.7), we can estimate $\boldsymbol{a}$ by

$$\widehat{\boldsymbol{a}}_r = \widehat{\boldsymbol{\Gamma}}_r^{-1} (\widehat{\boldsymbol{\gamma}}_r + \widetilde{\nu}^2 \widetilde{\boldsymbol{c}}_r), \tag{1.4.9}$$

where r is a much smaller differencing order than q. Specifically, in case (A), we can choose r to be any fixed number $r \geq 1$. Unlike q, the parameter r thus remains bounded as T increases. In case (B), our theory allows to choose any number r with $r \geq (1+\delta)p$ for some small $\delta > 0$. Since $q/p \to \infty$, it holds that $q/r \to \infty$ as well, which means that r is of smaller order than q. Hence, in both cases (A) and (B), the estimator $\widehat{\boldsymbol{a}}_r$ is based on a differencing order r that is much smaller than q; only the pilot estimator $\widetilde{\boldsymbol{a}}_q$ relies on differences of the larger order q. As a consequence, $\widehat{\boldsymbol{a}}_r$ should eliminate the trend m more appropriately and should thus be less biased than the pilot estimator $\widetilde{\boldsymbol{a}}_q$. In order to make the method more robust against estimation errors in $\widetilde{\boldsymbol{c}}_r$, we finally average the estimators $\widehat{\boldsymbol{a}}_r$ for a few values of r. In particular, we define

$$\widehat{\boldsymbol{a}} = \frac{1}{\overline{r} - \underline{r} + 1} \sum_{r=\underline{r}}^{\overline{r}} \widehat{\boldsymbol{a}}_r, \tag{1.4.10}$$

where $\underline{r}$ and $\overline{r}$ are chosen as follows: In case (A), we let $\underline{r}$ and $\overline{r}$ be small natural numbers. In case (B), we set $\underline{r} = (1-\delta)p$ for some small $\delta > 0$ and choose $\overline{r}$ such that $\overline{r} - \underline{r}$ remains

bounded. For ease of notation, we suppress the dependence of $\widehat{\boldsymbol{a}}$ on the parameters $\underline{r}$ and $\overline{r}$. Once $\widehat{\boldsymbol{a}} = (\widehat{a}_1, \ldots, \widehat{a}_p)^\top$ is computed, the long-run variance σ^2 can be estimated by

$$\widehat{\sigma}^2 = \frac{\widehat{\nu}^2}{(1 - \sum_{j=1}^p \widehat{a}_j)^2}, \tag{1.4.11}$$

where $\widehat{\nu}^2 = (2T)^{-1} \sum_{t=p+2}^T \widehat{r}_{t,T}^2$ with $\widehat{r}_{t,T} = \Delta_1 Y_{t,T} - \sum_{j=1}^p \widehat{a}_j \Delta_1 Y_{t-j,T}$ is an estimator of the innovation variance ν^2 and we make use of the fact that $\sigma^2 = \nu^2/(1 - \sum_{j=1}^{p^*} a_j)^2$ for the AR(p^*) process $\{\varepsilon_t\}$.

We briefly compare the estimator $\widehat{\boldsymbol{a}}$ to competing methods. Presumably closest to our approach is that of Hall and Van Keilegom (2003) which is designed for AR(p^*) processes of known finite order p^*. For comparing the two methods, we thus assume p^* to be known and set $p = p^*$. The two main advantages of our method are as follows:

(a) Our estimator produces accurate estimation results even when the AR process $\{\varepsilon_t\}$ is quite persistent, that is, even when the AR polynomial $A(z) = 1 - \sum_{j=1}^{p^*} a_j z^j$ has a root close to the unit circle. The estimator of Hall and Van Keilegom (2003), in contrast, may have very high variance and may thus produce unreliable results when the AR polynomial $A(z)$ is close to having a unit root. This difference in behaviour can be explained as follows: Our pilot estimator $\widetilde{\boldsymbol{a}}_q = (\widetilde{a}_1, \ldots, \widetilde{a}_{p^*})^\top$ has the property that the estimated AR polynomial $\widetilde{A}(z) = 1 - \sum_{j=1}^{p^*} \widetilde{a}_j z^j$ has no root inside the unit disc, that is, $\widetilde{A}(z) \neq 0$ for all complex numbers z with $|z| \leq 1$.[3] Hence, the fitted AR model with the coefficients $\widetilde{\boldsymbol{a}}_q$ is ensured to be stationary and causal. Even though this may seem to be a minor technical detail, it has a huge effect on the performance of the estimator $\widetilde{\boldsymbol{a}}_q$: It keeps the estimator stable even when the AR process is very persistent and the AR polynomial $A(z)$ has almost a unit root. This in turn results in a reliable behaviour of the estimator $\widehat{\boldsymbol{a}}$ in the case of high persistence. The estimator of Hall and Van Keilegom (2003), in contrast, may produce non-causal results when the AR polynomial $A(z)$ is close to having a unit root. As a consequence, it may have unnecessarily high variance in the case of high persistence. We illustrate this difference between the estimators by the simulation exercises in Section 1.5.2. A striking example is Figure 1.6, which presents the simulation results for the case of an AR(1) process $\varepsilon_t = a_1 \varepsilon_{t-1} + \eta_t$ with $a_1 = -0.95$ and clearly shows the much better performance of our method.

(b) Both our pilot estimator $\widetilde{\boldsymbol{a}}_q$ and the estimator of Hall and Van Keilegom (2003) tend to have a substantial bias when the trend m is pronounced. Our estimator $\widehat{\boldsymbol{a}}$ reduces this bias considerably as demonstrated in the simulations of Section 1.5.2. Unlike the estimator of Hall and Van Keilegom (2003), it thus produces accurate results even in the presence of a very strong trend.

[3] More precisely, $\widetilde{A}(z) \neq 0$ for all z with $|z| \leq 1$, whenever the covariance matrix $(\widehat{\gamma}_q(i-j) : 1 \leq i, j \leq p^*+1)$ is non-singular. Moreover, $(\widehat{\gamma}_q(i-j) : 1 \leq i, j \leq p^*+1)$ is non-singular whenever $\widehat{\gamma}_q(0) > 0$, which is the generic case.

We close this section by deriving some basic asymptotic properties of the estimators $\widetilde{\boldsymbol{a}}_q$, $\widehat{\boldsymbol{a}}$ and $\widehat{\sigma}^2$. To formulate the following result, we use the shorthand $v_T \ll w_T$ which means that $v_T/w_T \to 0$ as $T \to \infty$.

Proposition 1.4.1. *Let m be Lipschitz continuous and suppose that $\{\varepsilon_t\}$ is an AR(p^*) process of the form* (1.4.3) *with the following properties: $A(z) \neq 0$ for all $|z| \leq 1+\delta$ with some small $\delta > 0$ and the innovations η_t have a finite fourth moment. Assume that p, q, $\underline{r}$ and $\overline{r}$ satisfy the following conditions: In case (A), p, $\underline{r}$ and $\overline{r}$ are fixed natural numbers and $\log T \ll q \ll \sqrt{T}$. In case (B), $C \log T \leq p \ll \min\{T^{1/5}, q\}$ for some sufficiently large C, $q \ll \sqrt{T}$, $\underline{r} = (1+\delta)p$ for some small $\delta > 0$ and $\overline{r} - \underline{r}$ remains bounded. Under these conditions, $\widetilde{\boldsymbol{a}}_q - \boldsymbol{a} = O_p(\sqrt{p/T})$ as well as $\widehat{\boldsymbol{a}} - \boldsymbol{a} = O_p(\sqrt{p^3/T})$ and $\widehat{\sigma}^2 - \sigma^2 = O_p(\sqrt{p^4/T})$.*

The proof is provided in the Appendix. As one can see, the convergence rate of the second-step estimator $\widehat{\boldsymbol{a}}$ is somewhat slower than that of the pilot estimator $\widetilde{\boldsymbol{a}}_q$. Hence, from an asymptotic perspective, there is no gain from using the second-step estimator. Nevertheless, in finite samples, the estimator $\widehat{\boldsymbol{a}}$ vastly outperforms $\widetilde{\boldsymbol{a}}_q$ since it considerably reduces the bias of the latter.

1.5 Simulations

1.5.1 Small sample properties of the multiscale test

In what follows, we investigate the performance of our multiscale test and compare it to the dependent SiZer methods from Park et al. (2004), Rondonotti et al. (2007) and Park et al. (2009a). We consider the following versions of our multiscale test and SiZer:

$\mathcal{T}_{\mathrm{MS}}$: our multiscale test with the statistic $\widehat{\Psi}_T = \max_{h \in H_T}\{\widehat{\Psi}_T(h) - \lambda(h)\}$, where $\widehat{\Psi}_T(h) = \max_{u \in U_T} |\widehat{\psi}_T(u,h)/\widehat{\sigma}|$. Here and in what follows, we write $\mathcal{G}_T = U_T \times H_T$, where U_T is the set of locations and H_T the set of bandwidths.

$\mathcal{T}_{\mathrm{UC}}$: the uncorrected version of our multiscale test with the test statistic $\widehat{\Psi}_{T,\mathrm{uncorrected}} = \max_{h \in H_T} \widehat{\Psi}_T(h)$, which was already introduced in (1.3.3). The uncorrected test is carried out in exactly the same way as $\mathcal{T}_{\mathrm{MS}}$. The only difference is that the correction terms $\lambda(h)$ are removed.

$\mathcal{T}_{\mathrm{RW}}$: a row-wise (or scale-wise) version of our multiscale test as briefly mentioned in Section 1.3.4. This version carries out a test scale-wise, that is, separately for each scale $h \in H_T$ based on the statistic $\widehat{\Psi}_T(h)$. Note: (i) For each $h \in H_T$, the test based on $\widehat{\Psi}_T(h)$ can be performed in the same way as the multiscale test $\mathcal{T}_{\mathrm{MS}}$, since it is a degenerate version of the latter with the set of scales H_T replaced by the singleton $\{h\}$. (ii) It does not matter whether we correct the statistic $\widehat{\Psi}_T(h)$ by subtracting $\lambda(h)$ or not, since $\lambda(h)$ acts as a fixed constant when only one bandwidth h is taken into account.

$\mathcal{T}_{\text{SiZer}}$: the row-wise version of dependent SiZer from Park et al. (2004), Rondonotti et al. (2007) and Park et al. (2009a). We do not consider a global version of dependent SiZer as such a version was not fully developed in the aforementioned papers.

The simulation setup is as follows: We generate data from the model $Y_{t,T} = m(t/T)+\varepsilon_t$ for different trends m, error processes $\{\varepsilon_t\}$ and sample sizes T. The error terms are supposed to have the AR(1) structure $\varepsilon_t = a_1\varepsilon_{t-1} + \eta_t$, where $a_1 \in \{-0.9, -0.5, -0.25, 0.25, 0.5, 0.9\}$, η_t are i.i.d. standard normal and the AR order $p^* = 1$ is treated as known. To simulate data under the null, 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 different non-constant trend functions which are specified below. For each model specification, we simulate $S = 1000$ data samples and carry out the tests $\mathcal{T}_{\text{MS}}$, $\mathcal{T}_{\text{UC}}$, $\mathcal{T}_{\text{RW}}$ and $\mathcal{T}_{\text{SiZer}}$ for each simulated sample.

To implement our multiscale test $\mathcal{T}_{\text{MS}}$, we choose K to be an Epanechnikov kernel and let $\mathcal{G}_T = U_T \times H_T$ with

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

We thus take into account all locations u on an equidistant grid U_T with step length $5/T$ and all bandwidths $h = 5/T, 10/T, 15/T, \ldots$ with $\log T/T \leq h \leq 1/4$. Note that the lower bound $\log T/T$ is motivated by (C13) which requires that $\log T/T \ll h_{\min}$ (given that all moments of ε_t exist). As a robustness check, we have re-run the simulations for a number of other grids. As the results are very similar, we do however not report them here. To estimate the long-run error variance σ^2, we apply the procedure from Section 1.4 with $\underline{r} = 1$, $\overline{r} = 10$ and the following choices of q: For $a_1 \in \{-0.5, -0.25, 0.25, 0.5\}$, we set $q = 25$. As already discussed in Section 1.4, this should be an appropriate choice for AR(1) errors that are not too strongly correlated, in particular, for $a_1 \in \{-0.5, -0.25, 0.25, 0.5\}$. When the errors are very strongly correlated, larger values of q are required to produce precise estimates of σ^2. In the case of AR(1) errors with $a_1 \in \{-0.9, 0.9\}$, we thus set $q = 50$. The dependence of our long-run variance estimator on the tuning parameters q, $\underline{r}$ and $\overline{r}$ is explored more systematically in Section 1.5.2. To compute the critical values of the multiscale test $\mathcal{T}_{\text{MS}}$, we simulate 5000 values of the statistic Φ_T defined in Section 1.3.2 and compute their empirical $(1-\alpha)$ quantile $q_T(\alpha)$. The uncorrected and row-wise versions $\mathcal{T}_{\text{UC}}$ and $\mathcal{T}_{\text{RW}}$ of our multiscale test are implemented analogously. The SiZer test is implemented as described in Park et al. (2009a). The details are summarized in Section 1.C.1 of the Appendix.

1.5.1.1 Size simulations

The first part of our simulation study investigates the size properties of the four tests $\mathcal{T}_{\text{MS}}$, $\mathcal{T}_{\text{UC}}$, $\mathcal{T}_{\text{RW}}$ and $\mathcal{T}_{\text{SiZer}}$ under the null that the trend m is constant. To start with, we focus

Table 1.1: Size of $\mathcal{T}_{\mathrm{MS}}$ for the AR parameters $a_1 \in \{-0.5, -0.25, 0.25, 0.5\}$.

	$a_1 = -0.5$			$a_1 = -0.25$			$a_1 = 0.25$			$a_1 = 0.5$		
	nominal size α			nominal size α			nominal size α			nominal size α		
	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
$T = 250$	0.013	0.040	0.086	0.016	0.054	0.106	0.009	0.045	0.094	0.014	0.058	0.106
$T = 500$	0.013	0.044	0.102	0.008	0.041	0.089	0.013	0.057	0.107	0.014	0.056	0.101
$T = 1000$	0.011	0.052	0.090	0.007	0.057	0.114	0.011	0.049	0.106	0.007	0.050	0.098

Table 1.2: Size of $\mathcal{T}_{\mathrm{MS}}$ for the AR parameters $a_1 \in \{-0.9, 0.9\}$.

	$a_1 = -0.9$					$a_1 = 0.9$				
	sample size T					sample size T				
	250	500	1000	2000	3000	250	500	1000	2000	3000
$\alpha = 0.01$	0.040	0.032	0.017	0.009	0.012	0.003	0.016	0.015	0.021	0.017
$\alpha = 0.05$	0.137	0.093	0.067	0.061	0.047	0.017	0.038	0.055	0.059	0.057
$\alpha = 0.1$	0.218	0.160	0.124	0.108	0.098	0.040	0.054	0.095	0.096	0.106

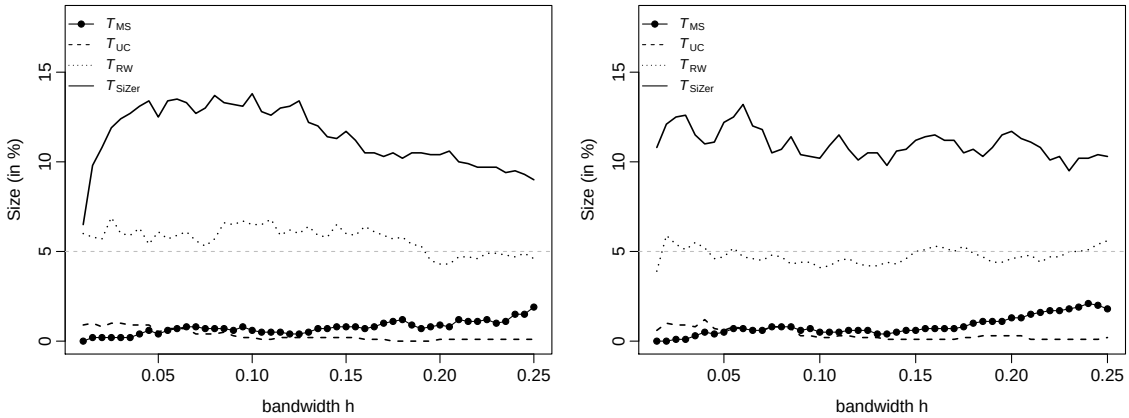
on the multiscale test $\mathcal{T}_{\mathrm{MS}}$. Table 1.1 reports the actual size of $\mathcal{T}_{\mathrm{MS}}$ for the AR parameters $a_1 \in \{-0.5, -0.25, 0.25, 0.5\}$, which is computed as the number of simulations in which $\mathcal{T}_{\mathrm{MS}}$ rejects the null divided by the total number of simulations. As can be seen, the actual size of the multiscale test $\mathcal{T}_{\mathrm{MS}}$ is fairly close to the nominal target α for all the considered AR parameters and sample sizes. Hence, the test has approximately the correct size.

In Table 1.1, we have explored the size of $\mathcal{T}_{\mathrm{MS}}$ when the errors are moderately autocorrelated. The case of strongly autocorrelated errors is investigated in Table 1.2, where we consider AR(1) errors with $a_1 \in \{-0.9, 0.9\}$. We first discuss the results for the positive AR parameter $a_1 = 0.9$. As can be seen, the size numbers are substantially downward biased for small sample sizes, in particular, for $T = 250$ and $T = 500$. As the sample size increases, this downward bias diminishes and the size numbers stabilize around their target α. In particular, for $T \geq 1000$, the size numbers give a decent approximation to α. An analogous picture arises for the negative AR parameter $a_1 = -0.9$. The size numbers, however, are upward rather than downward biased for small sample sizes T and the size distortions appear to vanish a bit more slowly as T increases. To summarize, in the case of strongly autocorrelated errors, our multiscale test has good size properties only for sufficiently large sample sizes. This is not very surprising: Statistical inference in the presence of strongly autocorrelated data is a very difficult problem in general and satisfying results can only be expected for fairly large sample sizes.

We next compare our multiscale test $\mathcal{T}_{\mathrm{MS}}$ with $\mathcal{T}_{\mathrm{UC}}$, $\mathcal{T}_{\mathrm{RW}}$ and $\mathcal{T}_{\mathrm{SiZer}}$ in terms of size. There is an important difference between $\mathcal{T}_{\mathrm{MS}}$ and $\mathcal{T}_{\mathrm{UC}}$ on the one hand and $\mathcal{T}_{\mathrm{RW}}$ and $\mathcal{T}_{\mathrm{SiZer}}$ on the other. $\mathcal{T}_{\mathrm{MS}}$ and its uncorrected version $\mathcal{T}_{\mathrm{UC}}$ are *global* test procedures: They test $H_0(u, h)$ simultaneously for all locations $u \in U_T$ and scales $h \in H_T$. Hence, they control the size simultaneously over both locations u and scales h. The methods $\mathcal{T}_{\mathrm{RW}}$ and $\mathcal{T}_{\mathrm{SiZer}}$, in contrast, are *row-wise* (or *scale-wise*) in nature: They test the hypothesis $H_0(u, h)$ simultaneously for all $u \in U_T$ but separately for each scale $h \in H_T$.

Table 1.3: Global size comparisons for the significance level $\alpha = 0.05$.

	$a_1 = -0.5$				$a_1 = 0.5$			
	$\mathcal{T}_{\text{MS}}$	$\mathcal{T}_{\text{UC}}$	$\mathcal{T}_{\text{RW}}$	$\mathcal{T}_{\text{SiZer}}$	$\mathcal{T}_{\text{MS}}$	$\mathcal{T}_{\text{UC}}$	$\mathcal{T}_{\text{RW}}$	$\mathcal{T}_{\text{SiZer}}$
$T = 250$	0.069	0.065	0.230	0.333	0.049	0.048	0.143	0.289
$T = 500$	0.054	0.065	0.288	0.448	0.042	0.026	0.187	0.397
$T = 1000$	0.046	0.051	0.318	0.522	0.052	0.049	0.276	0.509



(a) $a_1 = -0.5$ (b) $a_1 = 0.5$

Figure 1.2: Row-wise size comparisons for the significance level $\alpha = 5\%$ and the sample size $T = 1000$. Subfigure (a) corresponds to the case with $a_1 = -0.5$, subfigure (b) to the case with $a_1 = 0.5$. Each curve in the two subfigures shows the row-wise size (given in percentage % on the y-axis) as a function of the bandwidth h (specified on the x-axis) for one of the four tests $\mathcal{T}_{\text{MS}}$, $\mathcal{T}_{\text{UC}}$, $\mathcal{T}_{\text{RW}}$ and $\mathcal{T}_{\text{SiZer}}$.

Hence, they control the size for each scale $h \in H_T$ separately.

We conduct some simulation exercises to illustrate this important distinction. To keep the simulation study to a reasonable length, we restrict attention to the significance level $\alpha = 0.05$ and the AR parameters $a_1 \in \{-0.5, 0.5\}$. To simplify the implementation of $\mathcal{T}_{\text{SiZer}}$, we assume that the autocovariance function of the error process and thus the long-run error variance σ^2 is known. To keep the comparison fair, we treat σ^2 as known also when implementing $\mathcal{T}_{\text{MS}}$, $\mathcal{T}_{\text{UC}}$ and $\mathcal{T}_{\text{RW}}$. Moreover, we use exactly the same location-scale grid for all four methods. To achieve this, we start off with the grid $\mathcal{G}_T = U_T \times H_T$ with U_T and H_T defined above. We then follow Rondonotti et al. (2007) and Park et al. (2009a) and restrict attention to those points $(u, h) \in \mathcal{G}_T$ for which the effective sample size $\text{ESS}^*(u, h)$ for correlated data is not smaller than 5. This yields the grid $\mathcal{G}_T^* = \{(u, h) \in \mathcal{G}_T : \text{ESS}^*(u, h) \geq 5\}$. A definition of $\text{ESS}^*(u, h)$ is given in Section 1.C.1 of the Appendix.

For our simulation exercises, we distinguish between global and row-wise (or scale-wise) size: Global size is defined as the percentage of simulations in which the test under consideration rejects $H_0(u, h)$ for some $(u, h) \in \mathcal{G}_T^*$. Hence, it is identical to the size as computed in Tables 1.1 and 1.2. Row-wise size for scale $h^* \in H_T$, in contrast, is the percentage of simulations in which the test rejects $H_0(u, h^*)$ for some $(u, h^*) \in \mathcal{G}_T^*$.

Table 1.3 reports the global size of the four tests. As can be seen, the size numbers of our multiscale test $\mathcal{T}_{\text{MS}}$ and its uncorrected version $\mathcal{T}_{\text{UC}}$ are reasonably close to the target $\alpha = 0.05$. The global size numbers of the row-wise methods $\mathcal{T}_{\text{RW}}$ and $\mathcal{T}_{\text{SiZer}}$, in contrast, are much larger than the target $\alpha = 0.05$. Since the number of scales h in the grid $\mathcal{G}_T^*$ increases with T, they even move away from α as the sample size T increases. To summarize, as expected, the global tests $\mathcal{T}_{\text{MS}}$ and $\mathcal{T}_{\text{UC}}$ hold the size reasonably well, whereas the row-wise methods $\mathcal{T}_{\text{RW}}$ and $\mathcal{T}_{\text{SiZer}}$ are much too liberal.

Figure 1.2 reports the row-wise size of the four tests by so-called parallel coordinate plots [Inselberg (1985)] for the sample size $T = 1000$. Each curve in the figure specifies the row-wise size of one of the tests for the scales h under consideration. As can be seen, the row-wise version $\mathcal{T}_{\text{RW}}$ of our multiscale test holds the size quite accurately across scales. The row-wise size of $\mathcal{T}_{\text{SiZer}}$ also gives an acceptable approximation to the target $\alpha = 5\%$, even though the size numbers are upward biased quite a bit. The global tests $\mathcal{T}_{\text{MS}}$ and $\mathcal{T}_{\text{UC}}$, in contrast, have a row-wise size much smaller than the target $\alpha = 5\%$, which reflects the fact that they control global rather than row-wise size.

1.5.1.2 Power comparisons

In the second part of our simulation study, we compare the tests $\mathcal{T}_{\text{MS}}$, $\mathcal{T}_{\text{UC}}$, $\mathcal{T}_{\text{RW}}$ and $\mathcal{T}_{\text{SiZer}}$ in terms of power. As above, we use the location-scale grid $\mathcal{G}_T^*$ and treat the autocovariance function of the error terms as known when implementing the tests. Moreover, we restrict attention to the significance level $\alpha = 0.05$ and the AR parameters $a_1 \in \{-0.5, 0.5\}$. Our simulation exercises investigate the ability of the four tests to detect local increases in the trend m. (The same could of course be done for decreases.) The tests indicate a local increase in m according to the following decision rules: For each $(u, h) \in \mathcal{G}_T^*$,

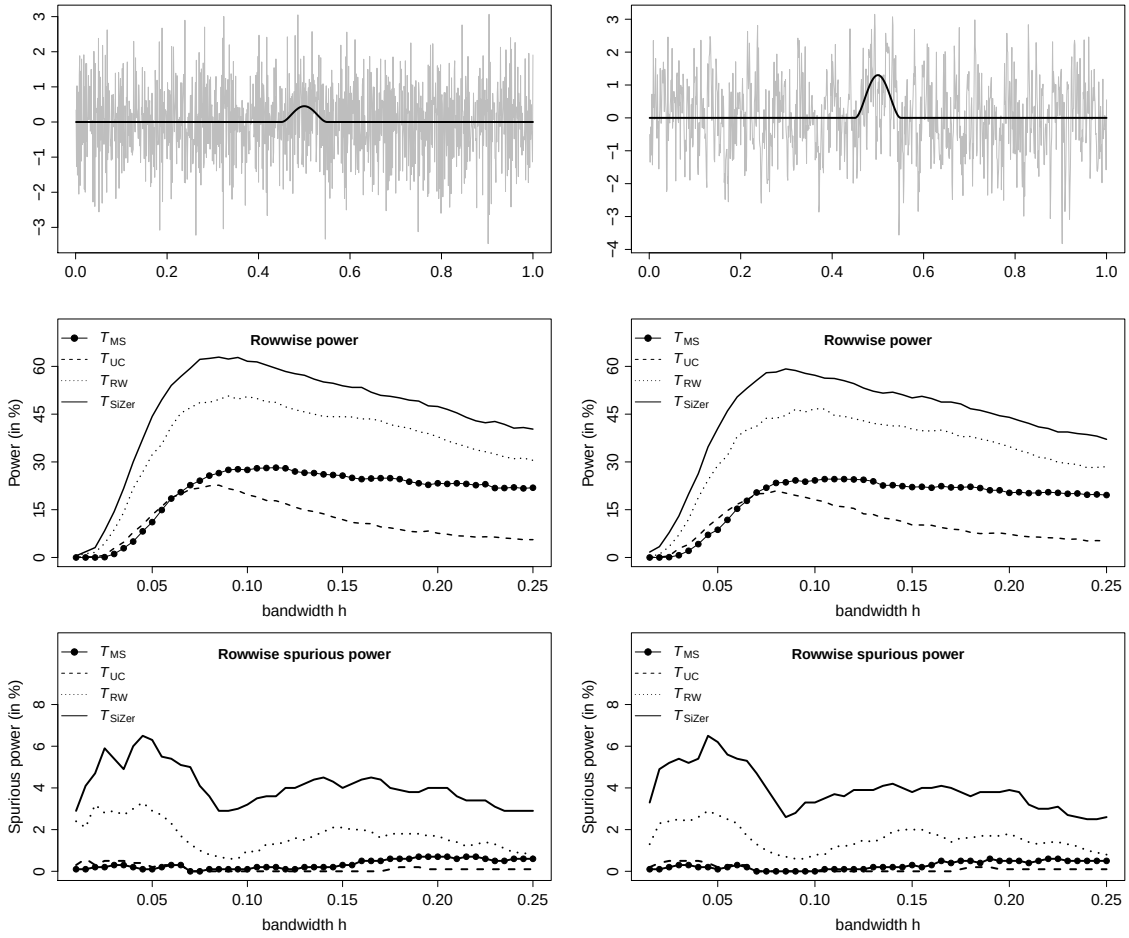
$$\begin{aligned}
\mathcal{T}_{\text{MS}} \text{ indicates an increase on } [u-h, u+h] &\iff \widehat{\psi}_T(u,h)/\widehat{\sigma} > q_T(\alpha) + \lambda(h) \\
\mathcal{T}_{\text{UC}} \text{ indicates an increase on } [u-h, u+h] &\iff \widehat{\psi}_T(u,h)/\widehat{\sigma} > q_T^{\text{UC}}(\alpha) \\
\mathcal{T}_{\text{RW}} \text{ indicates an increase on } [u-h, u+h] &\iff \widehat{\psi}_T(u,h)/\widehat{\sigma} > q_T^{\text{RW}}(\alpha, h) \\
\mathcal{T}_{\text{SiZer}} \text{ indicates an increase on } [u-h, u+h] &\iff \widehat{s}_T(u,h) > q_T^{\text{SiZer}}(\alpha, h),
\end{aligned}$$

where $q_T^{\text{UC}}(\alpha)$, $q_T^{\text{RW}}(\alpha, h)$, $q_T^{\text{SiZer}}(\alpha, h)$ are the critical values of $\mathcal{T}_{\text{UC}}$, $\mathcal{T}_{\text{RW}}$, $\mathcal{T}_{\text{SiZer}}$, respectively. Note that the critical values of $\mathcal{T}_{\text{RW}}$ and $\mathcal{T}_{\text{SiZer}}$ depend on the scale h as these are row-wise procedures.

To be able to make systematic power comparisons, we consider a very simple trend function m. More complicated signals m are analysed in Section 1.C.2 of the Appendix. The trend function we are considering here is defined as $m(u) = c \cdot 1(u \in [0.45, 0.55]) \cdot (1 - \{\frac{(u-0.5}{0.05}\}^2)^2$, where $c = 0.45$ in the AR case with $a_1 = -0.5$ and $c = 1.3$ in the case with $a_1 = 0.5$. The function m is increasing on $I^+ = (0.45, 0.5)$, decreasing on $I^- = (0.5, 0.55)$ and constant elsewhere. The two upper panels of Figure 1.3 give a graphical illustration of m, where the grey line in the background is the time series path of a representative simulated data sample. As can be seen, m is a small bump around $u = 0.5$, where c determines the height of the bump. The constant c is chosen such that the bump is difficult

Table 1.4: Global power and global spurious power comparisons for $\alpha = 0.05$.

		$a_1 = -0.5$				$a_1 = 0.5$			
		$\mathcal{T}_{\mathrm{MS}}$	$\mathcal{T}_{\mathrm{UC}}$	$\mathcal{T}_{\mathrm{RW}}$	$\mathcal{T}_{\mathrm{SiZer}}$	$\mathcal{T}_{\mathrm{MS}}$	$\mathcal{T}_{\mathrm{UC}}$	$\mathcal{T}_{\mathrm{RW}}$	$\mathcal{T}_{\mathrm{SiZer}}$
$T = 250$	Power	0.102	0.086	0.228	0.328	0.096	0.079	0.190	0.295
	Spurious power	0.021	0.032	0.109	0.166	0.012	0.017	0.054	0.131
$T = 500$	Power	0.212	0.166	0.464	0.617	0.186	0.160	0.406	0.587
	Spurious power	0.020	0.024	0.137	0.212	0.016	0.016	0.082	0.192
$T = 1000$	Power	0.575	0.425	0.817	0.901	0.526	0.394	0.780	0.884
	Spurious power	0.023	0.024	0.158	0.283	0.020	0.019	0.123	0.252



(a) $a_1 = -0.5$ (b) $a_1 = 0.5$

Figure 1.3: Row-wise power and row-wise spurious power comparisons for $\alpha = 5\%$ and $T = 1000$. Subfigure (a) corresponds to the case with $a_1 = -0.5$, subfigure (b) to the case with $a_1 = 0.5$. The upper panel of each subfigure shows the bump function m with a representative data sample in the background. The parallel coordinate plot in the middle panel reports row-wise power. In particular, each curve shows the row-wise power (given in percentage % on the y-axis) as a function of the bandwidth h (specified on the x-axis) for one of the four tests $\mathcal{T}_{\mathrm{MS}}$, $\mathcal{T}_{\mathrm{UC}}$, $\mathcal{T}_{\mathrm{RW}}$ and $\mathcal{T}_{\mathrm{SiZer}}$. The parallel coordinate plot in the lower panel reports row-wise spurious power in an analogous fashion.

but not impossible to detect for the four tests. We distinguish between the following types of power for the tests $\mathcal{T}_j$ with $j \in \{\text{MS}, \text{UC}, \text{RW}, \text{SiZer}\}$, where we restrict attention to increases in m:

(i) global power: the precentage of simulation runs in which the test $\mathcal{T}_j$ indicates an increase on some interval $I_{u,h} = [u-h, u+h]$ where m is indeed increasing, that is, on some $I_{u,h}$ with $I_{u,h} \cap I^+ \neq \emptyset$.

(ii) spurious global power: the percentage of simulation runs in which the test $\mathcal{T}_j$ indicates an increase on some interval $I_{u,h} = [u-h, u+h]$ where m is not increasing, that is, on some $I_{u,h}$ with $I_{u,h} \cap I^+ = \emptyset$.

(iii) row-wise power on scale h^*: the percentage of simulation runs in which the test $\mathcal{T}_j$ indicates an increase on some interval $I_{u,h^*} = [u-h^*, u+h^*]$ where m is indeed increasing, that is, on some I_{u,h^*} with $I_{u,h^*} \cap I^+ \neq \emptyset$.

(iv) spurious row-wise power on scale h^*: the percentage of simulation runs in which the test $\mathcal{T}_j$ indicates an increase on some interval $I_{u,h^*} = [u-h^*, u+h^*]$ where m is not increasing, that is, on some I_{u,h^*} with $I_{u,h^*} \cap I^+ = \emptyset$.

Table 1.4 reports the global power and global spurious power of the four tests. As can be seen, our multiscale test $\mathcal{T}_{\text{MS}}$ has higher power than the uncorrected version $\mathcal{T}_{\text{UC}}$. This confirms the theoretical optimality theory in Dümbgen and Spokoiny (2001) [see also Dümbgen and Walther (2008) and Rufibach and Walther (2010)] according to which the aggregation scheme of $\mathcal{T}_{\text{MS}}$ with its additive correction term should yield better power properties than the simpler scheme of $\mathcal{T}_{\text{UC}}$. As expected, the row-wise methods $\mathcal{T}_{\text{RW}}$ and $\mathcal{T}_{\text{SiZer}}$ have substantially more power than the global tests. Indeed, $\mathcal{T}_{\text{SiZer}}$ is even a bit more powerful than $\mathcal{T}_{\text{RW}}$, which is presumably due to the fact that it is somewhat too liberal in terms of row-wise size as observed in Figure 1.2. The higher power of the row-wise procedures comes at some cost: Their spurious global power is much higher than that of the global tests. For the sample size $T = 1000$ and the AR parameter $a_1 = -0.5$, for example, $\mathcal{T}_{\text{SiZer}}$ spuriously finds an increase in the trend m in more than 28% of the simulations, $\mathcal{T}_{\text{RW}}$ in more than 15%. The multiscale test $\mathcal{T}_{\text{MS}}$ (as well as its uncorrected version $\mathcal{T}_{\text{UC}}$), in contrast, controls the probability of finding a spurious increase. In particular, as implied by Proposition 1.3.3, its spurious global power is below $100 \cdot \alpha\% = 5\%$.

Figure 1.3 gives a more detailed picture of the power properties of the four tests for the sample size $T = 1000$. The parallel coordinate plots of the figure show how power and spurious power are distributed across scales h. Let us first have a look at the row-wise methods. As can be seen, $\mathcal{T}_{\text{SiZer}}$ is more powerful than $\mathcal{T}_{\text{RW}}$ on all scales under consideration. As already mentioned when discussing the global power results, this is presumably due to the fact that $\mathcal{T}_{\text{SiZer}}$ is a bit too liberal in terms of row-wise size. Comparing the power curves of the two global methods gives an interesting insight: Our

multiscale test $\mathcal{T}_{\mathrm{MS}}$ has substantially more power than the uncorrected version $\mathcal{T}_{\mathrm{UC}}$ on medium and large scales. On small scales, in contrast, it is slightly less powerful than $\mathcal{T}_{\mathrm{UC}}$. This again illustrates the theoretical optimality theory in Dümbgen and Spokoiny (2001) which suggests that, asymptotically, the multiscale test $\mathcal{T}_{\mathrm{MS}}$ should be as powerful as $\mathcal{T}_{\mathrm{UC}}$ on small scales but more powerful on large scales. This is essentially what we see in the two middle panels of Figure 1.3. Of course, $\mathcal{T}_{\mathrm{MS}}$ does not have exactly as much power as $\mathcal{T}_{\mathrm{UC}}$ on fine scales. However, the loss of power on fine scales is very small compared to the gain of power on larger scales (which is also reflected by the fact that $\mathcal{T}_{\mathrm{MS}}$ has more global power than $\mathcal{T}_{\mathrm{UC}}$).

The main findings of our simulation exercises can be summarized as follows: If one is interested in an exploratory data tool for finding local increases/decreases of a trend, the row-wise methods $\mathcal{T}_{\mathrm{RW}}$ and $\mathcal{T}_{\mathrm{SiZer}}$ both do a good job. However, if one wants to make rigorous statistical inference simultaneously across locations and scales, one needs to go for a global method. Our simulation exercises have demonstrated that our multiscale test $\mathcal{T}_{\mathrm{MS}}$ is a global method which enjoys good size and power properties. In particular, as predicted by the theory, it is a more effective test than the uncorrected version $\mathcal{T}_{\mathrm{UC}}$.

1.5.2 Small sample properties of the long-run variance estimator

In the final part of our simulation study, we analyse the estimators of the AR parameters and of the long-run error variance from Section 1.4 and compare them to the estimators of Hall and Van Keilegom (2003). We simulate data from the model $Y_{t,T} = m(t/T)+\varepsilon_t$, where $\{\varepsilon_t\}$ is an AR(1) process of the form $\varepsilon_t = a_1\varepsilon_{t-1} + \eta_t$. We consider the AR parameters $a_1 \in \{-0.95, -0.75, -0.5, -0.25, 0.25, 0.5, 0.75, 0.95\}$ and let η_t be i.i.d. standard normal innovation terms. Throughout the simulation study, the AR order $p^* = 1$ is treated as known. We report our findings for the sample size $T = 500$, the results for other sample sizes being very similar. For simplicity, m is chosen to be a linear function of the form $m(u) = \beta u$ with the slope parameter β. For each value of a_1, we consider two different slopes β, one corresponding to a moderate and one to a pronounced trend m. In particular, we let $\beta = s_\beta\sqrt{\mathrm{Var}(\varepsilon_t)}$ with $s_\beta \in \{1, 10\}$. When $s_\beta = 1$, the slope β is equal to the standard deviation $\sqrt{\mathrm{Var}(\varepsilon_t)}$ of the error process, which yields a moderate trend m. When $s_\beta = 10$, in contrast, the slope β is 10 times as large as $\sqrt{\mathrm{Var}(\varepsilon_t)}$, which results in a quite pronounced trend m.

For each model specification, we generate $S = 1000$ data samples and compute the following quantities for each simulated sample:

(i) the pilot estimator $\widetilde{a}_q$ from (1.4.8) with the tuning parameter q, the estimator $\widehat{a}$ from (1.4.10) with the tuning parameters $(\underline{r}, \overline{r})$ and the long-run variance estimator $\widehat{\sigma}^2$ from (1.4.11).

(ii) the estimators of a_1 and σ^2 from Hall and Van Keilegom (2003), which are denoted by $\widehat{a}_{\mathrm{HvK}}$ and $\widehat{\sigma}^2_{\mathrm{HvK}}$. The estimator $\widehat{a}_{\mathrm{HvK}}$ is computed as described in Section 2.2 of Hall and Van Keilegom (2003) and $\widehat{\sigma}^2_{\mathrm{HvK}}$ as defined at the bottom of p.447 in Section

2.3. The estimator $\widehat{a}_{\text{HvK}}$ (as well as $\widehat{\sigma}^2_{\text{HvK}}$) depends on two tuning parameters which we denote by m_1 and m_2 as in Hall and Van Keilegom (2003).

(iii) oracle estimators $\widehat{a}_{\text{oracle}}$ and $\widehat{\sigma}^2_{\text{oracle}}$ of a_1 and σ^2, which are constructed under the assumption that the error process $\{\varepsilon_t\}$ is observed. For each simulation run, we compute $\widehat{a}_{\text{oracle}}$ as the maximum likelihood estimator of a_1 from the time series of simulated error terms $\varepsilon_1, \ldots, \varepsilon_T$. We then calculate the residuals $r_t = \varepsilon_t - \widehat{a}_{\text{oracle}}\, \varepsilon_{t-1}$ and estimate the innovation variance $\nu^2 = \mathbb{E}[\eta_t^2]$ by $\widehat{\nu}^2_{\text{oracle}} = (T-1)^{-1} \sum_{t=2}^{T} r_t^2$. Finally, we set $\widehat{\sigma}^2_{\text{oracle}} = \widehat{\nu}^2_{\text{oracle}}/(1-\widehat{a}_{\text{oracle}})^2$.

Throughout the section, we set $q = 25$, $(\underline{r}, \overline{r}) = (1, 10)$ and $(m_1, m_2) = (20, 30)$. We in particular choose q to be in the middle of m_1 and m_2 to make the tuning parameters of the estimators $\widetilde{a}_q$ and $\widehat{a}_{\text{HvK}}$ more or less comparable. In order to assess how sensitive our estimators are to the choice of q and $(\underline{r}, \overline{r})$, we carry out a number of robustness checks, considering a range of different values for q and $(\underline{r}, \overline{r})$. In addition, we vary the tuning parameters m_1 and m_2 of the estimators from Hall and Van Keilegom (2003) to make sure that the results of our comparison study are not driven by the particular choice of any of the involved tuning parameters. The results of our robustness checks are reported in Section 1.C.3 of the Appendix. They show that the results of our comparison study are robust to different choices of the parameters q, $(\underline{r}, \overline{r})$ and (m_1, m_2).

For each estimator $\widehat{a}$, $\widehat{a}_{\text{HvK}}$, $\widehat{a}_{\text{oracle}}$ and $\widehat{\sigma}^2$, $\widehat{\sigma}^2_{\text{HvK}}$, $\widehat{\sigma}^2_{\text{oracle}}$ and for each model specification, the simulation output consists in a vector of length $S = 1000$ which contains the 1000 simulated values of the respective estimator. Figures 1.4 and 1.5 report the mean squared error (MSE) of these 1000 simulated values for each estimator. On the x-axis of each plot, the various values of the AR parameter a_1 are listed which are considered. The solid line in each plot gives the MSE values of our estimators. The dashed and dotted lines specify the MSE values of the HvK and the oracle estimators, respectively. Note that for the long-run variance estimators, the plots report the logarithm of the MSE rather than the MSE itself since the MSE values are too different across simulation scenarios to obtain a reasonable graphical presentation. In addition to the MSE values presented in Figures 1.4 and 1.5, we depict histograms of the 1000 simulated values produced by the estimators $\widehat{a}$, $\widehat{a}_{\text{HvK}}$, $\widehat{a}_{\text{oracle}}$ and $\widehat{\sigma}^2$, $\widehat{\sigma}^2_{\text{HvK}}$, $\widehat{\sigma}^2_{\text{oracle}}$ for two specific simulation scenarios in Figures 1.6 and 1.7. The main findings can be summarized as follows:

(a) In the simulation scenarios with a moderate trend ($s_\beta = 1$), the estimators $\widehat{a}_{\text{HvK}}$ and $\widehat{\sigma}^2_{\text{HvK}}$ of Hall and Van Keilegom (2003) exhibit a similar performance as our estimators $\widehat{a}$ and $\widehat{\sigma}^2$ as long as the AR parameter a_1 is not too close to -1. For strongly negative values of a_1 (in particular for $a_1 = -0.75$ and $a_1 = -0.95$), the estimators perform much worse than ours. This can be clearly seen from the much larger MSE values of the estimators $\widehat{a}_{\text{HvK}}$ and $\widehat{\sigma}^2_{\text{HvK}}$ for $a_1 = -0.75$ and $a_1 = -0.95$ in Figure 1.4. Figure 1.6 gives some further insights into what is happening here. It shows the histograms of the simulated values produced by the estimators $\widehat{a}$, $\widehat{a}_{\text{HvK}}$, $\widehat{a}_{\text{oracle}}$ and the corresponding long-run variance estimators in the scenario with $a_1 = -0.95$ and $s_\beta = 1$.

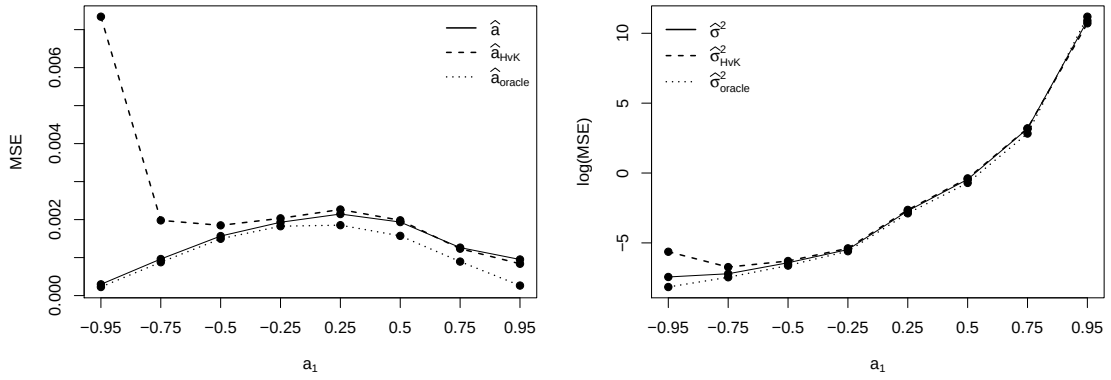


Figure 1.4: MSE values for the estimators $\widehat{a}$, $\widehat{a}_{\mathrm{HvK}}$, $\widehat{a}_{\mathrm{oracle}}$ and $\widehat{\sigma}^2$, $\widehat{\sigma}^2_{\mathrm{HvK}}$, $\widehat{\sigma}^2_{\mathrm{oracle}}$ in the simulation scenarios with a moderate trend ($s_\beta = 1$).

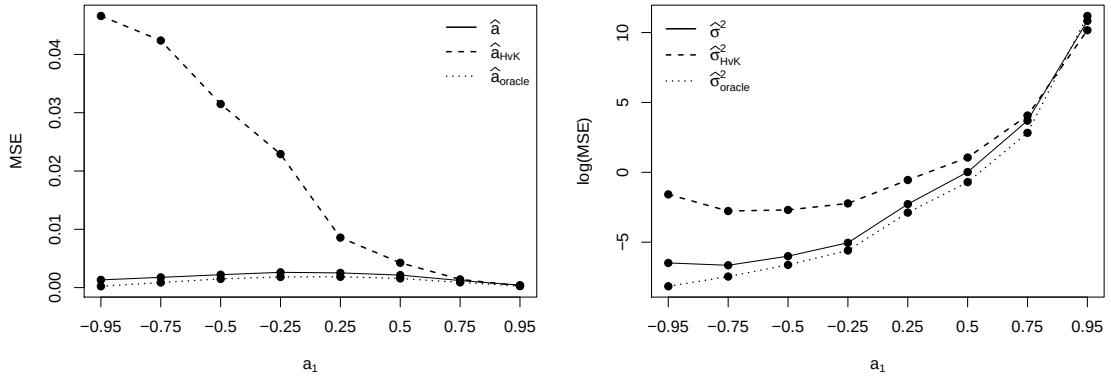


Figure 1.5: MSE values for the estimators $\widehat{a}$, $\widehat{a}_{\mathrm{HvK}}$, $\widehat{a}_{\mathrm{oracle}}$ and $\widehat{\sigma}^2$, $\widehat{\sigma}^2_{\mathrm{HvK}}$, $\widehat{\sigma}^2_{\mathrm{oracle}}$ in the simulation scenarios with a pronounced trend ($s_\beta = 10$).

As can be seen, the estimator $\widehat{a}_{\mathrm{HvK}}$ does not obey the causality restriction $|a_1| < 1$ but frequently takes values substantially smaller than -1. This results in a very large spread of the histogram and thus in a disastrous performance of the estimator.[4] A similar point applies to the histogram of the long-run variance estimator $\widehat{\sigma}^2_{\mathrm{HvK}}$. Our estimators $\widehat{a}$ and $\widehat{\sigma}^2$, in contrast, exhibit a stable behaviour in this case.
Interestingly, the estimator $\widehat{a}_{\mathrm{HvK}}$ (as well as the corresponding long-run variance estimator $\widehat{\sigma}^2_{\mathrm{HvK}}$) performs much worse than ours for large negative values but not for large positive values of a_1. This can be explained as follows: In the special case of an AR(1) process, the estimator $\widehat{a}_{\mathrm{HvK}}$ may produce estimates smaller than -1 but it cannot become larger than 1. This can be easily seen upon inspecting the definition of the estimator. Hence, for large positive values of a_1, the estimator $\widehat{a}_{\mathrm{HvK}}$ performs well as it satisfies the causality restriction that the estimated AR parameter should be smaller than 1.

(b) In the simulation scenarios with a pronounced trend ($s_\beta = 10$), the estimators of

[4] One could of course set $\widehat{a}_{\mathrm{HvK}}$ to $-(1-\delta)$ for some small $\delta > 0$ whenever it takes a value ≤ -1. This modified estimator, however, is still far from performing in a satisfactory way when a_1 is close to -1.

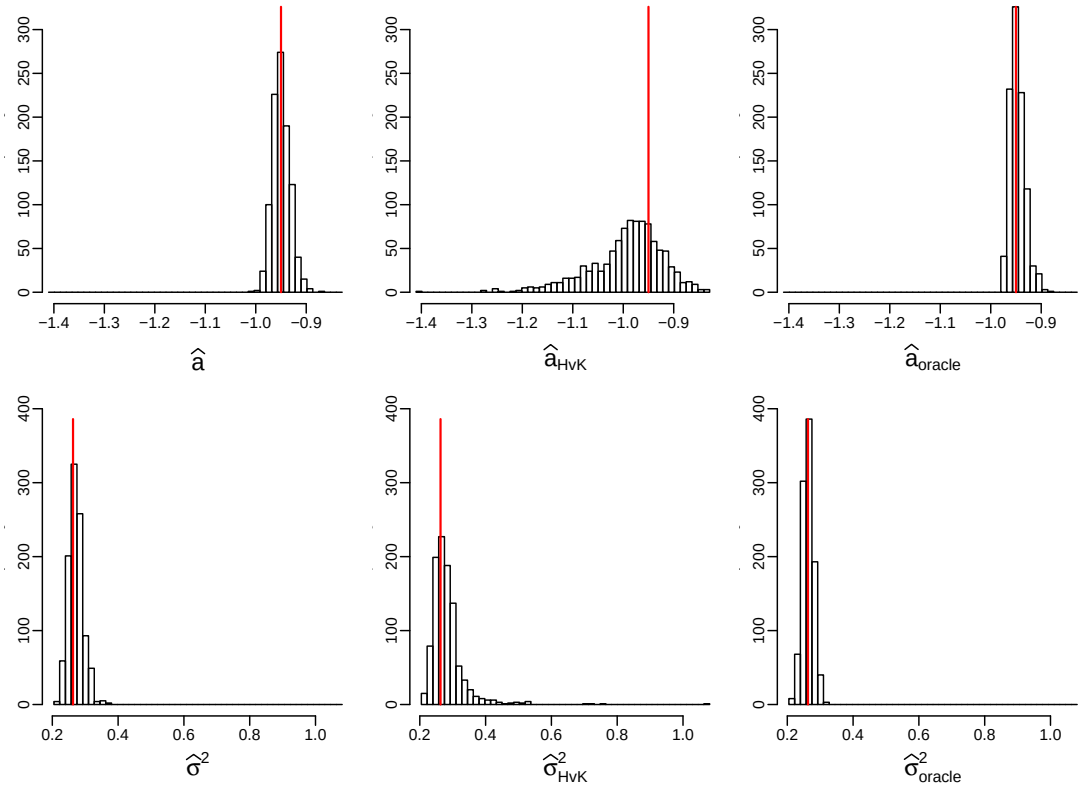


Figure 1.6: Histograms of the simulated values produced by the estimators $\widehat{a}$, $\widehat{a}_{\mathrm{HvK}}$, $\widehat{a}_{\mathrm{oracle}}$ and $\widehat{\sigma}^2$, $\widehat{\sigma}^2_{\mathrm{HvK}}$, $\widehat{\sigma}^2_{\mathrm{oracle}}$ in the scenario with $a_1 = -0.95$ and $s_\beta = 1$. The vertical red lines indicate the true values of a_1 and σ^2.

Hall and Van Keilegom (2003) are clearly outperformed by ours for most of the AR parameters a_1 under consideration. In particular, their MSE values reported in Figure 1.5 are much larger than the values produced by our estimators for most parameter values a_1. The reason is the following: The HvK estimators have a strong bias since the pronounced trend with $s_\beta = 10$ is not eliminated appropriately by the underlying differencing methods. This point is illustrated by Figure 1.7 which shows histograms of the simulated values for the estimators $\widehat{a}$, $\widehat{a}_{\mathrm{HvK}}$, $\widehat{a}_{\mathrm{oracle}}$ and the corresponding long-run variance estimators in the scenario with $a_1 = 0.25$ and $s_\beta = 10$. As can be seen, the histogram produced by our estimator $\widehat{a}$ is approximately centred around the true value $a_1 = 0.25$, whereas that of $\widehat{a}_{\mathrm{HvK}}$ is strongly biased upwards. A similar picture arises for the long-run variance estimators $\widehat{\sigma}^2$ and $\widehat{\sigma}^2_{\mathrm{HvK}}$.

Whereas the methods of Hall and Van Keilegom (2003) perform much worse than ours for negative and moderately positive values of a_1, the performance (in terms of MSE) is fairly similar for large values of a_1. This can be explained as follows: When the trend m is not eliminated appropriately by taking differences, this creates spurious persistence in the data. Hence, the estimator $\widehat{a}_{\mathrm{HvK}}$ tends to overestimate the AR parameter a_1, that is, $\widehat{a}_{\mathrm{HvK}}$ tends to be larger in absolute value than a_1. Very loosely speaking, when the parameter a_1 is close to 1, say $a_1 = 0.95$, there is not much room

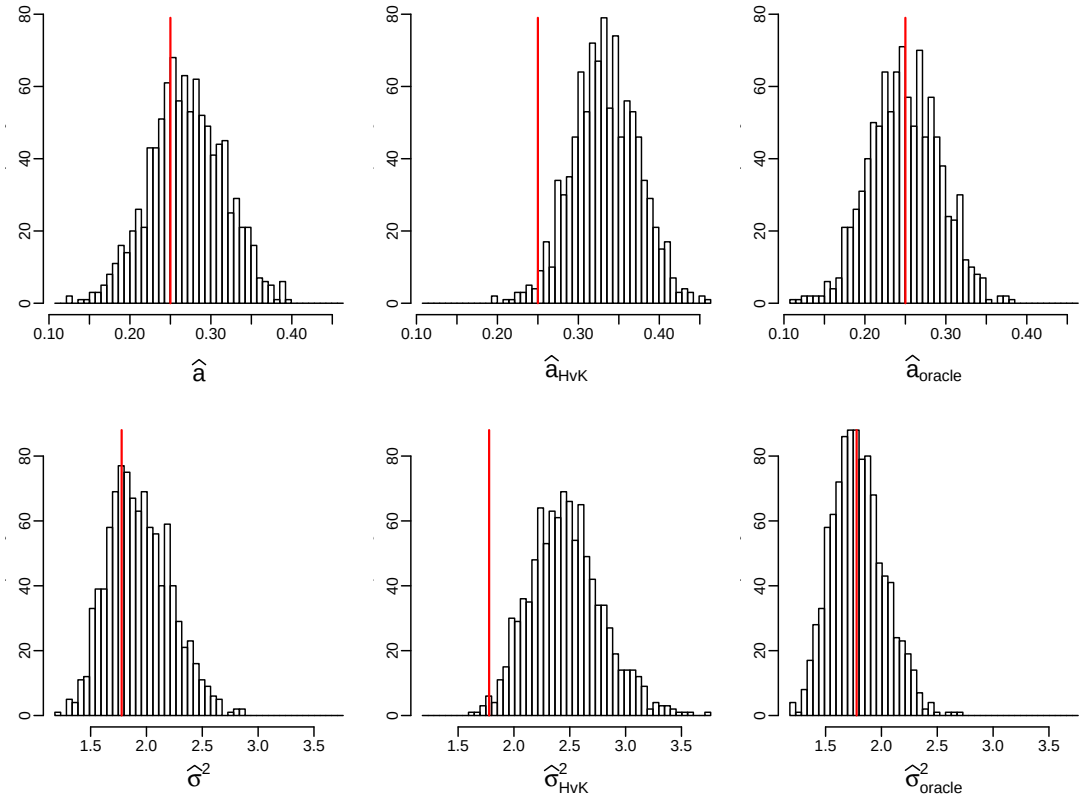


Figure 1.7: Histograms of the simulated values produced by the estimators $\widehat{a}$, $\widehat{a}_{\mathrm{HvK}}$, $\widehat{a}_{\mathrm{oracle}}$ and $\widehat{\sigma}^2$, $\widehat{\sigma}^2_{\mathrm{HvK}}$, $\widehat{\sigma}^2_{\mathrm{oracle}}$ in the scenario with $a_1 = 0.25$ and $s_\beta = 10$. The vertical red lines indicate the true values of a_1 and σ^2.

for overestimation since $\widehat{a}_{\mathrm{HvK}}$ cannot become larger than 1. Consequently, the effect of not eliminating the trend appropriately has a much smaller impact on $\widehat{a}_{\mathrm{HvK}}$ for large positive values of a_1.

1.6 Application

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. In what follows, we use our multiscale test $\mathcal{T}_{\mathrm{MS}}$ to analyse two long-term temperature records. Throughout the section, we set the significance level to $\alpha = 0.05$ and implement the multiscale test in exactly the same way as in the simulation study of Section 1.5.

1.6.1 Analysis of the Central England temperature record

The Central England temperature record is the longest instrumental temperature time series in the world. The data are 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

$Y_{t,T}$ covering the years from 1659 to 2017. A plot of the time series is given in panel (a) of Figure 1.8. We assume that the temperature data $Y_{t,T}$ follow the nonparametric trend model $Y_{t,T} = m(t/T) + \varepsilon_t$, where m is the unknown time trend of interest. The error process $\{\varepsilon_t\}$ is supposed to have the AR(p^*) structure $\varepsilon_t = \sum_{j=1}^{p^*} a_j \varepsilon_{t-j} + \eta_t$, where η_t are i.i.d. innovations with mean 0 and variance ν^2. As pointed out in Mudelsee (2010) among others, this is the most widely used error model for discrete climate time series. We select the AR order p^* by the Bayesian information criterion (BIC), which yields $p^* = 2$.[5] We then estimate the AR(2) parameters $\boldsymbol{a} = (a_1, a_2)$ and the long-run error variance σ^2 by the procedures from Section 1.4 with $q = 25$ and $(\underline{r}, \overline{r}) = (1, 10)$. This gives the estimators $\widehat{a}_1 = 0.164$, $\widehat{a}_2 = 0.175$ and $\widehat{\sigma}^2 = 0.737$.

With the help of our multiscale method, we now test the null hypothesis H_0 that m is constant on all intervals $[u-h, u+h]$ with $(u,h) \in \mathcal{G}_T^*$, where the grid $\mathcal{G}_T^*$ is defined in the same way as in Section 1.5. The results are presented in Figure 1.8. Panel (b) depicts the minimal intervals in the set Π_T^+ which is produced by our multiscale test $\mathcal{T}_{\mathrm{MS}}$. The set of intervals Π_T^- is empty in the present case. According to Proposition 1.3.3, we can make the following simultaneous confidence statement about the collection of minimal intervals plotted in panel (b). We can claim, with confidence of about 95%, that the trend 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 1870 onwards. Hence, our test in particular provides evidence that there has been some warming trend in the period over approximately the last 150 years. On the other hand, as the set Π_T^- is empty, there is no evidence of any downward movement of the trend.

Panel (c) presents the SiZer map produced by our multiscale test $\mathcal{T}_{\mathrm{MS}}$. For comparison, the SiZer map of the dependent SiZer test $\mathcal{T}_{\mathrm{SiZer}}$ is shown in panel (d). To produce panel (d), we have implemented SiZer as described in Section 1.C.1 of the Appendix, where the autocovariance function of the errors $\{\varepsilon_t\}$ is estimated with the help of our procedures from Section 1.4 under the assumption that $\{\varepsilon_t\}$ is an AR(2) process. The SiZer maps of panels (c) and (d) are to be read as follows: Each pixel of the map corresponds to a location-scale point (u,h), or put differently, to a time interval $[u-h, u+h]$. The pixel (u,h) is coloured blue if the test indicates an increase in the trend m on the interval $[u-h, u+h]$, red if the test indicates a decrease and purple if the test does not reject the null hypothesis that m is constant on $[u-h, u+h]$. As can be seen, the two SiZer maps in panels (c) and (d) have a similar structure. Both our multiscale test and SiZer indicate increases in the trend m during a short time period around 1700 and towards the end of the sample. However, in contrast to SiZer, our method allows to make formal

[5] More precisely, we proceed as follows: We estimate the AR parameters and the corresponding variance of the innovation terms for different AR orders by the methods from Section 1.4 and then choose p^* as the minimizer of the Bayesian information criterion (BIC). As a robustness check, we have repeated this procedure for a wide range of the tuning parameters q and $(\underline{r}, \overline{r})$, which produces the value $p^* = 2$ throughout. Moreover, we have considered other information criteria such as FPE, AIC and AICC, which gives the AR order $p^* = 2$ for almost all values of q and $(\underline{r}, \overline{r})$.

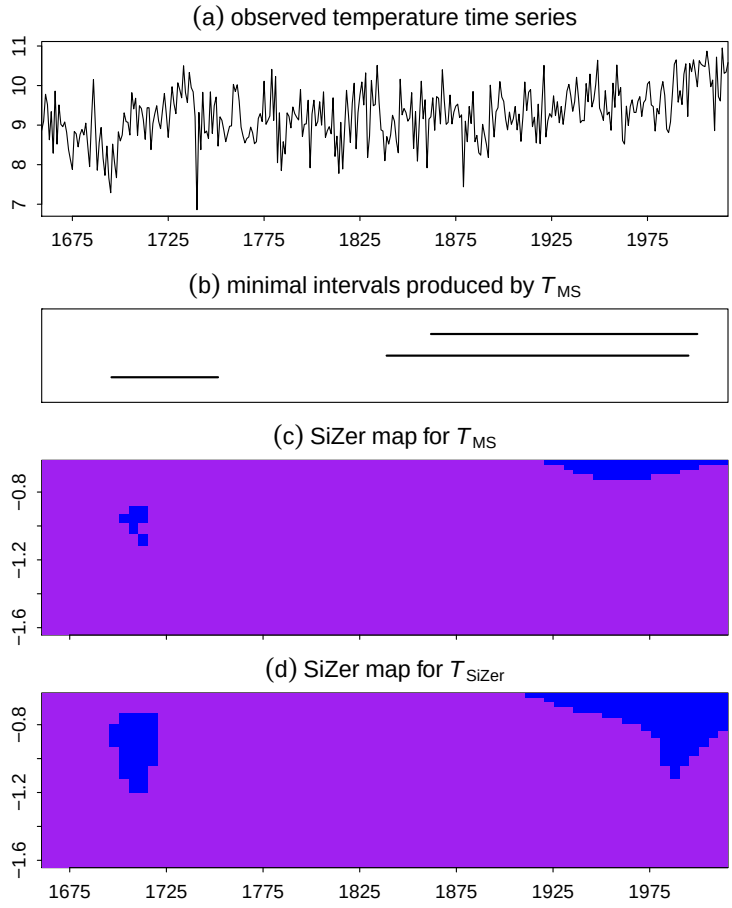


Figure 1.8: Summary of the results for the Central England temperature record. Panel (a) shows the observed temperature time series. Panel (b) depicts the minimal intervals in the set Π_T^+ produced by our multiscale test. These are $[1684, 1744]$, $[1839, 2009]$ and $[1864, 2014]$. Panels (c) and (d) present the SiZer maps produced by our multiscale test and SiZer.

confidence statements about the regions of blue pixels in the SiZer map. In particular, as the set of blue pixels in panel (c) exactly corresponds to the collection of intervals Π_T^+, we can claim, with confidence of about 95%, that the trend m has an increase on each time interval represented by a blue pixel in panel (c).

1.6.2 Analysis of global temperature data

We next analyse a data set which consists of annual global temperature anomalies from 1850 onwards. The data are publicly available on the webpage `https://cdiac.ess-dive.lbl.gov/trends/temp/jonescru/jones.html` and are plotted in panel (a) of Figure 1.9. As before, we assume that the data come from the model $Y_{t,T} = m(t/T) + \varepsilon_t$, where m is the trend and $\{\varepsilon_t\}$ the noise process. We apply our multiscale methods to test the null hypothesis H_0 that m is constant on all time intervals $[u - h, u + h]$ with $(u, h) \in \mathcal{G}_T$, where the grid $\mathcal{G}_T$ is defined as in Section 1.5. We compare our results with

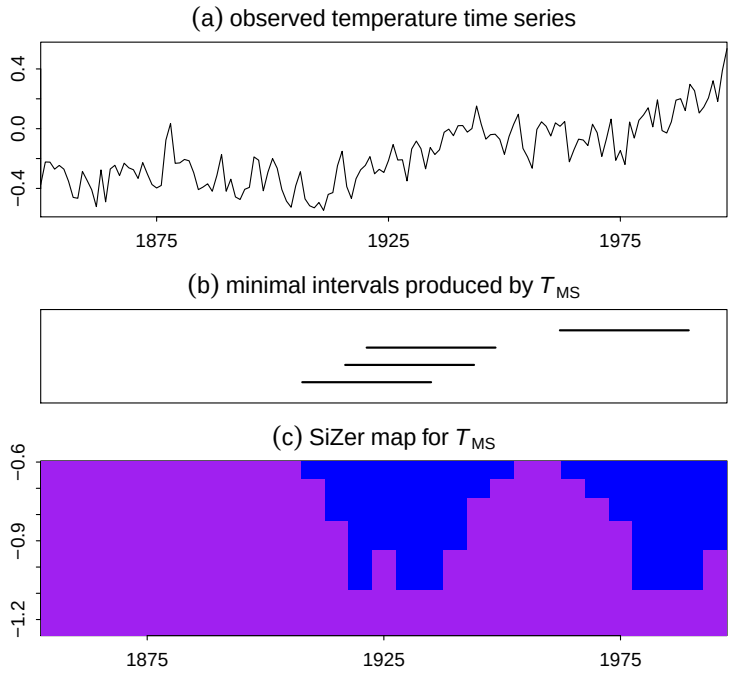


Figure 1.9: Summary of the results for the global temperature anomalies. Panel (a) shows the observed temperature time series. Panel (b) depicts the minimal intervals in the set Π_T^+ produced by the multiscale test. These are $[1905, 1935]$, $[1915, 1945]$, $[1920, 1950]$ and $[1965, 1995]$. Panel (c) presents the SiZer map of our test.

those obtained by Wu et al. (2001) who developed a method for testing the hypothesis that m is constant on $[0, 1]$ against the alternative that m is an arbitrary monotonic function. For comparability reasons, we use exactly the same data as in Wu et al. (2001), in particular, the yearly temperature anomalies from 1856 to 1998. Moreover, we use their estimate of the long-run error variance σ^2 which amounts to 0.01558. As we do not have an estimate available from Wu et al. (2001) for the autocovariance function of the error process, we do not consider dependent SiZer in the application example at hand.

The results produced by our multiscale test are reported in Figure 1.9. Panel (b) shows the minimal intervals in Π_T^+ and panel (c) the SiZer map of the test. As can be clearly seen from both panels (b) and (c), the test indicates an increase in the trend m during the first half of the 20th century followed by another increase during the second half. These findings are in line with those in Wu et al. (2001) who reject the null hypothesis that m is constant. In contrast to the test of Wu et al. (2001), however, our multiscale method does not only allow to test whether the null is violated. It also allows to make formal confidence statements about where violations occur, that is, about where the trend m is increasing. In particular, we can claim, with confidence of about 95%, that the trend has an increase on each interval plotted in panel (b) of Figure 1.9.

APPENDICES

1.A Proofs of the results from Section 1.3

In this section, we prove the theoretical results from Section 1.3. 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 (1.3.5) which is close to a Gaussian statistic whose distribution is known. More specifically, we prove the following result.

Proposition 1.A.1. *Under the conditions of Theorem 1.3.1, 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)*

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

where Φ_T is a Gaussian statistic as defined in (1.3.4).

Proof of Proposition 1.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 $[\widetilde{\varepsilon}_1, \ldots, \widetilde{\varepsilon}_T] \overset{\mathcal{D}}{=} [\varepsilon_1, \ldots, \varepsilon_T]$ for each T and

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

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 $Y_{t,T} = m(t/T) + \varepsilon_t$ replaced by $\widetilde{Y}_{t,T} = m(t/T) + \widetilde{\varepsilon}_t$ 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

$$\big|\widetilde{\Phi}_T - \Phi_T\big| \leq \big|\widetilde{\Phi}_T - \Phi_T^\diamond\big| + \big|\Phi_T^\diamond - \Phi_T\big| = \big|\widetilde{\Phi}_T - \Phi_T^\diamond\big| + o_p\big(\rho_T\sqrt{\log T}\big), \tag{1.A.2}$$

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

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

Using summation by parts, we further obtain that

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

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 $\max_{(u,h)\in\mathcal{G}_T} W_T(u,h) = O(1/\sqrt{Th_{\min}})$. Applying the strong approximation result (3.C.4), we can thus infer that

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

Plugging (1.A.3) into (3.C.5) completes the proof. □

Auxiliary results using anti-concentration bounds

In this section, we establish some properties of the Gaussian statistic Φ_T defined in (1.3.4). 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 1.A.2. *Under the conditions of Theorem 1.3.1, it holds that*

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

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

Proof of Proposition 1.A.2. The main technical tool for proving Proposition 1.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 1.A.3. *Let $(X_1, \dots, X_p)^\top$ 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 $1 \leq j \leq p$. Define $\overline{\mu} = \max_{1\leq j\leq p} |\mu_j|$ together with $\underline{\sigma} = \min_{1\leq j\leq p} \sigma_j$ and $\overline{\sigma} = \max_{1\leq j\leq p} \sigma_j$. Moreover, set $a_p = \mathbb{E}[\max_{1\leq j\leq p}(X_j - \mu_j)/\sigma_j]$ and $b_p = \mathbb{E}[\max_{1\leq j\leq p}(X_j - \mu_j)]$. For every $\delta > 0$, it holds that*

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

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

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

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

where $(X_1, \dots, X_{2p})^\top$ is a Gaussian random vector with the following properties: (i) $\mu_j := \mathbb{E}[X_j] = -\lambda(x_{j2})$ and thus $\overline{\mu} = \max_{1\leq j\leq 2p} |\mu_j| \leq C\sqrt{\log T}$, 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 $a_{2p} = b_{2p} \leq \sqrt{2\log(2p)} \leq C\sqrt{\log T}$. With this notation at hand, we can apply Proposition 1.A.3 to obtain that

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

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

Proof of Theorem 1.3.1

To prove Theorem 1.3.1, we make use of the two auxiliary results derived above. By Proposition 1.A.1, there exist statistics $\widetilde{\Phi}_T$ for $T = 1, 2, \dots$ which are distributed as $\widehat{\Phi}_T$ for any $T \geq 1$ and which have the property that

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

where Φ_T is a Gaussian statistic as defined in (1.3.4). The approximation result (1.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

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

which immediately implies the statement of Theorem 1.3.1. For the proof of (1.A.5), we use the following simple lemma:

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

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

then

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

The statement of Lemma 3.C.10 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 (1.A.6), then the distribution of W_T can be approximated by that of V_T in the sense of (1.A.7).

Proof of Lemma 3.C.10. It holds that

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

□

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 (1.A.4), we already know that $\widetilde{\Phi}_T - \Phi_T = o_p(\delta_T)$. Moreover, by Proposition 1.A.2, it holds that

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

Hence, the conditions of Lemma 3.C.10 are satisfied. Applying the lemma, we obtain (1.A.5), which completes the proof of Theorem 1.3.1.

Proof of Proposition 1.3.2

To start with, we introduce the notation $\widehat{\psi}_T(u,h) = \widehat{\psi}_T^A(u,h) + \widehat{\psi}_T^B(u,h)$ 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 $m_T'(w) \geq c_T\sqrt{\log T/(Th_0^3)}$ for all $w \in [u_0 - h_0, u_0 + h_0]$. (The case that $-m_T'(w) \geq c_T\sqrt{\log T/(Th_0^3)}$ for all w can be treated analogously.) Below, we prove that under this assumption,

$$\widehat{\psi}_T^B(u_0, h_0) \geq \frac{\kappa c_T\sqrt{\log T}}{2} \tag{1.A.9}$$

for sufficiently large T, where $\kappa = (\int K(\varphi)\varphi^2 d\varphi)/(\int K^2(\varphi)\varphi^2 d\varphi)^{1/2}$. Moreover, by arguments very similar to those for the proof of Proposition 1.A.1, it follows that

$$\max_{(u,h)\in\mathcal{G}_T} |\widehat{\psi}_T^A(u,h)| = O_p(\sqrt{\log T}). \tag{1.A.10}$$

With the help of (1.A.9), (1.A.10) and the fact that $\lambda(h) \leq \lambda(h_{\min}) \leq C\sqrt{\log T}$, we can infer that

$$\begin{aligned}
\widehat{\Psi}_T &\geq \max_{(u,h)\in\mathcal{G}_T} \frac{|\widehat{\psi}_T^B(u,h)|}{\widehat{\sigma}} - \max_{(u,h)\in\mathcal{G}_T} \Big\{ \frac{|\widehat{\psi}_T^A(u,h)|}{\widehat{\sigma}} + \lambda(h) \Big\} \\
&= \max_{(u,h)\in\mathcal{G}_T} \frac{|\widehat{\psi}_T^B(u,h)|}{\widehat{\sigma}} + O_p(\sqrt{\log T}) \\
&\geq \frac{\kappa c_T\sqrt{\log T}}{2\widehat{\sigma}} + O_p(\sqrt{\log T})
\end{aligned} \tag{1.A.11}$$

for sufficiently large T. Since $q_T(\alpha) = O(\sqrt{\log T})$ for any fixed $\alpha \in (0,1)$, (1.A.11) immediately yields that $\mathbb{P}(\widehat{\Psi}_T \leq q_T(\alpha)) = o(1)$, which is the statement of Proposition 1.3.2.

Proof of (1.A.9). Write $m_T(\frac{t}{T}) = m_T(u_0) + m_T'(\xi_{u_0,t,T})(\frac{t}{T} - u_0)$, where $\xi_{u_0,t,T}$ is an intermediate point between u_0 and t/T. The local linear weights $w_{t,T}(u_0, h_0)$ are constructed such that $\sum_{t=1}^T w_{t,T}(u_0, h_0) = 0$. We thus obtain that

$$\widehat{\psi}_T^B(u_0, h_0) = \sum_{t=1}^T w_{t,T}(u_0, h_0)\Big(\frac{\frac{t}{T} - u_0}{h_0}\Big) h_0 m_T'(\xi_{u_0,t,T}). \tag{1.A.12}$$

Moreover, since the kernel K is symmetric and $u_0 = t/T$ for some t, it holds that $S_{T,1}(u_0, h_0) = 0$, which in turn implies that

$$\begin{aligned}
&w_{t,T}(u_0, h_0)\Big(\frac{\frac{t}{T} - u_0}{h_0}\Big) \\
&\quad = K\Big(\frac{\frac{t}{T} - u_0}{h_0}\Big)\Big(\frac{\frac{t}{T} - u_0}{h_0}\Big)^2 \Big/ \Big\{ \sum_{t=1}^T K^2\Big(\frac{\frac{t}{T} - u_0}{h_0}\Big)\Big(\frac{\frac{t}{T} - u_0}{h_0}\Big)^2 \Big\}^{1/2} \geq 0.
\end{aligned} \tag{1.A.13}$$

From (1.A.12), (1.A.13) and the assumption that $m_T'(w) \geq c_T\sqrt{\log T/(Th_0^3)}$ for all $w \in$

$[u_0 - h_0, u_0 + h_0]$, we get that

$$\widehat{\psi}_T^B(u_0, h_0) \geq c_T \sqrt{\frac{\log T}{T h_0}} \sum_{t=1}^{T} w_{t,T}(u_0, h_0) \Big(\frac{\frac{t}{T} - u_0}{h_0} \Big). \tag{1.A.14}$$

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,

$$\Big| \frac{1}{Th} \sum_{t=1}^{T} K\Big(\frac{\frac{t}{T} - u}{h} \Big) \Big(\frac{\frac{t}{T} - u}{h} \Big)^{\ell} - \int_0^1 \frac{1}{h} K\Big(\frac{w - u}{h} \Big) \Big(\frac{w - u}{h} \Big)^{\ell} dw \Big| \leq \frac{C}{Th}, \tag{1.A.15}$$

where the constant C does not depend on u, h and T. With the help of (1.A.13) and (3.C.41), we obtain that for any $(u, h) \in \mathcal{G}_T$ with $[u - h, u + h] \subseteq [0, 1]$,

$$\Big| \sum_{t=1}^{T} w_{t,T}(u, h) \Big(\frac{\frac{t}{T} - u}{h} \Big) - \kappa \sqrt{Th} \Big| \leq \frac{C}{\sqrt{Th}}, \tag{1.A.16}$$

where the constant C does once again not depend on u, h and T. (1.A.16) implies that $\sum_{t=1}^{T} w_{t,T}(u, h)(\frac{t}{T} - u)/h \geq \kappa \sqrt{Th}/2$ for sufficiently large T and any $(u, h) \in \mathcal{G}_T$ with $[u-h, u+h] \subseteq [0, 1]$. Using this together with (1.A.14), we immediately obtain (1.A.9). □

Proof of Proposition 1.3.3

In what follows, we show that

$$\mathbb{P}(E_T^+) \geq (1 - \alpha) + o(1). \tag{1.A.17}$$

The other statements of Proposition 1.3.3 can be verified by analogous arguments. (1.A.17) is a consequence of the following two observations:

(i) For all $(u, h) \in \mathcal{G}_T$ with

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

it holds that $\mathbb{E}[\widehat{\psi}_T(u, h)] > 0$.

(ii) For all $(u, h) \in \mathcal{G}_T$ with $[u - h, u + h] \subseteq [0, 1]$, $\mathbb{E}[\widehat{\psi}_T(u, h)] > 0$ implies that $m'(v) > 0$ for some $v \in [u - h, u + h]$.

Observation (i) is trivial, (ii) can be seen as follows: Let (u, h) be any point with $(u, h) \in \mathcal{G}_T$ and $[u - h, u + h] \subseteq [0, 1]$. It holds that $\mathbb{E}[\widehat{\psi}_T(u, h)] = \widehat{\psi}_T^B(u, h)$, where $\widehat{\psi}_T^B(u, h)$ has been defined in the proof of Proposition 1.3.2. As already shown in (1.A.12),

$$\widehat{\psi}_T^B(u, h) = \sum_{t=1}^{T} w_{t,T}(u, h) \Big(\frac{\frac{t}{T} - u}{h} \Big) h m'(\xi_{u,t,T}),$$

where $\xi_{u,t,T}$ is some intermediate point between u and t/T. Moreover, by (1.A.13), it holds that $w_{t,T}(u,h)(\frac{t}{T}-u)/h \geq 0$ for any t. Hence, $\mathbb{E}[\widehat{\psi}_T(u,h)] = \widehat{\psi}_T^B(u,h)$ can only take a positive value if $m'(v) > 0$ for some $v \in [u-h, u+h]$.

From observations (i) and (ii), we can draw the following conclusions: On the event

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

it holds that for all $(u,h) \in \mathcal{A}_T^+$ with $[u-h, u+h] \subseteq [0,1]$, $m'(v) > 0$ for some $v \in I_{u,h} = [u-h, u+h]$. We thus obtain that $\{\widehat{\Phi}_T \leq q_T(\alpha)\} \subseteq E_T^+$. This in turn implies that

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

where the last equality holds by Theorem 1.3.1.

Proof of Corollary 1.3.1

Let $\alpha = \alpha_T \to 0$ and let $\widetilde{\Phi}_T$ be defined as in the proof of Theorem 1.3.1. It holds that

$$\begin{aligned} \big|\mathbb{P}(\widetilde{\Phi}_T \leq q_T(\alpha_T)) - (1-\alpha_T)\big| &= \big|\mathbb{P}(\widetilde{\Phi}_T \leq q_T(\alpha_T)) - \mathbb{P}(\Phi_T \leq q_T(\alpha_T))\big| \\ &\leq \sup_{x\in\mathbb{R}} \big|\mathbb{P}(\widetilde{\Phi}_T \leq x) - \mathbb{P}(\Phi_T \leq x)\big| = o(1), \end{aligned}$$

where the last equality is due to (1.A.5). From this, it immediately follows that $\mathbb{P}(\widetilde{\Phi}_T \leq q_T(\alpha_T)) \to 1$. Moreover, since $\widetilde{\Phi}_T$ and $\widehat{\Phi}_T$ have the same distribution by construction, we obtain that

$$\mathbb{P}(\widehat{\Phi}_T \leq q_T(\alpha_T)) \to 1. \tag{1.A.18}$$

Taking into account (1.A.18), Corollary 1.3.1 can be proven in exactly the same way as Proposition 1.3.3.

Proof of Proposition 1.A.3

The proof makes use of the following three lemmas, which correspond to Lemmas 5–7 in Chernozhukov et al. (2015).

Lemma 1.A.2. *Let $(W_1, \ldots, W_p)^\top$ be a (not necessarily centred) Gaussian random vector in $\mathbb{R}^p$ with $\mathrm{Var}(W_j) = 1$ for all $1 \leq j \leq p$. Suppose that $\mathrm{Corr}(W_j, W_k) < 1$ whenever $j \neq k$. Then the distribution of $\max_{1\leq j\leq p} W_j$ is absolutely continuous with respect to Lebesgue measure and a version of the density is given by*

$$f(x) = f_0(x) \sum_{j=1}^{p} e^{\mathbb{E}[W_j]x - \mathbb{E}[W_j]^2/2}\, \mathbb{P}\big(W_k \leq x \text{ for all } k \neq j \,\big|\, W_j = x\big).$$

Lemma 1.A.3. *Let $(W_0, W_1, \dots, W_p)^\top$ be a (not necessarily centred) Gaussian random vector with $\mathrm{Var}(W_j) = 1$ for all $0 \le j \le p$. Suppose that $\mathbb{E}[W_0] \ge 0$. Then the map*

$$x \mapsto e^{\mathbb{E}[W_0]x - \mathbb{E}[W_0]^2/2}\, \mathbb{P}\big(W_j \le x \text{ for } 1 \le j \le p \,\big|\, W_0 = x\big)$$

is non-decreasing on $\mathbb{R}$.

Lemma 1.A.4. *Let $(X_1, \dots, X_p)^\top$ be a centred Gaussian random vector in $\mathbb{R}^p$ with $\max_{1\le j\le p} \mathbb{E}[X_j^2] \le \sigma_X^2$ for some $\sigma_X^2 > 0$. Then for any $r > 0$,*

$$\mathbb{P}\Big(\max_{1\le j\le p} X_j \ge \mathbb{E}\Big[\max_{1\le j\le p} X_j \Big] + r \Big) \le e^{-r^2/(2\sigma_X^2)}.$$

The proof of Lemmas 1.A.2 and 1.A.3 can be found in Chernozhukov et al. (2015). Lemma 1.A.4 is a standard result on Gaussian concentration whose proof is given e.g. in Ledoux (2001); see Theorem 7.1 therein. We now closely follow the arguments for the proof of Theorem 3 in Chernozhukov et al. (2015). The proof splits up into three steps.

Step 1. Pick any $x \ge 0$ and set

$$W_j = \frac{X_j - x}{\sigma_j} + \frac{\overline{\mu} + x}{\underline{\sigma}}.$$

By construction, $\mathbb{E}[W_j] \ge 0$ and $\mathrm{Var}(W_j) = 1$. Defining $Z = \max_{1\le j\le p} W_j$, it holds that

$$\begin{aligned}
\mathbb{P}\Big(\Big| \max_{1\le j\le p} X_j - x \Big| \le \delta \Big) &\le \mathbb{P}\Big(\Big| \max_{1\le j\le p} \frac{X_j - x}{\sigma_j} \Big| \le \frac{\delta}{\underline{\sigma}} \Big) \\
&\le \sup_{y\in\mathbb{R}} \mathbb{P}\Big(\Big| \max_{1\le j\le p} \frac{X_j - x}{\sigma_j} + \frac{\overline{\mu} + x}{\underline{\sigma}} - y \Big| \le \frac{\delta}{\underline{\sigma}} \Big) \\
&= \sup_{y\in\mathbb{R}} \mathbb{P}\Big(|Z - y| \le \frac{\delta}{\underline{\sigma}} \Big).
\end{aligned}$$

Step 2. We now bound the density of Z. Without loss of generality, we assume that $\mathrm{Corr}(W_j, W_k) < 1$ for $k \ne j$. The marginal distribution of W_j is $N(\nu_j, 1)$ with $\nu_j = \mathbb{E}[W_j] = (\mu_j/\sigma_j + \overline{\mu}/\underline{\sigma}) + (x/\underline{\sigma} - x/\sigma_j) \ge 0$. Hence, by Lemmas 1.A.2 and 1.A.3, the random variable Z has a density of the form

$$f_p(z) = f_0(z) G_p(z), \tag{1.A.19}$$

where the map $z \mapsto G_p(z)$ is non-decreasing. Define $\overline{Z} = \max_{1\le j\le p}(W_j - \mathbb{E}[W_j])$ and set $\overline{z} = 2\overline{\mu}/\underline{\sigma} + x(1/\underline{\sigma} - 1/\overline{\sigma})$ such that $\mathbb{E}[W_j] \le \overline{z}$ for any $1 \le j \le p$. With these definitions at hand, we obtain that

$$\begin{aligned}
\int_z^\infty f_0(u)du\, G_p(z) \le \int_z^\infty f_0(u) G_p(u) du &= \mathbb{P}(Z > z) \\
&\le P(\overline{Z} > z - \overline{z}) \le \exp\Big(- \frac{(z - \overline{z} - \mathbb{E}[\overline{Z}])_+^2}{2} \Big),
\end{aligned}$$

where the last inequality follows from Lemma 1.A.4. Since $W_j - \mathbb{E}[W_j] = (X_j - \mu_j)/\sigma_j$, it holds that

$$\mathbb{E}[\overline{Z}] = \mathbb{E}\Big[\max_{1\le j\le p}\Big\{\frac{X_j-\mu_j}{\sigma_j}\Big\}\Big] =: a_p.$$

Hence, for every $z \in \mathbb{R}$,

$$G_p(z) \le \frac{1}{1-F_0(z)}\exp\Big(-\frac{(z-\overline{z}-a_p)_+^2}{2}\Big). \tag{1.A.20}$$

Mill's inequality states that for $z > 0$,

$$z \le \frac{f_0(z)}{1-F_0(z)} \le z\frac{1+z^2}{z^2}.$$

Since $(1+z^2)/z^2 \le 2$ for $z \ge 1$ and $f_0(z)/\{1-F_0(z)\} \le 1.53 \le 2$ for $z \in (-\infty, 1)$, we can infer that

$$\frac{f_0(z)}{1-F_0(z)} \le 2(z \vee 1) \quad \text{for any } z \in \mathbb{R}.$$

This together with (1.A.19) and (1.A.20) yields that

$$f_p(z) \le 2(z\vee 1)\exp\Big(-\frac{(z-\overline{z}-a_p)_+^2}{2}\Big) \quad \text{for any } z \in \mathbb{R}.$$

Step 3. By Step 2, we get that for any $y \in \mathbb{R}$ and $u > 0$,

$$\mathbb{P}(|Z-y| \le u) = \int_{y-u}^{y+u} f_p(z)dz \le 2u \max_{z\in[y-u,y+u]} f_p(z) \le 4u(\overline{z}+a_p+1),$$

where the last inequality follows from the fact that the map $z \mapsto ze^{-(z-a)^2/2}$ (with $a > 0$) is non-increasing on $[a+1, \infty)$. Combining this bound with Step 1, we further obtain that for any $x \ge 0$ and $\delta > 0$,

$$\mathbb{P}\Big(\Big|\max_{1\le j\le p} X_j - x\Big| \le \delta\Big) \le 4\delta\Big\{\frac{2\overline{\mu}}{\underline{\sigma}} + |x|\Big(\frac{1}{\underline{\sigma}} - \frac{1}{\overline{\sigma}}\Big) + a_p + 1\Big\}/\underline{\sigma}. \tag{1.A.21}$$

This inequality also holds for $x < 0$ by an analogous argument, and hence for all $x \in \mathbb{R}$.

Now let $0 < \delta \le \underline{\sigma}$ and define $b_p = \mathbb{E}\max_{1\le j\le p}\{X_j - \mu_j\}$. For any $|x| \le \delta + \overline{\mu} + b_p + \overline{\sigma}\sqrt{2\log(\underline{\sigma}/\delta)}$, (1.A.21) yields that

$$\begin{aligned}\mathbb{P}\Big(\Big|\max_{1\le j\le p} X_j - x\Big| \le \delta\Big) &\le \frac{4\delta}{\underline{\sigma}}\Big\{\overline{\mu}\Big(\frac{3}{\underline{\sigma}} - \frac{1}{\overline{\sigma}}\Big) + a_p + \Big(\frac{1}{\underline{\sigma}} - \frac{1}{\overline{\sigma}}\Big)b_p \\ &\qquad + \Big(\frac{\overline{\sigma}}{\underline{\sigma}} - 1\Big)\sqrt{2\log\Big(\frac{\overline{\sigma}}{\delta}\Big)} + 2 - \frac{\underline{\sigma}}{\overline{\sigma}}\Big\} \\ &\le C\delta\big\{\overline{\mu} + a_p + b_p + \sqrt{1 \vee \log(\underline{\sigma}/\delta)}\big\}\end{aligned} \tag{1.A.22}$$

with a sufficiently large constant $C > 0$ that depends only on $\underline{\sigma}$ and $\overline{\sigma}$. For $|x| \ge \delta + \overline{\mu} +$

$b_p + \overline{\sigma}\sqrt{2\log(\underline{\sigma}/\delta)}$, we obtain that

$$\mathbb{P}\Big(\Big|\max_{1\le j\le p} X_j - x\Big| \le \delta\Big) \le \frac{\delta}{\underline{\sigma}}, \tag{1.A.23}$$

which can be seen as follows: If $x > \delta + \overline{\mu}$, then $|\max_j X_j - x| \le \delta$ implies that $|x| - \delta \le \max_j X_j \le \max_j\{X_j - \mu_j\} + \overline{\mu}$ and thus $\max_j\{X_j - \mu_j\} \ge |x| - \delta - \overline{\mu}$. Hence, it holds that

$$\mathbb{P}\Big(\Big|\max_{1\le j\le p} X_j - x\Big| \le \delta\Big) \le \mathbb{P}\Big(\max_{1\le j\le p}\big\{X_j - \mu_j\big\} \ge |x| - \delta - \overline{\mu}\Big). \tag{1.A.24}$$

If $x < -(\delta + \overline{\mu})$, then $|\max_j X_j - x| \le \delta$ implies that $\max_j\{X_j - \mu_j\} \le -|x| + \delta + \overline{\mu}$. Hence, in this case,

$$\begin{aligned}\mathbb{P}\Big(\Big|\max_{1\le j\le p} X_j - x\Big| \le \delta\Big) &\le \mathbb{P}\Big(\max_{1\le j\le p}\big\{X_j - \mu_j\big\} \le -|x| + \delta + \overline{\mu}\Big)\\ &\le \mathbb{P}\Big(\max_{1\le j\le p}\big\{X_j - \mu_j\big\} \ge |x| - \delta - \overline{\mu}\Big),\end{aligned} \tag{1.A.25}$$

where the last inequality follows from the fact that for centred Gaussian random variables V_j and $v > 0$, $\mathbb{P}(\max_j V_j \le -v) \le \mathbb{P}(V_1 \le -v) = P(V_1 \ge v) \le \mathbb{P}(\max_j V_j \ge v)$. With (1.A.24) and (1.A.25), we obtain that for any $|x| \ge \delta + \overline{\mu} + b_p + \overline{\sigma}\sqrt{2\log(\underline{\sigma}/\delta)}$,

$$\begin{aligned}\mathbb{P}\Big(\Big|\max_{1\le j\le p} X_j - x\Big| \le \delta\Big) &\le \mathbb{P}\Big(\max_{1\le j\le p}\big\{X_j - \mu_j\big\} \ge |x| - \delta - \overline{\mu}\Big)\\ &\le \mathbb{P}\Big(\max_{1\le j\le p}\big\{X_j - \mu_j\big\} \ge \mathbb{E}\Big[\max_{1\le j\le p}\big\{X_j - \mu_j\big\}\Big] + \overline{\sigma}\sqrt{2\log(\underline{\sigma}/\delta)}\Big) \le \frac{\delta}{\underline{\sigma}},\end{aligned}$$

the last inequality following from Lemma 1.A.4. To sum up, we have established that for any $0 < \delta \le \underline{\sigma}$ and any $x \in \mathbb{R}$,

$$\mathbb{P}\Big(\Big|\max_{1\le j\le p} X_j - x\Big| \le \delta\Big) \le C\delta\big\{\overline{\mu} + a_p + b_p + \sqrt{1 \vee \log(\underline{\sigma}/\delta)}\big\} \tag{1.A.26}$$

with some constant $C > 0$ that does only depend on $\underline{\sigma}$ and $\overline{\sigma}$. For $\delta > \underline{\sigma}$, (1.A.26) trivially follows upon setting $C \ge 1/\underline{\sigma}$. This completes the proof.

1.B Proofs of the results from Section 1.4

In what follows, we prove Proposition 1.4.1 from Section 1.4. Throughout the section, we assume that m is Lipschitz and that $\{\varepsilon_t\}$ is an AR(p^*) process of the form (1.4.3) with the following properties: $A(z) \ne 0$ for all $|z| \le 1 + \delta$ with some small $\delta > 0$ and the innovations η_t have a finite fourth moment. As in the previous section, the symbol C denotes a generic real constant which may take a different value on each occurrence. Moreover, the notation $v_T \ll w_T$ means that $v_T/w_T \to 0$ as $T \to \infty$ and the symbol $\|\cdot\|_2$ denotes the usual Euclidean/spectral norm for vectors/matrices.

Auxiliary results

To start with, we derive some auxiliary results on the sample autocovariances

$$\widehat{\gamma}_q(\ell) = \frac{1}{T-q} \sum_{t=q+\ell+1}^{T} \Delta_q Y_{t,T}\, \Delta_q Y_{t-\ell,T}.$$

The first result bounds the L_2-distance between the ℓ-th sample autocovariance $\widehat{\gamma}_q(\ell)$ and its true counterpart $\gamma_q(\ell) = \mathrm{Cov}(\Delta_q \varepsilon_t, \Delta_q \varepsilon_{t-\ell})$ uniformly over ℓ.

Lemma 1.B.5. *Let $1 \le p \ll \sqrt{T}$ and $1 \le q \ll \sqrt{T}$. For any $1 \le \ell \le p$,*

$$\mathbb{E}\Big[\big(\widehat{\gamma}_q(\ell) - \gamma_q(\ell)\big)^2\Big] \le C\Big\{\frac{1}{T-q} + \Big(\frac{p}{T-q}\Big)^2 + \Big(\frac{q}{T}\Big)^2\Big\},$$

where the constant C is independent of ℓ, p, q and T.

Proof of Lemma 1.B.5. The expression $\widehat{\gamma}_q(\ell) - \gamma_q(\ell)$ can be decomposed as

$$\widehat{\gamma}_q(\ell) - \gamma_q(\ell) = \big(\widehat{\gamma}_q^*(\ell) - \gamma_q(\ell)\big) + R_{q,A}(\ell) + R_{q,B}(\ell) + R_{q,C}(\ell),$$

where

$$\widehat{\gamma}_q^*(\ell) = \frac{1}{T-q} \sum_{t=q+\ell+1}^{T} \Delta_q \varepsilon_t\, \Delta_q \varepsilon_{t-\ell}$$

and

$$R_{q,A}(\ell) = \frac{1}{T-q} \sum_{t=q+\ell+1}^{T} \Delta_q m_t \Delta_q \varepsilon_{t-\ell}$$

$$R_{q,B}(\ell) = \frac{1}{T-q} \sum_{t=q+\ell+1}^{T} \Delta_q \varepsilon_t \Delta_q m_{t-\ell}$$

$$R_{q,C}(\ell) = \frac{1}{T-q} \sum_{t=q+\ell+1}^{T} \Delta_q m_t \Delta_q m_{t-\ell}$$

with $\Delta_q m_t = m(\frac{t}{T}) - m(\frac{t-q}{T})$. In what follows, we prove that for any $1 \le \ell \le p$,

$$\mathbb{E}\Big[\big(\widehat{\gamma}_q^*(\ell) - \gamma_q(\ell)\big)^2\Big] \le C\Big\{\frac{1}{T-q} + \Big(\frac{p}{T-q}\Big)^2\Big\}, \tag{1.B.1}$$

where the constant C does not depend on ℓ, p, q and T. Applying the Cauchy-Schwarz inequality and exploiting the Lipschitz continuity of m, we further obtain that

$$\mathbb{E}R_{q,A}^2(\ell) \le \Big(\frac{1}{T-q} \sum_{t=q+\ell+1}^{T} \{\Delta_q m_t\}^2\Big)\Big(\frac{1}{T-q} \sum_{t=q+\ell+1}^{T} \{\Delta_q \varepsilon_{t-\ell}\}^2\Big) \le C\Big(\frac{q}{T}\Big)^2,$$

where C is independent of ℓ, p, q and T. Analogously, we get that $\mathbb{E}R_{q,k}^2(\ell) \le C(q/T)^2$

for $k = B, C$. From this and (1.B.1), it immediately follows that

$$\begin{aligned}\mathbb{E}\Big[\big(\widehat{\gamma}_q(\ell) - \gamma_q(\ell)\big)^2\Big] &\leq 16\Big\{\mathbb{E}\Big[\big(\widehat{\gamma}_q^*(\ell) - \gamma_q(\ell)\big)^2\Big] + \mathbb{E}R_{q,A}^2(\ell) + \mathbb{E}R_{q,B}^2(\ell) + \mathbb{E}R_{q,C}^2(\ell)\Big\} \\ &\leq C\Big\{\frac{1}{T-q} + \Big(\frac{p}{T-q}\Big)^2 + \Big(\frac{q}{T}\Big)^2\Big\},\end{aligned}$$

which completes the proof.

It remains to verify (1.B.1). To do so, we decompose $\widehat{\gamma}_q^*(\ell) - \gamma_q(\ell)$ as

$$\widehat{\gamma}_q^*(\ell) - \gamma_q(\ell) = \Sigma_{q,1}(\ell) - \Sigma_{q,2}(\ell) - \Sigma_{q,3}(\ell) + \Sigma_{q,4}(\ell) - \Big(1 - \frac{T-q-\ell}{T-q}\Big)\gamma_q(\ell)$$

with

$$\begin{aligned}\Sigma_{q,1}(\ell) &= \frac{1}{T-q}\sum_{t=q+\ell+1}^{T}\big\{\varepsilon_t\varepsilon_{t-\ell} - \mathbb{E}\varepsilon_t\varepsilon_{t-\ell}\big\} \\ \Sigma_{q,2}(\ell) &= \frac{1}{T-q}\sum_{t=q+\ell+1}^{T}\big\{\varepsilon_t\varepsilon_{t-\ell-q} - \mathbb{E}\varepsilon_t\varepsilon_{t-\ell-q}\big\} \\ \Sigma_{q,3}(\ell) &= \frac{1}{T-q}\sum_{t=q+\ell+1}^{T}\big\{\varepsilon_{t-q}\varepsilon_{t-\ell} - \mathbb{E}\varepsilon_{t-q}\varepsilon_{t-\ell}\big\} \\ \Sigma_{q,4}(\ell) &= \frac{1}{T-q}\sum_{t=q+\ell+1}^{T}\big\{\varepsilon_{t-q}\varepsilon_{t-\ell-q} - \mathbb{E}\varepsilon_{t-q}\varepsilon_{t-\ell-q}\big\}\end{aligned}$$

and prove that

$$\mathbb{E}\Sigma_{q,k}^2(\ell) \leq \frac{C}{T-q} \tag{1.B.2}$$

for $1 \leq k \leq 4$, where the constant C only depends on the coefficients $c_0, c_1, c_2, \ldots$ of the MA(∞) representation of $\{\varepsilon_t\}$ and the innovation variance ν^2. From this, we get that

$$\begin{aligned}\mathbb{E}\Big[\big(\widehat{\gamma}_q^*(\ell) - \gamma_q(\ell)\big)^2\Big] &\leq 25\Big\{\sum_{k=1}^{4}\mathbb{E}\Sigma_{q,k}^2(\ell) + \Big(1 - \frac{T-q-\ell}{T-q}\Big)^2\gamma_q^2(\ell)\Big\} \\ &\leq C\Big\{\frac{1}{T-q} + \Big(\frac{p}{T-q}\Big)^2\Big\}\end{aligned}$$

with C independent of ℓ, p, q and T, which shows (1.B.1).

We now turn to the proof of (1.B.2). As the proof is completely analogous for $k \in \{1, \ldots, 4\}$, we restrict attention to $k = 1$. Since the variables ε_t have the MA(∞) expansion $\varepsilon_t = \sum_{k=0}^{\infty} c_k\eta_{t-k}$ and the autocovariances $\gamma_\varepsilon(\ell) = \mathrm{Cov}(\varepsilon_t, \varepsilon_{t-\ell})$ can be written as $\gamma_\varepsilon(\ell) =$

$(\sum_{k=0}^{\infty} c_k c_{k+\ell})\nu^2$, it holds that

$$\begin{aligned}
&\mathbb{E}\big[\varepsilon_t \varepsilon_{t-\ell} \varepsilon_{t'} \varepsilon_{t'-\ell}\big] \\
&= \Big(\sum_{k=0}^{\infty} c_k\, c_{k+t'-t}\, c_{k+\ell}\, c_{k+\ell+t'-t}\Big)\kappa + \Big(\sum_{k=0}^{\infty} c_k\, c_{k+\ell}\Big)^2 \nu^4 \\
&\quad + \Big(\sum_{k=0}^{\infty} c_k\, c_{k+t'-t}\Big)^2 \nu^4 + \Big(\sum_{k=0}^{\infty} c_k\, c_{k+\ell+t'-t}\Big)\Big(\sum_{k=0}^{\infty} c_k\, c_{k+\ell+t-t'}\Big)\nu^4 \\
&= \Big(\sum_{k=0}^{\infty} c_k\, c_{k+t'-t}\, c_{k+\ell}\, c_{k+\ell+t'-t}\Big)\kappa + \gamma_\varepsilon^2(\ell) + \gamma_\varepsilon^2(t'-t) + \gamma_\varepsilon(t'-t+\ell)\gamma_\varepsilon(t'-t-\ell)
\end{aligned}$$

with $\kappa = \mathbb{E}[\eta_0^4] - 3\nu^4$ and $c_k = 0$ for $k < 0$. From this, we obtain that $\mathbb{E}\Sigma_{q,1}^2(\ell) = \sum_{k=1}^{3} \mathbb{E}\Sigma_{q,1,k}^2(\ell)$ with

$$\begin{aligned}
\mathbb{E}\Sigma_{q,1,1}^2(\ell) &= \frac{\kappa}{(T-q)^2} \sum_{t,t'=q+\ell+1}^{T} \Big(\sum_{k=0}^{\infty} c_k\, c_{k+t'-t}\, c_{k+\ell}\, c_{k+\ell+t'-t}\Big) \\
\mathbb{E}\Sigma_{q,1,2}^2(\ell) &= \frac{1}{(T-q)^2} \sum_{t,t'=q+\ell+1}^{T} \gamma_\varepsilon^2(t'-t) \\
\mathbb{E}\Sigma_{q,1,3}^2(\ell) &= \frac{1}{(T-q)^2} \sum_{t,t'=q+\ell+1}^{T} \gamma_\varepsilon(t'-t+\ell)\gamma_\varepsilon(t'-t-\ell).
\end{aligned}$$

Let $\#\{t'-t=r\}$ be the number of pairs (t,t') with $q+1 \le t,t' \le T$ such that $t'-t=r$ and note that $\#\{t'-t=r\} \le T-q$ for any r. It holds that

$$\begin{aligned}
|\mathbb{E}\Sigma_{q,1,1}^2(\ell)| &\le \frac{\kappa}{(T-q)^2} \sum_{r=-T}^{T} \#\big\{t'-t=r\big\} \sum_{k=0}^{\infty} |c_k\, c_{k+r}\, c_{k+\ell}\, c_{k+\ell+r}| \\
&\le \frac{\kappa}{T-q} \sum_{k=0}^{\infty} |c_k\, c_{k+\ell}| \sum_{r=-\infty}^{\infty} |c_{k+r}\, c_{k+\ell+r}| \\
&\le \frac{\kappa\{\max_j |c_j|\}^2\{\sum_{k=0}^{\infty} |c_k|\}^2}{T-q} \le \frac{C}{T-q},
\end{aligned} \tag{1.B.3}$$

where C only depends on the parameters $c_0, c_1, c_2, \ldots$ of the MA(∞) representation of $\{\varepsilon_t\}$. Moreover,

$$\mathbb{E}\Sigma_{q,1,2}^2(\ell) \le \frac{1}{(T-q)^2} \sum_{r=-T}^{T} \#\big\{t'-t=r\big\}\gamma_\varepsilon^2(r) \le \frac{\{\sum_{r=-\infty}^{\infty} \gamma_\varepsilon^2(r)\}}{T-q} \le \frac{C}{T-q} \tag{1.B.4}$$

and analogously $|\mathbb{E}\Sigma_{q,1,3}^2(\ell)| \le C/(T-q)$, where C only depends on the MA parameters $c_0, c_1, c_2, \ldots$ and the innovation variance ν^2 (noting that $\gamma_\varepsilon(k) = \sum_{j=0}^{\infty} c_j c_{j+k} \nu^2$). With (1.B.3) and (1.B.4), we immediately arrive at (1.B.2). □

With the help of Lemma 1.B.5, we can derive the following bounds on $\mathbb{E}\|\widehat{\boldsymbol{\gamma}}_q - \boldsymbol{\gamma}_q\|_2$

and $\mathbb{E}\|\widehat{\mathbf{\Gamma}}_q - \mathbf{\Gamma}_q\|_2$, where $\widehat{\boldsymbol{\gamma}}_q = (\widehat{\gamma}_q(1), \dots, \widehat{\gamma}_q(p))^\top$ and $\boldsymbol{\gamma}_q = (\gamma_q(1), \dots, \gamma_q(p))^\top$ as well as $\widehat{\mathbf{\Gamma}}_q = (\widehat{\gamma}_q(i-j) : 1 \le i, j \le p)$ and $\mathbf{\Gamma}_q = (\gamma_q(i-j) : 1 \le i, j \le p)$.

Lemma 1.B.6. *Let $1 \le p \ll \sqrt{T}$ and $1 \le q \ll \sqrt{T}$. It holds that*

$$\mathbb{E}\|\widehat{\boldsymbol{\gamma}}_q - \boldsymbol{\gamma}_q\|_2 \le C\sqrt{p}\Big\{\frac{1}{T-q} + \Big(\frac{p}{T-q}\Big)^2 + \Big(\frac{q}{T}\Big)^2\Big\}^{1/2}$$
$$\mathbb{E}\|\widehat{\mathbf{\Gamma}}_q - \mathbf{\Gamma}_q\|_2 \le Cp\Big\{\frac{1}{T-q} + \Big(\frac{p}{T-q}\Big)^2 + \Big(\frac{q}{T}\Big)^2\Big\}^{1/2}$$

with some constant C independent of p, q and T.

Proof of Lemma 1.B.6. The first statement immediately follows from Lemma 1.B.5. The second statement is obtained by using Lemma 1.B.5 and the bound

$$\|\widehat{\mathbf{\Gamma}}_q - \mathbf{\Gamma}_q\|_2 \le \max_{1\le i\le p}\Big(\sum_{j=1}^{p}\big|\widehat{\gamma}_q(i-j) - \gamma_q(i-j)\big|\Big) \le \sum_{\ell=-p}^{p}\big|\widehat{\gamma}_q(\ell) - \gamma_q(\ell)\big|,$$

which follows from Gershgorin's theorem. □

The final auxiliary result summarizes some properties of the inverse autocovariance matrices $\widehat{\mathbf{\Gamma}}_q^{-1}$ and $\mathbf{\Gamma}_q^{-1}$.

Lemma 1.B.7. *Let $p \to \infty$ and $(1+\delta)p \le q \ll \sqrt{T}$ for some small $\delta > 0$. Then*

(i) $\|\mathbf{\Gamma}_q^{-1}\|_2 \le C$ *for sufficiently large T with C independent of p, q and T*

(ii) $\|\widehat{\mathbf{\Gamma}}_q^{-1} - \mathbf{\Gamma}_q^{-1}\|_2 = O_p(p/\sqrt{T})$

(iii) $\|\widehat{\mathbf{\Gamma}}_q^{-1}\|_2 = O_p(1)$.

Proof of Lemma 1.B.7. We first prove (i). As $\gamma_q(\ell) = 2\gamma_\varepsilon(\ell) - \gamma_\varepsilon(q-\ell) - \gamma_\varepsilon(q+\ell)$ with $\gamma_\varepsilon(\ell) = \mathrm{Cov}(\varepsilon_t, \varepsilon_{t-\ell})$, it holds that

$$\mathbf{\Gamma}_q = 2\mathbf{\Gamma} - \boldsymbol{R},$$

where $\mathbf{\Gamma} = (\gamma_\varepsilon(i-j) : 1 \le i, j \le p)$ and $\boldsymbol{R} = (\gamma_\varepsilon(q+i-j) + \gamma_\varepsilon(q-i+j) : 1 \le i, j \le p)$. Since the spectral density f_ε of the AR(p^*) process $\{\varepsilon_t\}$ is bounded away from zero and infinity, we can use Proposition 4.5.3 in Brockwell and Davis (1991) to obtain that the eigenvalues of the autocovariance matrix $\mathbf{\Gamma}$ lie in some interval $[c_\Gamma, C_\Gamma]$ with constants $0 < c_\Gamma \le C_\Gamma < \infty$ that are independent of p. From this, we can infer that

$$\|\mathbf{\Gamma}^{-1}\|_2 \le C \text{ with some constant } C \text{ independent of } p. \tag{1.B.5}$$

Moreover, it holds that

$$\|\mathbf{\Gamma}_q/2 - \mathbf{\Gamma}\|_2 = \|\boldsymbol{R}/2\|_2 \le \sum_{\ell=-p}^{p}|\gamma_\varepsilon(q-\ell) + \gamma_\varepsilon(q+\ell)|/2 \le Cp\xi^{q-p} = o(1), \tag{1.B.6}$$

where we have used Gershgorin's theorem and the fact that $|\gamma_\varepsilon(\ell)| \leq C\xi^\ell$ with some $\xi \in (0,1)$. We now make use of the inequality

$$\|\boldsymbol{A}^{-1} - \boldsymbol{B}^{-1}\|_2 \leq \frac{\|\boldsymbol{B}^{-1}\|_2^2 \, \|\boldsymbol{B} - \boldsymbol{A}\|_2}{1 - \|\boldsymbol{B} - \boldsymbol{A}\|_2 \, \|\boldsymbol{B}^{-1}\|_2}, \tag{1.B.7}$$

which holds for general invertible matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ and which is a simple consequence of the fact that $\boldsymbol{A}^{-1} - \boldsymbol{B}^{-1} = (\boldsymbol{A}^{-1} - \boldsymbol{B}^{-1} + \boldsymbol{B}^{-1})(\boldsymbol{B} - \boldsymbol{A})\boldsymbol{B}^{-1}$. Setting $\boldsymbol{A} = \boldsymbol{\Gamma}_q/2$ and $\boldsymbol{B} = \boldsymbol{\Gamma}$ in (1.B.7) and applying (1.B.5) together with (1.B.6), we get that

$$\|(\boldsymbol{\Gamma}_q/2)^{-1} - \boldsymbol{\Gamma}^{-1}\|_2 \leq \frac{\|\boldsymbol{\Gamma}^{-1}\|_2^2 \, \|\boldsymbol{\Gamma}_q/2 - \boldsymbol{\Gamma}\|_2}{1 - \|\boldsymbol{\Gamma}_q/2 - \boldsymbol{\Gamma}\|_2 \, \|\boldsymbol{\Gamma}^{-1}\|_2} \leq C \tag{1.B.8}$$

for sufficiently large T with C independent of p, q and T. Statement (i) easily follows upon combining (1.B.5) and (1.B.8).

We next turn to the proof of (ii). With (1.B.7), we obtain that

$$\|\widehat{\boldsymbol{\Gamma}}_q^{-1} - \boldsymbol{\Gamma}_q^{-1}\|_2 \leq \frac{\|\boldsymbol{\Gamma}_q^{-1}\|_2^2 \, \|\widehat{\boldsymbol{\Gamma}}_q - \boldsymbol{\Gamma}_q\|_2}{1 - \|\widehat{\boldsymbol{\Gamma}}_q - \boldsymbol{\Gamma}_q\|_2 \, \|\boldsymbol{\Gamma}_q^{-1}\|_2} = O_p\Big(\frac{p}{\sqrt{T}}\Big),$$

since $\|\boldsymbol{\Gamma}_q^{-1}\|_2 \leq C$ by (i) and $\|\widehat{\boldsymbol{\Gamma}}_q - \boldsymbol{\Gamma}_q\|_2 = O_p(p/\sqrt{T})$ by Lemma 1.B.6. Finally, (iii) is an immediately consequence of (i) and (ii). □

Proof of Proposition 1.4.1

We have to prove the following three statements:

$$\|\widetilde{\boldsymbol{a}}_q - \boldsymbol{a}\|_2 = O_p\Big(\sqrt{\frac{p}{T}}\Big) \tag{1.B.9}$$

$$\|\widehat{\boldsymbol{a}} - \boldsymbol{a}\|_2 = O_p\Big(\sqrt{\frac{p^3}{T}}\Big) \tag{1.B.10}$$

$$\widehat{\sigma}^2 - \sigma^2 = O_p\Big(\sqrt{\frac{p^4}{T}}\Big). \tag{1.B.11}$$

We carry out the proof for case (B), that is, we impose the following conditions on the parameters p, q, $\underline{r}$ and $\overline{r}$: $q \ll \sqrt{T}$, $C \log T \leq p \ll \min\{T^{1/5}, q\}$ for some sufficiently large C, $\underline{r} = (1+\delta)p$ for some small $\delta > 0$ and $\overline{r} - \underline{r}$ remains bounded as $T \to \infty$. The proof for case (A) is a simplified version of that for case (B).

Proof of (1.B.9). It holds that

$$\begin{aligned}\widetilde{\boldsymbol{a}}_q - \boldsymbol{a} = \widehat{\boldsymbol{\Gamma}}_q^{-1}\widehat{\boldsymbol{\gamma}}_q - \boldsymbol{a} &= \widehat{\boldsymbol{\Gamma}}_q^{-1}\big[(\widehat{\boldsymbol{\gamma}}_q - \boldsymbol{\gamma}_q) + (\boldsymbol{\gamma}_q - \boldsymbol{\Gamma}_q\boldsymbol{a}) + (\boldsymbol{\Gamma}_q - \widehat{\boldsymbol{\Gamma}}_q)\boldsymbol{a}\big] \\ &= \widehat{\boldsymbol{\Gamma}}_q^{-1}\big[(\widehat{\boldsymbol{\gamma}}_q - \boldsymbol{\gamma}_q) - \nu^2\boldsymbol{c}_q + \boldsymbol{\rho}_q + (\boldsymbol{\Gamma}_q - \widehat{\boldsymbol{\Gamma}}_q)\boldsymbol{a}\big]\end{aligned}$$

and thus

$$\|\widetilde{\boldsymbol{a}}_q - \boldsymbol{a}\|_2 \leq \|\widehat{\boldsymbol{\Gamma}}_q^{-1}\|_2 \Big[\|\widehat{\boldsymbol{\gamma}}_q - \boldsymbol{\gamma}_q\|_2 + \nu^2 \|\boldsymbol{c}_q\|_2 + \|\boldsymbol{\rho}_q\|_2 + \|(\boldsymbol{\Gamma}_q - \widehat{\boldsymbol{\Gamma}}_q)\boldsymbol{a}\|_2 \Big], \tag{1.B.12}$$

where we have used that $\boldsymbol{\gamma}_q - \boldsymbol{\Gamma}_q \boldsymbol{a} = -\nu^2 \boldsymbol{c}_q + \boldsymbol{\rho}_q$. We now make use of the following facts: (i) By Lemmas 1.B.6 and 1.B.7, it holds that $\|\widehat{\boldsymbol{\gamma}}_q - \boldsymbol{\gamma}_q\|_2 = O_p(\sqrt{p/T})$ and $\|\widehat{\boldsymbol{\Gamma}}_q^{-1}\|_2 = O_p(1)$. (ii) Since $\boldsymbol{c}_q = (c_{q-1}, \ldots, c_{q-p})^\top$ and $|c_j| \leq C\xi^j$ for some $\xi \in (0,1)$, it holds that $\|\boldsymbol{c}_q\|_2 \leq C\sqrt{p}\xi^{q-p}$. Similarly, as $\boldsymbol{\rho}_q = (\rho_q(1), \ldots, \rho_q(p))^\top$ with $\rho_q(\ell) = \sum_{j=p+1}^{p^*} a_j \gamma_q(\ell - j)$, we have $|\rho_q(\ell)| \leq C \sum_{j=p+1}^{p^*} \xi^j \leq C\xi^p$ and thus $\|\boldsymbol{\rho}_q\|_2 \leq C\sqrt{p}\xi^p$. (iii) It holds that $\|(\boldsymbol{\Gamma}_q - \widehat{\boldsymbol{\Gamma}}_q)\boldsymbol{a}\|_2 = O_p(\sqrt{p/T})$, which can be seen as follows:

$$\begin{aligned}
&\|(\boldsymbol{\Gamma}_q - \widehat{\boldsymbol{\Gamma}}_q)\boldsymbol{a}\|_2^2 \\
&= \sum_{i,j,j'=1}^{p} \big\{\gamma_q(i-j) - \widehat{\gamma}_q(i-j)\big\}\big\{\gamma_q(i-j') - \widehat{\gamma}_q(i-j')\big\} a_j a_{j'} \\
&\leq \sum_{j,j'=1}^{p} |a_j a_{j'}| \Big(\sum_{i=1}^{p} \big\{\gamma_q(i-j) - \widehat{\gamma}_q(i-j)\big\}^2 \Big)^{1/2} \Big(\sum_{i=1}^{p} \big\{\gamma_q(i-j') - \widehat{\gamma}_q(i-j')\big\}^2 \Big)^{1/2} \\
&\leq \Big(\max_{1\leq j\leq p} \sum_{i=1}^{p} \big\{\gamma_q(i-j) - \widehat{\gamma}_q(i-j)\big\}^2 \Big) \Big(\sum_{j,j'=1}^{p} |a_j a_{j'}| \Big) \\
&\leq C \sum_{\ell=-p}^{p} \big\{\gamma_q(\ell) - \widehat{\gamma}_q(\ell)\big\}^2 = O_p\Big(\frac{p}{T}\Big),
\end{aligned}$$

where the last equality follows by Lemma 1.B.5. Plugging (i)–(iii) into (1.B.12), we arrive at

$$\|\widetilde{\boldsymbol{a}}_q - \boldsymbol{a}\|_2 = O_p\Big(\sqrt{\frac{p}{T}} + \sqrt{p}\xi^{q-p} + \sqrt{p}\xi^p\Big) = O_p\Big(\sqrt{\frac{p}{T}}\Big),$$

which completes the proof. □

Proof of (1.B.10). It suffices to show that $\|\widehat{\boldsymbol{a}}_r - \boldsymbol{a}\|_2 = O_p(\sqrt{p^3/T})$ for $r = (1+\delta)p$ with some small $\delta > 0$. By the same arguments as for (1.B.9), we obtain the bound

$$\begin{aligned}
\|\widehat{\boldsymbol{a}}_r - \boldsymbol{a}\|_2 &\leq \|\widehat{\boldsymbol{\Gamma}}_r^{-1}\|_2 \Big[\|\widehat{\boldsymbol{\gamma}}_r - \boldsymbol{\gamma}_r\|_2 + \|\widetilde{\nu}^2 \widetilde{\boldsymbol{c}}_r - \nu^2 \boldsymbol{c}_r\|_2 + \|\boldsymbol{\rho}_r\|_2 + \|(\boldsymbol{\Gamma}_r - \widehat{\boldsymbol{\Gamma}}_r)\boldsymbol{a}\|_2 \Big] \\
&= \|\widehat{\boldsymbol{\Gamma}}_r^{-1}\|_2 \, \|\widetilde{\nu}^2 \widetilde{\boldsymbol{c}}_r - \nu^2 \boldsymbol{c}_r\|_2 + O_p\Big(\sqrt{\frac{p}{T}}\Big).
\end{aligned} \tag{1.B.13}$$

Straightforward but lengthy calculations yield that $\widetilde{\nu}^2 - \nu^2 = O_p(\sqrt{p/T})$. Moreover, as proven below,

$$\max_{1\leq j\leq p} |\widetilde{c}_j - c_j| = O_p\Big(\frac{p}{\sqrt{T}}\Big), \tag{1.B.14}$$

which implies that $\|\widetilde{\nu}^2 \widetilde{\boldsymbol{c}}_r - \nu^2 \boldsymbol{c}_r\|_2 = O_p(\sqrt{p^3/T})$. Plugging this into (1.B.13) completes the proof of (1.B.10).

It remains to verify (1.B.14). By assumption, $A(z) \neq 0$ for all complex $|z| \leq 1 + \delta$ with some small $\delta > 0$. Hence, $A^{-1}(z)$ has a (unique) series expansion of the form

$A^{-1}(z) = \sum_{j=0}^{\infty} c_j z^j$ for all $|z| < 1 + \delta$, where the coefficients c_j satisfy the integral representation

$$c_j = \frac{1}{2\pi i} \oint_{|w|=1} \frac{A^{-1}(w)}{w^{j+1}} dw \qquad \text{for all } j \geq 0. \tag{1.B.15}$$

Here and in what follows, the symbol $\oint_{|w|=1}$ denotes the line integral along the unit circle. An analogous integral representation can be derived for the estimates $\widetilde{c}_j$: By classical time series results, $\widetilde{A}(z) = 1 - \sum_{j=1}^{p} \widetilde{a}_j z^j \neq 0$ for all complex $|z| \leq 1$ whenever $\widehat{\gamma}_q(0) > 0$. Since $\widetilde{A}(z)$ has exactly p (and thus finitely many) roots, it even holds that $\widetilde{A}(z) \neq 0$ for all $|z| \leq 1 + \widetilde{\delta}$ with some small $\widetilde{\delta} > 0$ whenever $\widehat{\gamma}_q(0) > 0$. Note that $\widetilde{\delta} = \widetilde{\delta}(\widetilde{a}_1, \ldots, \widetilde{a}_p)$ is a function of the parameter estimates $\widetilde{a}_1, \ldots, \widetilde{a}_p$ and thus random. As $\widehat{\gamma}_q(0) > 0$ with probability approaching 1, we can infer that $\widetilde{A}(z) \neq 0$ for all $|z| \leq 1 + \widetilde{\delta}$ with probability tending to 1. This in turn implies that with probability tending to 1, $\widetilde{A}^{-1}(z)$ has a (unique) series expansion of the form $\widetilde{A}^{-1}(z) = \sum_{j=0}^{\infty} \widetilde{\phi}_j z^j$ for all $|z| < 1 + \widetilde{\delta}$, where the coefficients $\widetilde{\phi}_j$ satisfy the integral representation $\widetilde{\phi}_j = (2\pi i)^{-1} \oint_{|w|=1} \widetilde{A}^{-1}(w)/w^{j+1} dw$ for all $j \geq 0$. By definition, the estimates $\widetilde{c}_j$ are the solution to the recursive equations (1.4.2). Importantly, this is equivalent to them being the coefficients in the series expansion of $A^{-1}(z)$. Hence, $\widetilde{c}_j = \widetilde{\phi}_j$ for all $j \geq 0$ and thus

$$\widetilde{c}_j = \frac{1}{2\pi i} \oint_{|w|=1} \frac{\widetilde{A}^{-1}(w)}{w^{j+1}} dw \qquad \text{for all } j \geq 0 \tag{1.B.16}$$

with probability tending to 1. Combining (1.B.15) and (1.B.16), we finally obtain that

$$\max_{1 \leq j \leq p} |\widetilde{c}_j - c_j| = \max_{1 \leq j \leq p} \Big| \frac{1}{2\pi i} \oint_{|w|=1} \frac{\widetilde{A}^{-1}(w) - A^{-1}(w)}{w^{j+1}} dw \Big|$$

with probability tending to 1. The right-hand side of this formula can be bounded with the help of (1.B.9) and straightforward arguments to arrive at (1.B.14). □

Proof of (1.B.11). With the help of (1.B.10) and straightforward but lengthy calculations, it can be shown that $\widehat{\nu}^2 = \nu^2 + O_p(\sqrt{p^3/T})$ and $(1 - \sum_{j=1}^{p} \widehat{a}_j)^2 - (1 - \sum_{j=1}^{p^*} a_j)^2 = O_p(\sqrt{p^4/T})$. The statement (1.B.11) is a simple consequence of these two facts. □

1.C Supplementary material for the simulation study of Section 1.5

1.C.1 Implementation of SiZer in Section 1.5.1

(a) Computation of the grid $\mathcal{G}_T^*$:

To start with, we compute the variance of $\bar{Y} = T^{-1} \sum_{t=1}^{T} Y_{t,T}$, which is given by

$$\mathrm{Var}(\bar{Y}) = \frac{\gamma_\varepsilon(0)}{T} + \frac{2}{T} \sum_{k=1}^{T-1} \Big(1 - \frac{k}{T}\Big) \gamma_\varepsilon(k).$$

Since the autocovariance function $\gamma_\varepsilon(\cdot)$ of the error process $\{\varepsilon_t\}$ is assumed to be known, we can calculate the value of $\mathrm{Var}(\bar{Y})$ by using the formula $\gamma_\varepsilon(k) = \nu^2 a_1^{|k|}/(1 - a_1^2)$ together with the true parameters a_1 and $\nu^2 = \mathbb{E}[\eta_t^2]$. We next compute

$$T^* = \frac{\gamma_\varepsilon(0)}{\mathrm{Var}(\bar{Y})},$$

which can be interpreted as a measure of information in the data. For each point $(u, h) \in \mathcal{G}_T$, we finally calculate the effective sample size for dependent data

$$\mathrm{ESS}^*(u, h) = \frac{T^*}{T} \frac{\sum_{t=1}^T K_h(t/T - u)}{K_h(0)}$$

with $K_h(v) = h^{-1}K(v/h)$ and set $\mathcal{G}_T^* = \{(u, h) \in \mathcal{G}_T : \mathrm{ESS}^*(u, h) \geq 5\}$.

(b) Computation of the local linear estimator and its standard deviation:

For each $(u, h) \in \mathcal{G}_T^*$, we compute a standard local linear estimator $\widehat{m}'_h(u)$ of the derivative $m'(u)$ together with its standard deviation $\mathrm{sd}(\widehat{m}'_h(u))$. The latter is given by $\mathrm{sd}(\widehat{m}'_h(u)) = \{\mathrm{Var}(\widehat{m}'_h(u))\}^{1/2}$, where $\mathrm{Var}(\widehat{m}'_h(u)) = e^\top V e$ with $e = (0 \quad 1)^\top$ and

$$V = (X^T W X)^{-1}(X^T \Sigma X)(X^T W X)^{-1}.$$

The matrices X, W and Σ are defined as follows: Σ is a $T \times T$ matrix with the elements

$$\Sigma_{st} = \gamma_\varepsilon(s - t) K_h\Big(\frac{s}{T} - u\Big) K_h\Big(\frac{t}{T} - u\Big),$$

W is a $T \times T$ diagonal matrix with the diagonal entries $K_h(t/T - u)$ and

$$X = \begin{pmatrix} 1 & (1/T - u) \\ 1 & (2/T - u) \\ \vdots & \vdots \\ 1 & (1 - u) \end{pmatrix}.$$

(c) Computation of the critical values:

For a given significance level α and for each bandwidth h with $(u, h) \in \mathcal{G}_T^*$, we compute the term

$$q(h) = \Phi^{-1}\Big(\Big(1 - \frac{\alpha}{2}\Big)^{1/(\theta g)}\Big),$$

where Φ is the distribution function of a standard normal random variable, g is the number of locations u with $(u, h) \in \mathcal{G}_T^*$, and the cluster index θ is defined on p.1519 in Park et al. (2009a). The SiZer test is carried out as follows: For each $(u, h) \in \mathcal{G}_T^*$, the test rejects the null hypothesis $H_0(u, h)$ that m is constant on $[u - h, u + h]$ if $|\widehat{m}'_h(u)/\mathrm{sd}(\widehat{m}'_h(u))| > q(h)$. More specifically, the test indicates an increase in m on the interval $[u-h, u+h]$ if $\widehat{m}'_h(u)/\mathrm{sd}(\widehat{m}'_h(u)) > q(h)$ and a decrease if $-\widehat{m}'_h(u)/\mathrm{sd}(\widehat{m}'_h(u)) >$

$q(h)$.

1.C.2 Power simulations additional to Section 1.5.1.2

In the following simulation exercises, we compare the performance of the tests $\mathcal{T}_{\mathrm{MS}}$, $\mathcal{T}_{\mathrm{UC}}$, $\mathcal{T}_{\mathrm{RW}}$ and $\mathcal{T}_{\mathrm{SiZer}}$ (i) when m is the blocks signal of Donoho and Johnstone (1995) that was investigated in detail by Hannig and Marron (2006) in the SiZer context and (ii) when m is the sine curve $m(u) = \sin(6\pi u)$ that was considered in Park et al. (2009a). We define the blocks signal exactly as Marron et al. (1998) and Hannig and Marron (2006). Specifically, we set

$$\begin{aligned} m(x) = (0.6/9.2)\big\{&4\mathrm{ssgn}(x-0.1) - 5\mathrm{ssgn}(x-0.13) + 3\mathrm{ssgn}(x-0.15) \\ &- 4\mathrm{ssgn}(x-0.23) + 5\mathrm{ssgn}(x-0.25) - 4.2\mathrm{ssgn}(x-0.4) \\ &+ 2.1\mathrm{ssgn}(x-0.44) + 4.3\mathrm{ssgn}(x-0.65) - 3.1\mathrm{ssgn}(x-0.76) \\ &+ 2.1\mathrm{ssgn}(x-0.78) - 4.2\mathrm{ssgn}(x-0.81) + 2\big\} + 0.2, \end{aligned}$$

where $\mathrm{ssgn}(x) = (1 + \mathrm{sgn}(x))/2$ and $\mathrm{sgn}(x)$ is the standard sign function. In both the blocks and the sine case, we model the error terms as an AR(1) process $\varepsilon_t = a_1\varepsilon_{t-1} + \eta_t$, where $a_1 \in \{-0.5, 0.5\}$ and η_t are i.i.d. normal with $\mathbb{E}[\eta_t] = 0$ and $\mathbb{E}[\eta_t^2] = \nu^2$. In the blocks example, we set $\nu^2 = (1 - a_1^2)/100$. This implies that $\mathrm{Var}(\varepsilon_t) = (0.1)^2$, which matches the variance of the i.i.d. errors in the blocks example of Hannig and Marron (2006). In the sine example, we choose $\nu^2 = (1 - a_1^2)$, which implies that $\mathrm{Var}(\varepsilon_t) = 1$. A plot of the blocks signal is given in the two top panels of Figure 1.C.1. As can be seen, the signal is a piecewise constant function with several jumps. We could replace this jump function by a slightly smoothed and thus differentiable version with very steep increases and decreases. However, as this would leave the simulation results essentially unchanged, we stick to the original blocks signal.

For both the blocks and the sine example, we simulate a representative data sample of length $T = 1000$ and carry out the four tests on the simulated sample for the significance level $\alpha = 0.05$. The results are presented by SiZer maps in Figures 1.C.1 and 1.C.2 which are to be read as follows: Each pixel of the SiZer map corresponds to a location-scale point (u, h), or put differently, to a time interval $[u-h, u+h]$. The pixel (u, h) is coloured blue if the test indicates an increase in the trend m on the interval $[u - h, u + h]$, red if the test indicates a decrease and purple if the test does not reject the null hypothesis that m is constant on $[u - h, u + h]$. Moreover, a pixel (u, h) is coloured grey if the effective sample size $\mathrm{ESS}^*(u, h)$ is smaller than 5, in which case the pixel (u, h) is not included in the location-scale grid $\mathcal{G}_T^*$.

The results for the blocks example are reported in Figure 1.C.1, the left-hand panels of subfigure (a) corresponding to the case with $a_1 = -0.5$ and the right-hand panels of subfigure (b) to the case with $a_1 = 0.5$. Let us first have a closer look at subfigure (a). The top panel depicts the blocks signal with the simulated data sample in the background. The

other panels show the SiZer maps produced by the four tests $\mathcal{T}_{\text{MS}}$, $\mathcal{T}_{\text{UC}}$, $\mathcal{T}_{\text{RW}}$ and $\mathcal{T}_{\text{SiZer}}$. As can be seen, the SiZer maps are fairly similar. In particular, all four tests pick up the increases and decreases (that is, the upward and downward jumps) in the signal m quite accurately. The situation is a bit different in subfigure (b), that is, in the case with $a_1 = 0.5$. Overall, the colour patterns in the four SiZer maps look fairly similar. However, on closer inspection, the following differences become apparent:

(i) In the SiZer maps of the two row-wise methods $\mathcal{T}_{\text{RW}}$ and $\mathcal{T}_{\text{SiZer}}$, there is a small stripe of blue pixels around the jump location $u = 0.44$. Hence, the row-wise methods detect the small upward jump in the blocks signal at $u = 0.44$, whereas the global methods do not pick up this jump.

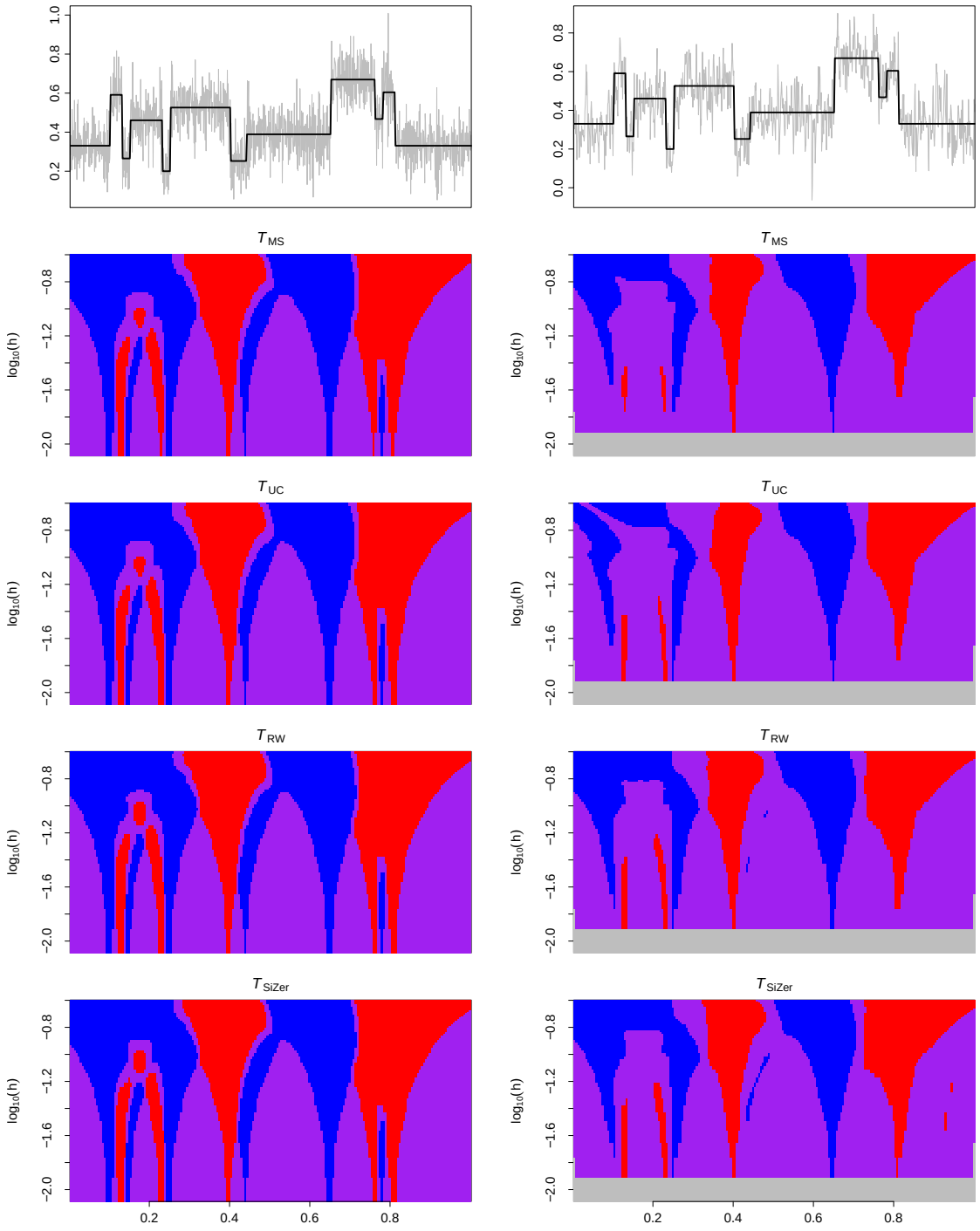
(ii) In the SiZer map of $\mathcal{T}_{\text{SiZer}}$, there are two small stripes of red pixels near the location $u = 0.95$ corresponding to scales h with $\log_{10}(h)$ between -1.2 and -1.6. Hence, the row-wise SiZer test $\mathcal{T}_{\text{SiZer}}$ spuriously finds a decrease in the trend m on a short time interval around $u = 0.95$. As a specific example, the pixel $(u, h) = (0.93, 0.03)$ is coloured red, implying that $\mathcal{T}_{\text{SiZer}}$ spuriously finds a decrease on the interval $[0.90, 0.96]$. Inspecting the grey time series plot in the top panel of subfigure (b), it indeed looks as if there is a short downward trend in the time series towards the end of the sample. However, this downward movement of the time series is not due to an actual decrease in the trend function m. It is rather produced by the autocorrelation structure in the error terms.

(i) indicates that the row-wise methods $\mathcal{T}_{\text{RW}}$ and $\mathcal{T}_{\text{SiZer}}$ tend to be more powerful than the global tests $\mathcal{T}_{\text{MS}}$ and $\mathcal{T}_{\text{UC}}$. (ii) shows that this gain of power comes at a cost: The row-wise methods tend to find spurious increases/decreases more often than the global ones. Hence, the SiZer maps of subfigure (b) nicely illustrate the main findings of our power simulations in Section 1.5.1.2.

The SiZer maps for the sine example are depicted in Figure 1.C.2. Overall, they convey a picture very similar to the SiZer maps of the blocks example. The SiZer maps for the case with $a_1 = -0.5$ show that the increases and decreases of the sine curve m are picked up appropriately by all four tests. In the case with $a_1 = 0.5$, in contrast, the four tests only detect the increases and decreases of m in the interior of the support $[0, 1]$. The increase of m at the left-hand boundary of the support is not picked up by any of the tests, the increase at the right-hand boundary is only detected by row-wise SiZer $\mathcal{T}_{\text{SiZer}}$, which is indicated by the small blue area at the right-hand boundary of the SiZer map. This again illustrates that the row-wise methods tend to be more powerful than the global tests.

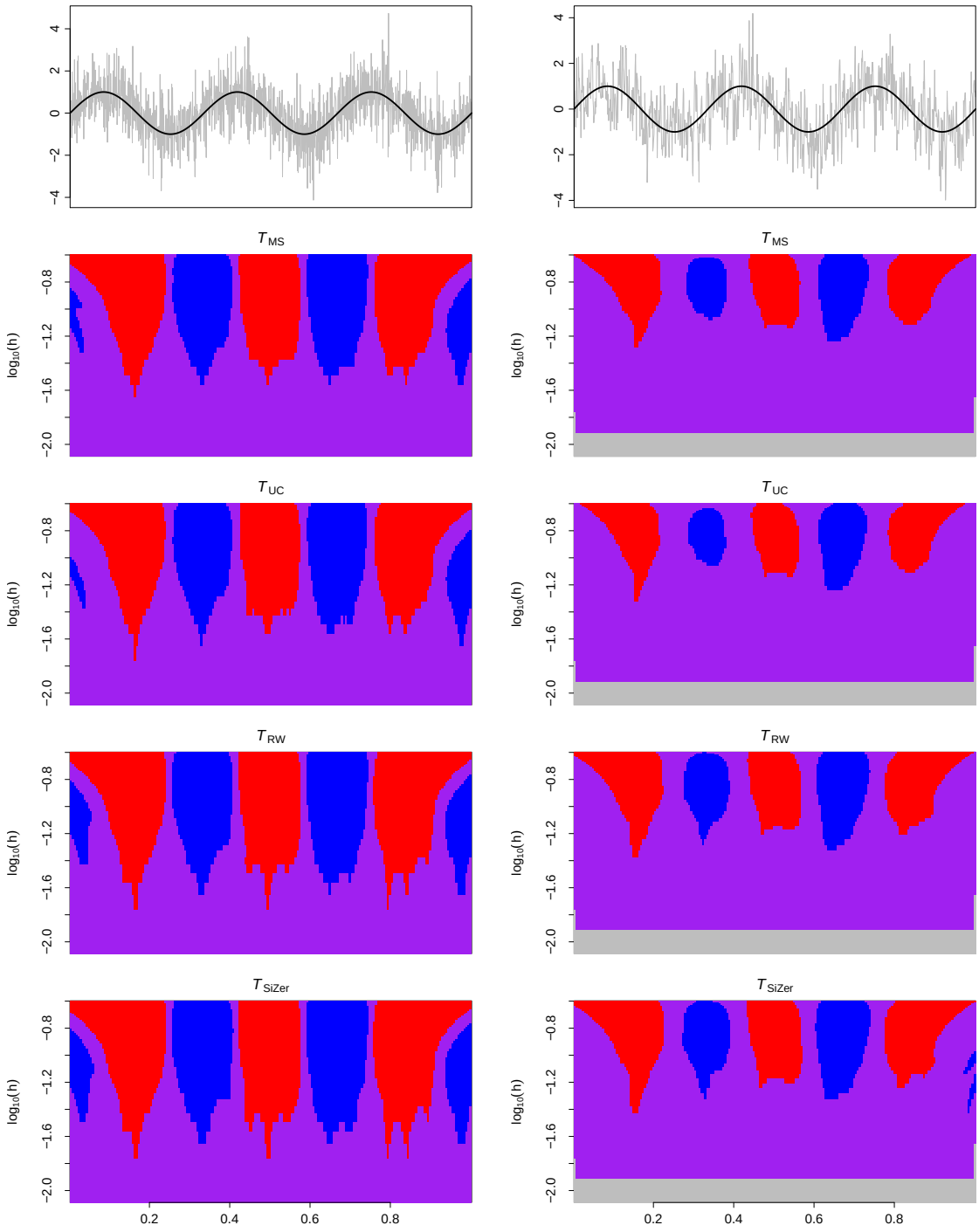
1.C.3 Robustness checks for Section 1.5.2

In what follows, we carry out some robustness checks to assess how sensitive the estimators $\widehat{a}$ and $\widehat{\sigma}^2$ are to the choice of the tuning parameters q and $(\underline{r}, \overline{r})$. To do so, we repeat the simulation exercises of Section 1.5.2 for different values of q and $\overline{r}$. The parameter $\underline{r}$, in



(a) $a = -0.5$ (b) $a = 0.5$

Figure 1.C.1: SiZer maps for the blocks example. The left-hand panels of subfigure (a) show the results for $a_1 = -0.5$, the right-hand panels of subfigure (b) those for $a_1 = 0.5$. The two upper panels depict the trend curve m with the simulated data sample in the background. The other panels show the SiZer maps produced by the four tests $\mathcal{T}_{\mathrm{MS}}$, $\mathcal{T}_{\mathrm{UC}}$, $\mathcal{T}_{\mathrm{RW}}$ and $\mathcal{T}_{\mathrm{SiZer}}$.



(a) $a = -0.5$ (b) $a = 0.5$

Figure 1.C.2: SiZer maps for the sine example. The left-hand panels of subfigure (a) show the results for $a_1 = -0.5$, the right-hand panels of subfigure (b) those for $a_1 = 0.5$. The two upper panels depict the sine curve with the simulated data sample in the background. The other panels show the SiZer maps produced by the four tests $\mathcal{T}_{\mathrm{MS}}$, $\mathcal{T}_{\mathrm{UC}}$, $\mathcal{T}_{\mathrm{RW}}$ and $\mathcal{T}_{\mathrm{SiZer}}$.

contrast, is not varied but set to $\underline{r} = 1$ throughout. The reason is as follows: Our estimators can be expected to perform well as long as the differencing orders r with $\underline{r} \leq r \leq \overline{r}$ are sufficiently small. Choosing $\underline{r}$ as small as possible is thus completely unproblematic (and indeed most natural). The interesting issue is rather how strongly our estimators depend on the choice of $\overline{r}$. In addition, we consider different choices of the tuning parameters (m_1, m_2) on which the estimators of Hall and Van Keilegom (2003) depend. As in Section 1.5.2, we choose m_1 and m_2 such that q lies between these values. We thus keep the parameters q and (m_1, m_2) roughly comparable.

To start with, we consider the simulation scenarios with a moderate trend ($s_\beta = 1$). The MSE values of the estimators $\widehat{a}$, $\widehat{a}_{\text{HvK}}$, $\widehat{a}_{\text{oracle}}$ and $\widehat{\sigma}^2$, $\widehat{\sigma}^2_{\text{HvK}}$, $\widehat{\sigma}^2_{\text{oracle}}$ for these scenarios are presented in Figure 1.4 of Section 1.5.2. These MSEs are re-calculated in Figures 1.C.3 and 1.C.4 for a range of different choices of q, $\overline{r}$ and (m_1, m_2). As one can see, the MSEs in the different plots of Figures 1.C.3 and 1.C.4 are very similar. Hence, the MSE results reported in Section 1.5.2 for the scenarios with a moderate trend appear to be fairly robust to different choices of the tuning parameters. In particular, our estimators $\widehat{a}$ and $\widehat{\sigma}^2$ seem to be quite insensitive to the choice of tuning parameters, at least as far as their MSEs are concerned.

We next turn to the simulation designs with a pronounced trend ($s_\beta = 10$). The MSE values of the estimators in these scenarios are reported in Figure 1.5 of Section 1.5.2. Analogously as before, we re-calculate these MSEs for different tuning parameters in Figures 1.C.5–1.C.7. Figure 1.C.6 is a zoomed-in version of Figure 1.C.5 which is added for better visibility. As can be seen, our estimators appear to be barely influenced by the choice of q. However, the MSE values become somewhat larger when $\overline{r}$ is chosen bigger. This is of course not very surprising: The main reason why the estimator $\widehat{a}$ works well in the presence of a strong trend is that it is only based on differences of small orders. If we increase $\overline{r}$, we use larger differences to compute $\widehat{a}$, which results in not eliminating the trend m appropriately any more. This becomes visible in somewhat larger MSE values. Nevertheless, overall, our estimators appear not to be strongly influenced by the choice of tuning parameters (in terms of MSE) as long as these are chosen within reasonable bounds.

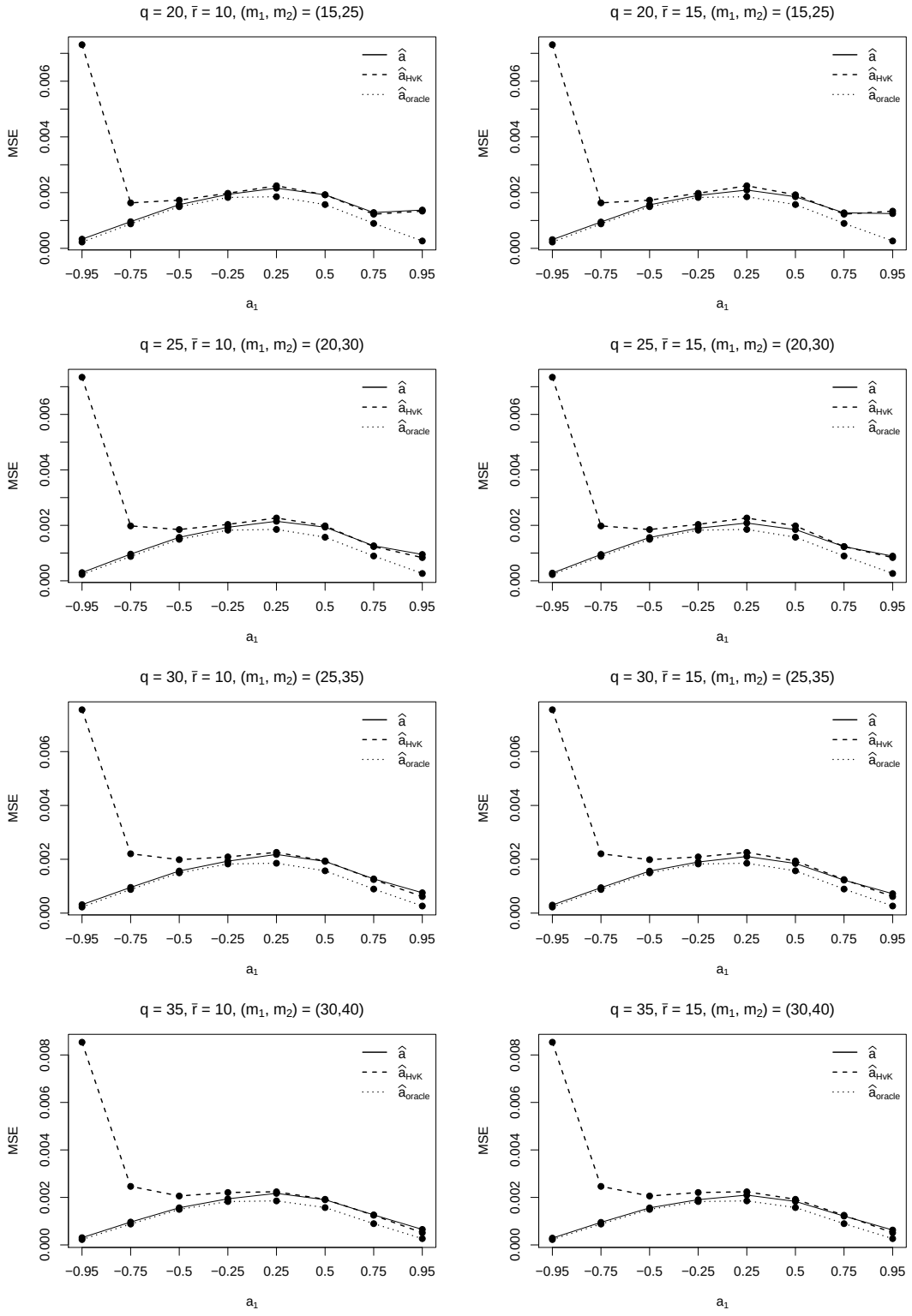


Figure 1.C.3: MSE values for the estimators $\widehat{a}$, $\widehat{a}_{\text{HvK}}$ and $\widehat{a}_{\text{oracle}}$ in the scenario with a moderate trend ($s_\beta = 1$).

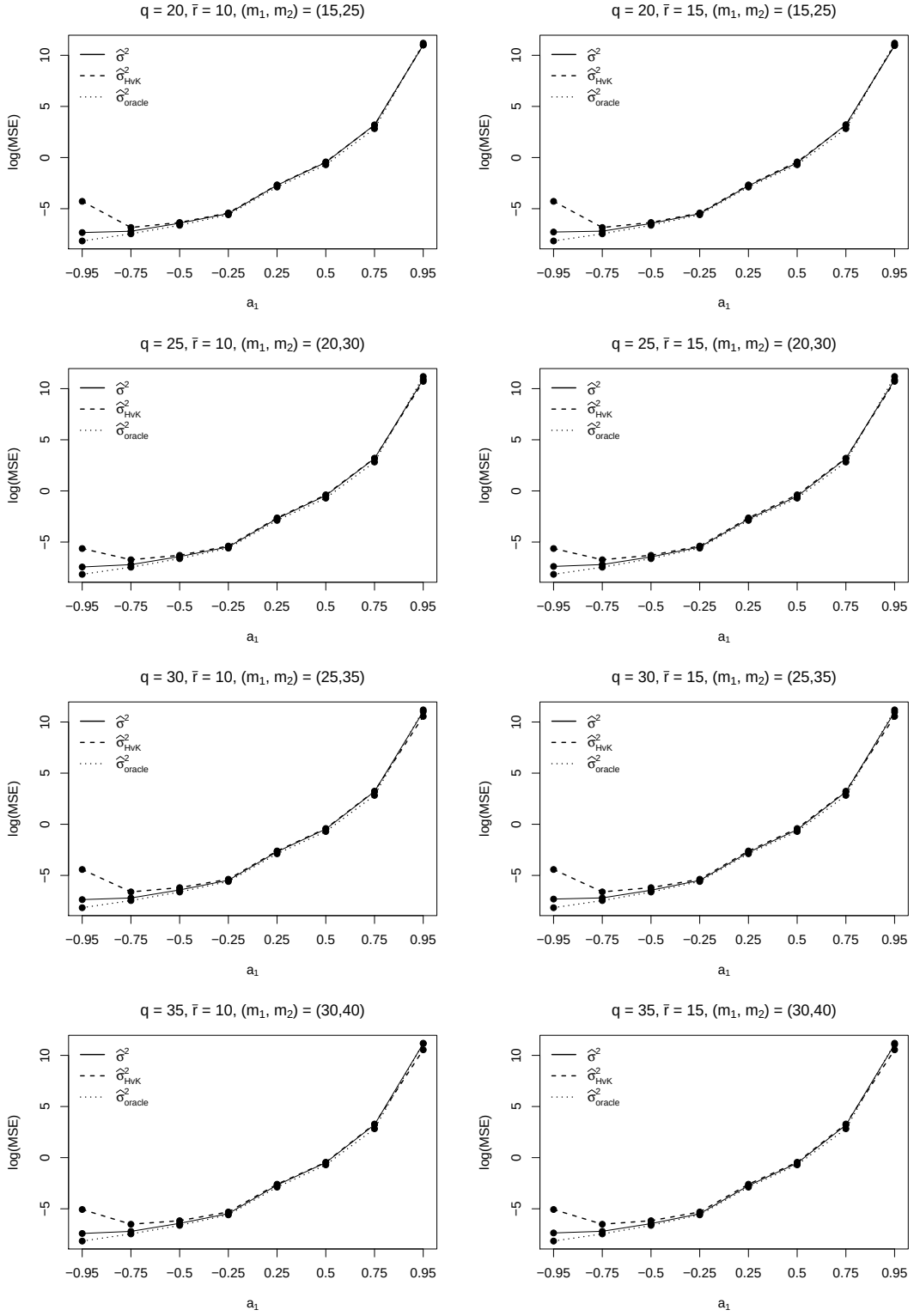


Figure 1.C.4: Logarithmic MSE values for the estimators $\widehat{\sigma}^2$, $\widehat{\sigma}^2_{\mathrm{HvK}}$ and $\widehat{\sigma}^2_{\mathrm{oracle}}$ in the scenario with a moderate trend ($s_\beta = 1$).

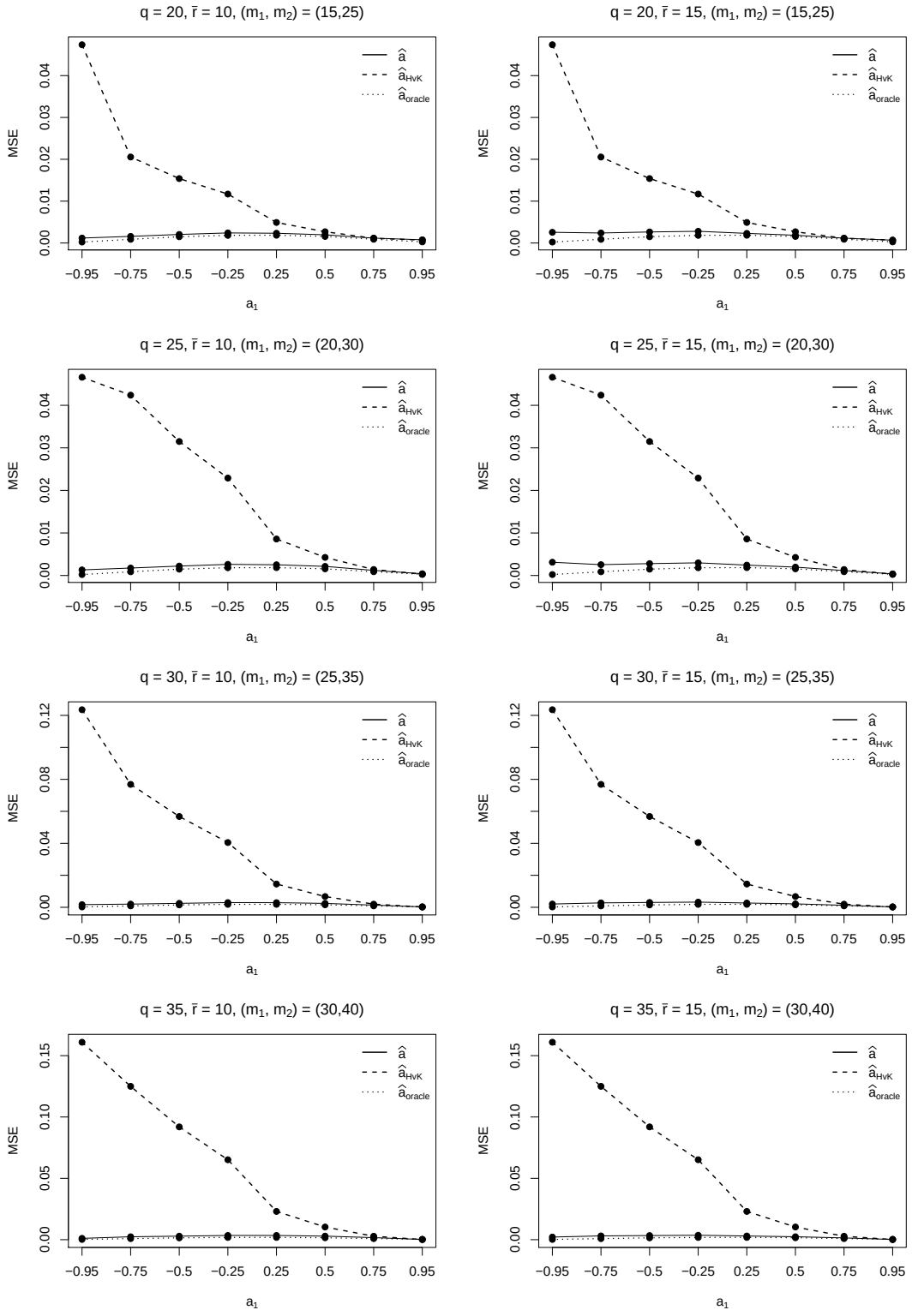


Figure 1.C.5: MSE values for the estimators $\widehat{a}$, $\widehat{a}_{\text{HvK}}$ and $\widehat{a}_{\text{oracle}}$ in the scenario with a pronounced trend ($s_\beta = 10$).

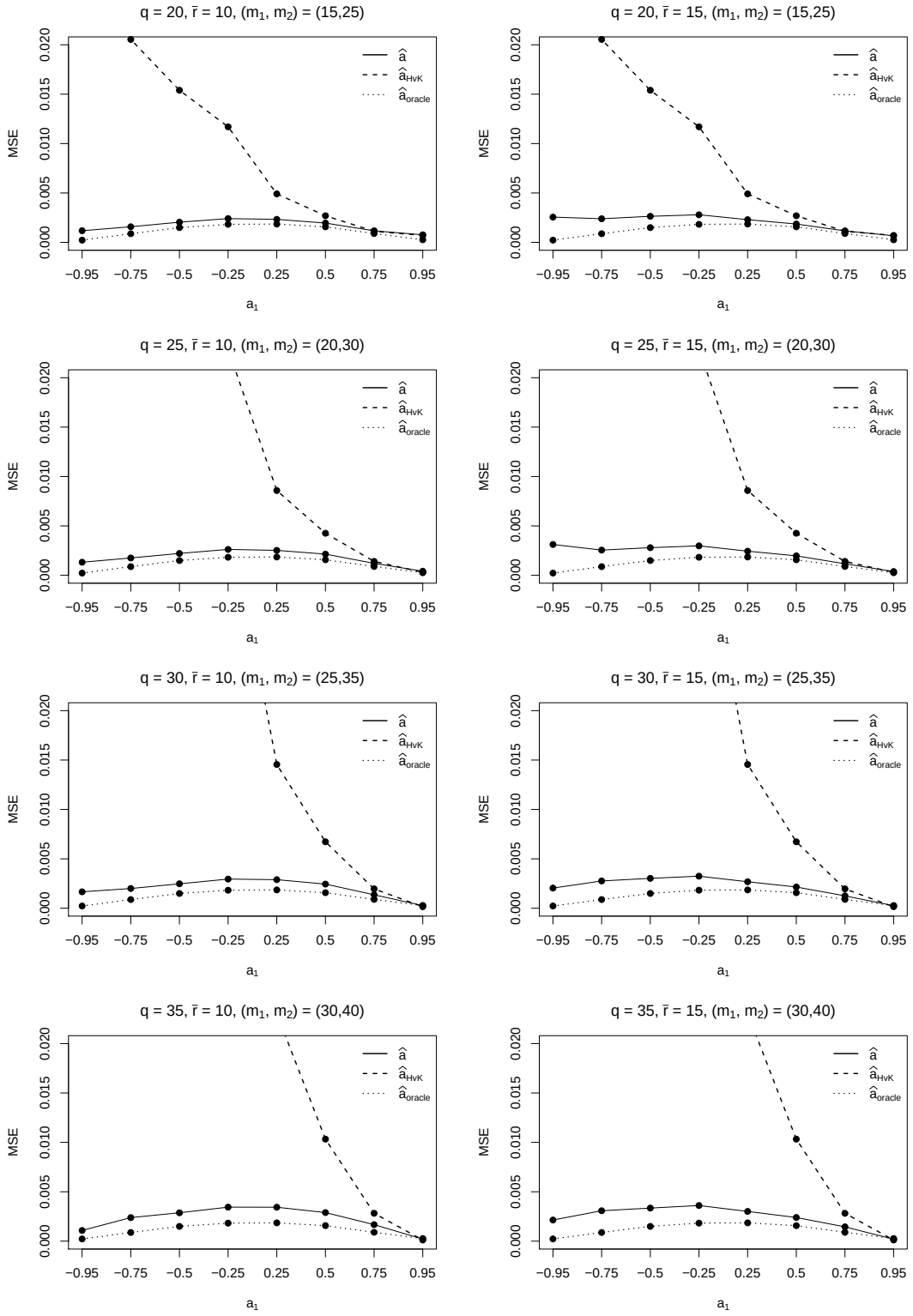


Figure 1.C.6: MSE values for the estimators $\widehat{a}$, $\widehat{a}_{\mathrm{HvK}}$ and $\widehat{a}_{\mathrm{oracle}}$ in the scenario with a pronounced trend ($s_\beta = 10$). The plots are zoomed-in versions of the respective plots in Figure 1.C.5.

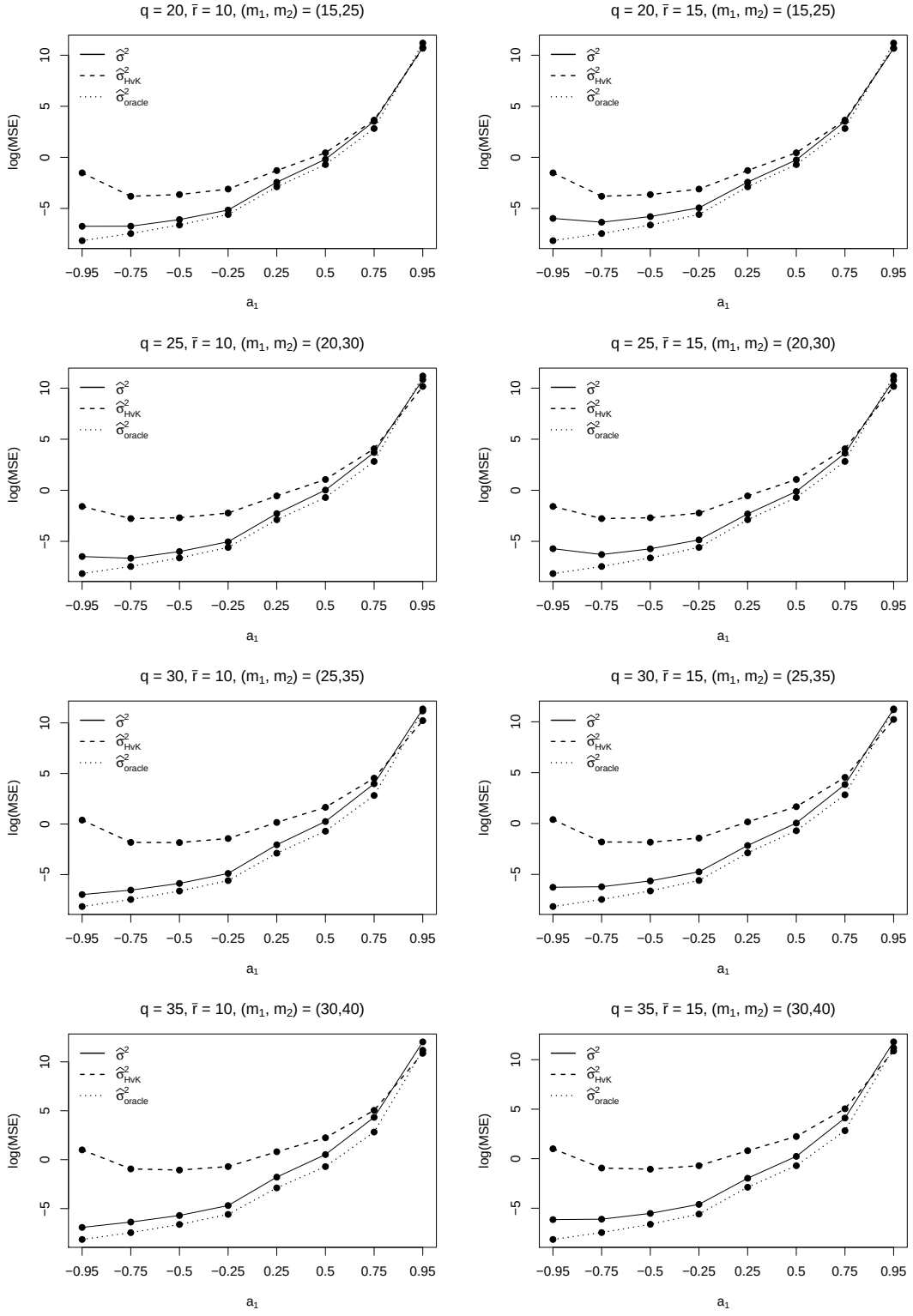


Figure 1.C.7: Logarithmic MSE values for the estimators $\widehat{\sigma}^2$, $\widehat{\sigma}^2_{\text{HvK}}$ and $\widehat{\sigma}^2_{\text{oracle}}$ in the scenario with a pronounced trend ($s_\beta = 10$).

Chapter 2

Nonparametric Comparison of Epidemic Time Trends: the Case of COVID-19

Joint with Michael Vogt

2.1 Introduction

There are many questions surrounding the current COVID-19 pandemic that are not well understood yet. A question which is particularly important for governments and policy makers is the following: How do the outbreak patterns of COVID-19 compare across countries? Are the time trends of daily new infections more or less the same across countries, or is the virus spreading differently in different regions of the world? Identifying differences between countries may help, for instance, to better understand which government policies have been more effective in containing the virus than others. The main aim of this paper is to develop new inference methods that allow to detect differences between time trends of COVID-19 infections in a statistically rigorous way.

Let X_{it} be the number of new infections on day t in country i and suppose we observe a sample of data $\mathcal{X}_i = \{X_{it} : 1 \leq t \leq T\}$ for n different countries i. In order to make the data comparable across countries, we take the starting date $t = 1$ to be the first Monday after reaching 100 confirmed cases in each country. Considering the dates after reaching a certain level of confirmed cases is a common practice of "normalizing" the data (see e.g. Cohen and Kupferschmidt, 2020). Starting on a Monday additionally aligns the data across countries by the day of the week. This allows us to take care of possible weekly cycles in the data which are produced by delays in reporting new cases over the weekend. A simple way to model the count data X_{it} is to use a Poisson distribution. Specifically, we may assume that the random variables X_{it} are Poisson distributed with time-varying intensity parameter $\lambda_i(t/T)$, that is, $X_{it} \sim P_{\lambda_i(t/T)}$. Since $\lambda_i(t/T) = \mathbb{E}[X_{it}] = \mathrm{Var}(X_{it})$, we can model the observations X_{it} by the nonparametric regression equation

$$X_{it} = \lambda_i\Big(\frac{t}{T}\Big) + u_{it} \tag{2.1.1}$$

for $1 \leq t \leq T$, where $u_{it} = X_{it} - \mathbb{E}[X_{it}]$ with $\mathbb{E}[u_{it}] = 0$ and $\mathrm{Var}(u_{it}) = \lambda_i(t/T)$. As usual in nonparametric regression (see e.g. Robinson, 1989), we let the regression function

λ_i in model (2.1.1) depend on rescaled time t/T rather than on real time t. Hence, $\lambda_i : [0,1] \to \mathbb{R}$ can be regarded as a function on the unit interval, which allows us to estimate it by standard techniques from nonparametric regression. Since λ_i is a function of rescaled time t/T, the variables X_{it} in model (2.1.1) depend on the time series length T in general, that is, $X_{it} = X_{it,T}$. To keep the notation simple, we however suppress this dependence throughout the paper. In Section 3.2, we introduce the model setting in detail which underlies our analysis. As we will see there, it is a generalized version of the Poisson model (2.1.1).

In model (2.1.1), the time trend of new COVID-19 infections in country i is described by the intensity function λ_i of the underlying Poisson distribution. Hence, the question whether the time trends are comparable across countries amounts to the question whether the intensity functions λ_i have the same shape across countries i. In this paper, we construct a multiscale test which allows to *identify* and *locate* the differences between the functions λ_i. More specifically, let $\mathcal{F} = \{\mathcal{I}_k \subseteq [0,1] : 1 \leq k \leq K\}$ be a family of (rescaled) time intervals $\mathcal{I}_k$ and let $H_0^{(ijk)}$ be the hypothesis that the functions λ_i and λ_j are the same on the interval $\mathcal{I}_k$, that is,

$$H_0^{(ijk)} : \lambda_i(w) = \lambda_j(w) \text{ for all } w \in \mathcal{I}_k.$$

We design a method to test the hypothesis $H_0^{(ijk)}$ *simultaneously* for all pairs of countries i and j under consideration and for all intervals $\mathcal{I}_k$ in the family $\mathcal{F}$. The main theoretical result of the paper shows that the method controls the familywise error rate, that is, the probability of wrongly rejecting at least one null hypothesis $H_0^{(ijk)}$. As we will see, this allows us to make simultaneous confidence statements of the following form for a given significance level $\alpha \in (0,1)$:

> *With probability at least $1-\alpha$, the functions λ_i and λ_j differ on the interval $\mathcal{I}_k$ for every (i,j,k) for which the test rejects $H_0^{(ijk)}$.*

Hence, the method allows us to make simultaneous confidence statements (a) about which time trend functions differ from each other and (b) about where, that is, in which time intervals $\mathcal{I}_k$ they differ.

Even though our multiscale test is motivated by the current COVID-19 crisis, its applicability is by no means restricted to this specific event. It is a general method to compare nonparametric trends in epidemiological (count) data. It thus contributes to the literature on statistical tests for equality of nonparametric regression and trend curves. Examples of such tests can be found in Härdle and Marron (1990), Hall and Hart (1990), King et al. (1991), Delgado (1993), Kulasekera (1995), Young and Bowman (1995), Munk and Dette (1998), Lavergne (2001), Neumeyer and Dette (2003) and Pardo-Fernández et al. (2007). More recent approaches were developed in Degras et al. (2012), Zhang et al. (2012), Hidalgo and Lee (2014) and Chen and Wu (2018). Compared to existing methods, our test has the following crucial advantage: it is much more informative. Most existing procedures allow to test *whether* the regression or trend curves under consideration are

all the same or not. However, they do not allow to infer *which* curves are different and *where* (that is, in which parts of the support) they differ. Our multiscale approach, in contrast, conveys this information. Indeed, it even allows to make rigorous confidence statements about which curves λ_i are different and where they differ. To the best of our knowledge, there is no other method available in the literature which allows to make such simultaneous confidence statements. As far as we know, the only other multiscale test for comparing trend curves has been developed in Park et al. (2009b). However, their analysis is mainly methodological and not backed up by a general theory. In particular, theory is only available for the special case $n = 2$. Moreover, the theoretical results are only valid under very severe restrictions on the family of time intervals $\mathcal{F}$.

The paper is structured as follows. As already mentioned above, Section 3.2 details the model setting which underlies our analysis. The multiscale test is developed step by step in Section 3.3. To keep the presentation as clear as possible, the technical details are deferred to the Appendix. Section 2.4 contains the empirical part of the paper. There, we run some simulation experiments to demonstrate that the multiscale test has the formal properties predicted by the theory. Moreover, we use the test to compare the outbreak patterns of the COVID-19 epidemic in a number of European countries.

2.2 Model setting

As already discussed in the Introduction, the assumption that $X_{it} \sim P_{\lambda_i(t/T)}$ leads to a nonparametric regression model of the form

$$X_{it} = \lambda_i\Big(\frac{t}{T}\Big) + u_{it} \quad \text{with} \quad u_{it} = \sqrt{\lambda_i\Big(\frac{t}{T}\Big)}\eta_{it}, \tag{2.2.1}$$

where η_{it} has zero mean and unit variance. In this model, both the mean and the variance are described by the same function λ_i. In empirical applications, however, the variance often tends to be much larger than the mean. To deal with this issue, which has been known for a long time in the literature (Cox, 1983) and which is commonly called overdispersion, so-called quasi-Poisson models (Efron, 1986; McCullagh and Nelder, 1989) are frequently used. In our context, a quasi-Poisson model of X_{it} has the form

$$X_{it} = \lambda_i\Big(\frac{t}{T}\Big) + \varepsilon_{it} \quad \text{with} \quad \varepsilon_{it} = \sigma\sqrt{\lambda_i\Big(\frac{t}{T}\Big)}\eta_{it}, \tag{2.2.2}$$

where σ is a scaling factor that allows the variance to be a multiple of the mean function λ_i. In what follows, we assume that the observed data X_{it} are produced by model (3.1.1), where the noise residuals η_{it} have zero mean and unit variance but we do not impose any further distributional assumptions on them.

Poisson and quasi-Poisson models are often used in the literature on epidemic modelling. De Salazar et al. (2020), for example, assume that the observed COVID-19 case count in country i follows a Poisson distribution with parameter λ_i being a linear function

of some covariate Z_i, that is, $\lambda_i = \beta Z_i$. Pellis et al. (2020) consider a quasi-Poisson model for the number of new COVID-19 cases. They in particular examine (a) a version of the model where the mean function is parametrically restricted to be exponentially growing with a constant growth rate and (b) a version where the mean function is modelled nonparametrically by splines. Tobías et al. (2020) analyze data on the accumulated number of cases using quasi-Poisson regression, where the mean function is modelled parametrically as a piecewise linear curve with known change points.

In order to derive our theoretical results, we impose the following regularity conditions on model (3.1.1):

(C1) The functions λ_i are uniformly Lipschitz continuous, that is, $|\lambda_i(u)-\lambda_i(v)| \leq L|u-v|$ for all $u, v \in [0,1]$, where the constant L does not depend on i. Moreover, they are uniformly bounded away from zero and infinity, that is, there exist constants $\lambda_{\min}$ and $\lambda_{\max}$ with $0 < \lambda_{\min} \leq \min_{w\in[0,1]} \lambda_i(w) \leq \max_{w\in[0,1]} \lambda_i(w) \leq \lambda_{\max} < \infty$ for all i.

(C2) The random variables η_{it} are independent both across i and t. Moreover, for any i and t, it holds that $\mathbb{E}[\eta_{it}] = 0$, $\mathbb{E}[\eta_{it}^2] = 1$ and $\mathbb{E}[|\eta_{it}|^\theta] \leq C_\theta < \infty$ for some $\theta > 4$.

We briefly comment on the above conditions.

- (C1) imposes some standard-type regularity conditions on the functions λ_i. In particular, the functions are assumed to be smooth, bounded from above and bounded away from zero. The latter restriction is required because the noise variance in model (3.1.1) equals 0 if λ_i is equal to 0. Since we normalize our test statistics by an estimate of the noise variance as detailed in Section 3.3, we need this variance and thus the functions λ_i to be bounded away from zero.

- (C2) assumes the noise terms η_{it} to fulfill some mild moment conditions and to be independent both across countries i and time t. We require the independence assumptions of (C2) in order to apply the Gaussian approximation results for hyperrectangles from Chernozhukov et al. (2017) in our proofs. We in particular need independence across t, but it would in principle be possible to allow for certain forms of dependence across i at the cost of a more complicated test procedure and more involved technical arguments.

- In the current COVID-19 crisis, independence across countries i seems to be a fairly reasonable assumption due to severe travel restrictions, the closure of borders, etc. Note that this assumption can in principle be tested, for example, with the help of the tests in Blum et al. (1961), Sinha and Wieand (1977) and Bakirov et al. (2006).

- Independence across time t is more debatable than independence across countries i, but it is by no means unreasonable in our model framework: The time series process $\mathcal{X}_i = \{X_{it} : 1 \leq t \leq T\}$ produced by model (3.1.1) is nonstationary for each i. Specifically, both the mean $\mathbb{E}[X_{it}] = \lambda_i(t/T)$ and the variance $\mathrm{Var}(X_{it}) = \sigma^2\lambda_i(t/T)$ are

time-varying. A well-known fact in the time series literature is that nonstationarities such as a time-varying mean may produce spurious sample autocorrelations (see e.g. Fryzlewicz et al., 2008; Mikosch and Stărică, 2004). Hence, the observed persistence of a time series (captured by the sample autocorrelation function) may be due to nonstationarities rather than real autocorrelations. This insight has led researchers to prefer simple nonstationary models over intricate stationary time series models in some application areas such as finance (see e.g. Fryzlewicz et al., 2006; Hafner and Linton, 2010; Mikosch and Stărică, 2000, 2004). In a similar vein, our model accounts for the persistence in the observed time series $\mathcal{X}_i$ via nonstationarities rather than autocorrelations in the error terms.

2.3 The multiscale test

Let $\mathcal{S} \subseteq \{(i,j) : 1 \leq i < j \leq n\}$ be the set of all pairs of countries (i,j) whose trend functions λ_i and λ_j we want to compare. Moreover, as already introduced above, let $\mathcal{F} = \{\mathcal{I}_k : 1 \leq k \leq K\}$ be the family of (rescaled) time intervals under consideration. Finally, write $\mathcal{M} := \mathcal{S} \times \{1, \ldots, K\}$ and let $p := |\mathcal{M}|$ be the cardinality of $\mathcal{M}$. In this section, we devise a method to test the null hypothesis $H_0^{(ijk)}$ simultaneously for all pairs of countries $(i,j) \in \mathcal{S}$ and all time intervals $\mathcal{I}_k \in \mathcal{F}$, that is, for all $(i,j,k) \in \mathcal{M}$. The value $p = |\mathcal{M}|$ is the dimensionality of the simultaneous test problem we are dealing with. It amounts to the number of tests that we carry out simultaneously. As shown by our theoretical results in the Appendix, p may grow as a polynomial T^γ of the time series length T, where the exponent γ depends on the number of error moments θ defined in (C2) and on the minimal length of the rescaled time intervals in the family $\mathcal{F}$. Precise conditions on the exponent γ are given in the statement of Theorem 2.A.1. These conditions show that γ can be very large provided that the error terms have sufficiently many moments θ. Consequently, p may be much larger than the time series length T, which means that the simultaneous test problem under consideration can be very high-dimensional.

2.3.1 Construction of the test statistics

A statistic to test the hypothesis $H_0^{(ijk)}$ for a given triple (i,j,k) can be constructed as follows. To start with, we consider the expression

$$\hat{s}_{ijk,T} = \frac{1}{Th_k} \sum_{t=1}^{T} 1\Big(\frac{t}{T} \in \mathcal{I}_k\Big)(X_{it} - X_{jt}),$$

where h_k is the length of the time interval $\mathcal{I}_k$, $1(\cdot)$ denotes the indicator function and $1(t/T \in \mathcal{I}_k)$ can be regarded as a rectangular kernel weight. Inserting the model equation

(3.1.1) into the definition of $\hat{s}_{ijk,T}$ yields that $\hat{s}_{ijk,T} = \Delta_{ijk,T} + R_{ijk,T}$, where

$$\Delta_{ijk,T} = \frac{1}{Th_k} \sum_{t=1}^{T} 1\Big(\frac{t}{T} \in \mathcal{I}_k\Big)\Big\{\lambda_i\Big(\frac{t}{T}\Big) - \lambda_j\Big(\frac{t}{T}\Big)\Big\}$$

is the average distance between the functions λ_i and λ_j on the interval $\mathcal{I}_k$ and

$$R_{ijk,T} = \frac{1}{Th_k} \sum_{t=1}^{T} 1\Big(\frac{t}{T} \in \mathcal{I}_k\Big)\sigma\Big\{\sqrt{\lambda_i\Big(\frac{t}{T}\Big)}\eta_{it} - \sqrt{\lambda_j\Big(\frac{t}{T}\Big)}\eta_{jt}\Big\}$$

is a remainder term that is asymptotically negligible in the following sense: A simple application of the law of large numbers for a fixed i gives that $(Th_k)^{-1}\sum_{t=1}^{T} 1(t/T \in \mathcal{I}_k)\sqrt{\lambda_i(t/T)}\eta_{it} = o_p(1)$, which in turn implies that $R_{ijk,T} = o_p(1)$. Hence, for any fixed triple (i, j, k), we obtain that

$$\hat{s}_{ijk,T} = \Delta_{ijk,T} + o_p(1),$$

which means that the statistic $\hat{s}_{ijk,T}$ estimates the average distance $\Delta_{ijk,T}$ between the functions λ_i and λ_j on the interval $\mathcal{I}_k$.

We next have a closer look at the variance of the statistic $\hat{s}_{ijk,T}$. Under (C2), the variance of $\hat{s}_{ijk,T}$ is given by

$$\nu^2_{ijk,T} := \mathrm{Var}(\hat{s}_{ijk,T}) = \frac{\sigma^2}{(Th_k)^2} \sum_{t=1}^{T} 1\Big(\frac{t}{T} \in \mathcal{I}_k\Big)\Big\{\lambda_i\Big(\frac{t}{T}\Big) + \lambda_j\Big(\frac{t}{T}\Big)\Big\}$$

and can be estimated by

$$\hat{\nu}^2_{ijk,T} = \frac{\hat{\sigma}^2}{(Th_k)^2} \sum_{t=1}^{T} 1\Big(\frac{t}{T} \in \mathcal{I}_k\Big)\{X_{it} + X_{jt}\}.$$

Here, $\hat{\sigma}^2$ is an estimator of σ^2 which is defined as $\hat{\sigma}^2 = |\mathcal{C}|^{-1}\sum_{i\in\mathcal{C}} \hat{\sigma}_i^2$, where $\mathcal{C} = \{\ell : \ell = i$ or $\ell = j$ for some $(i, j) \in \mathcal{S}\}$ denotes the set of countries that are taken into account by our test and

$$\hat{\sigma}_i^2 = \frac{\sum_{t=2}^{T}(X_{it} - X_{it-1})^2}{2\sum_{t=1}^{T} X_{it}}$$

for each country i. The idea behind the estimator $\hat{\sigma}_i^2$ is as follows: We can write

$$X_{it} - X_{it-1} = \sigma\sqrt{\lambda_i\Big(\frac{t}{T}\Big)}(\eta_{it} - \eta_{it-1}) + r_{it} \tag{2.3.1}$$

with

$$r_{it} = \lambda_i\Big(\frac{t}{T}\Big) - \lambda_i\Big(\frac{t-1}{T}\Big) + \sigma\Big\{\sqrt{\lambda_i\Big(\frac{t}{T}\Big)} - \sqrt{\lambda_i\Big(\frac{t-1}{T}\Big)}\Big\}\eta_{it-1}.$$

By the triangle inequality and since λ_i is Lipschitz continuous, we have that

$$\begin{aligned}|r_{it}| &\leq \Big|\lambda_i\Big(\frac{t}{T}\Big) - \lambda_i\Big(\frac{t-1}{T}\Big)\Big| + \sigma\Big|\sqrt{\lambda_i\Big(\frac{t}{T}\Big)} - \sqrt{\lambda_i\Big(\frac{t-1}{T}\Big)}\Big||\eta_{it-1}| \\ &\leq \frac{L}{T} + \frac{\sigma L}{2\sqrt{\lambda_{\min}}T}|\eta_{it-1}| \leq \frac{C(1+|\eta_{it-1}|)}{T}, \end{aligned} \tag{2.3.2}$$

where $C = \max\{L, \sigma L/(2\sqrt{\lambda_{\min}})\}$ with L and $\lambda_{\min}$ defined in (C1). From (2.3.1) and (2.3.2), we can infer that

$$\frac{1}{T}\sum_{t=2}^{T}(X_{it} - X_{it-1})^2 = 2\sigma^2\Big\{\frac{1}{T}\sum_{t=2}^{T}\lambda_i\Big(\frac{t}{T}\Big)\Big\} + o_p(1).$$

Moreover, since $T^{-1}\sum_{t=1}^{T} X_{it} = T^{-1}\sum_{t=1}^{T}\lambda_i(t/T) + o_p(1)$, we get that $\hat{\sigma}_i^2 = \sigma^2 + o_p(1)$ for any fixed i. In Lemma 2.B.8 of the Appendix, we further show that $\hat{\sigma}^2 = \sigma^2 + o_p(1)$ under our regularity conditions. Hence, $\hat{\sigma}^2$ is a consistent estimator of σ^2.

We now replace the statistic $\hat{s}_{ijk,T}$ by a normalized version whose variance is approximately equal to 1. To achieve this, we simply divide $\hat{s}_{ijk,T}$ by its estimated standard deviation $\hat{\nu}_{ijk,T}$. This results in the expression

$$\hat{\psi}_{ijk,T} := \frac{\hat{s}_{ijk,T}}{\hat{\nu}_{ijk,T}} = \frac{\sum_{t=1}^{T} 1(\frac{t}{T} \in \mathcal{I}_k)(X_{it} - X_{jt})}{\hat{\sigma}\{\sum_{t=1}^{T} 1(\frac{t}{T} \in \mathcal{I}_k)(X_{it} + X_{jt})\}^{1/2}}, \tag{2.3.3}$$

which serves as our test statistic of the hypothesis $H_0^{(ijk)}$. In addition to $\hat{\psi}_{ijk,T}$, we introduce the auxiliary statistic

$$\hat{\psi}^0_{ijk,T} = \frac{\sum_{t=1}^{T} 1(\frac{t}{T} \in \mathcal{I}_k)\,\sigma\overline{\lambda}_{ij}^{1/2}(\frac{t}{T})(\eta_{it} - \eta_{jt})}{\hat{\sigma}\{\sum_{t=1}^{T} 1(\frac{t}{T} \in \mathcal{I}_k)(X_{it} + X_{jt})\}^{1/2}} \tag{2.3.4}$$

with $\overline{\lambda}_{ij}(u) = \{\lambda_i(u) + \lambda_j(u)\}/2$, which by construction is identical to $\hat{\psi}_{ijk,T}$ under $H_0^{(ijk)}$. This auxiliary statistic is needed to define the critical values of our multiscale test in what follows.

2.3.2 Construction of the test

Our multiscale test is carried out as follows: For a given significance level $\alpha \in (0,1)$ and each $(i,j,k) \in \mathcal{M}$, we reject $H_0^{(ijk)}$ if

$$|\hat{\psi}_{ijk,T}| > c_{ijk,T}(\alpha),$$

where $c_{ijk,T}(\alpha)$ is the critical value for the (i,j,k)-th test problem. The critical values $c_{ijk,T}(\alpha)$ are chosen such that the familywise error rate (FWER) is controlled at level α, which is defined as the probability of wrongly rejecting $H_0^{(ijk)}$ for at least one (i,j,k).

More formally speaking, for a given significance level $\alpha \in (0,1)$, the FWER is

$$\begin{aligned}\mathrm{FWER}(\alpha) &= \mathbb{P}\Big(\exists (i,j,k) \in \mathcal{M}_0 : |\hat{\psi}_{ijk,T}| > c_{ijk,T}(\alpha)\Big) \\ &= 1 - \mathbb{P}\Big(\forall (i,j,k) \in \mathcal{M}_0 : |\hat{\psi}_{ijk,T}| \leq c_{ijk,T}(\alpha)\Big),\end{aligned}$$

where $\mathcal{M}_0 \subseteq \mathcal{M}$ is the set of triples (i,j,k) for which $H_0^{(ijk)}$ holds true.

There are different ways to construct critical values $c_{ijk,T}(\alpha)$ that ensure control of the FWER at level α. In the traditional approach, the same critical value $c_T(\alpha) = c_{ijk,T}(\alpha)$ is used for all (i,j,k). In this case, controlling the FWER at the level α requires to determine the critical value $c_T(\alpha)$ such that

$$\begin{aligned}\mathrm{FWER}(\alpha) &= 1 - \mathbb{P}\Big(\forall (i,j,k) \in \mathcal{M}_0 : |\hat{\psi}_{ijk,T}| \leq c_T(\alpha)\Big) \\ &= 1 - \mathbb{P}\Big(\max_{(i,j,k)\in\mathcal{M}_0} |\hat{\psi}_{ijk,T}| \leq c_T(\alpha)\Big) \leq \alpha. \qquad (2.3.5)\end{aligned}$$

This can be achieved by choosing $c_T(\alpha)$ as the $(1-\alpha)$-quantile of the statistic

$$\tilde{\Psi}_T = \max_{(i,j,k)\in\mathcal{M}} |\hat{\psi}^0_{ijk,T}|,$$

where the auxiliary statistic $\hat{\psi}^0_{ijk,T}$ was introduced in (2.3.4) and is equal to $\hat{\psi}_{ijk,T}$ under the null $H_0^{(ijk)}$ by construction.[1]

A more modern approach assigns different critical values $c_{ijk,T}(\alpha)$ to the test problems (i,j,k). In particular, the critical value for the hypothesis $H_0^{(ijk)}$ is allowed to depend on the length h_k of the time interval $\mathcal{I}_k$, that is, on the scale of the test problem. A general approach to construct scale-dependent critical values was pioneered by Dümbgen and Spokoiny (2001) and has been used in many other studies since then; see e.g. Rohde (2008), Dümbgen and Walther (2008), Rufibach and Walther (2010), Schmidt-Hieber et al. (2013), Eckle et al. (2017) and Dunker et al. (2019). In our context, the approach of Dümbgen and Spokoiny (2001) leads to the critical values

$$c_{ijk,T}(\alpha) = c_T(\alpha, h_k) := b_k + q_T(\alpha)/a_k,$$

where $a_k = \{\log(e/h_k)\}^{1/2}/\log\log(e^e/h_k)$ and $b_k = \sqrt{2\log(1/h_k)}$ are scale-dependent constants and the quantity $q_T(\alpha)$ is determined by the following consideration: Since

$$\begin{aligned}\mathrm{FWER}(\alpha) &= \mathbb{P}\Big(\exists (i,j,k) \in \mathcal{M}_0 : |\hat{\psi}_{ijk,T}| > c_T(\alpha, h_k)\Big) \\ &= 1 - \mathbb{P}\Big(\forall (i,j,k) \in \mathcal{M}_0 : |\hat{\psi}_{ijk,T}| \leq c_T(\alpha, h_k)\Big) \\ &= 1 - \mathbb{P}\Big(\forall (i,j,k) \in \mathcal{M}_0 : a_k\big(|\hat{\psi}_{ijk,T}| - b_k\big) \leq q_T(\alpha)\Big)\end{aligned}$$

[1] Note that both the statistic $\tilde{\Psi}_T$ and the quantile $c_T(\alpha)$ depend on the dimensionality p of the test problem in general. To keep the notation simple, we however suppress this dependence throughout the paper. We use the same convention for all other quantities that are defined in the sequel.

$$= 1 - \mathbb{P}\Big(\max_{(i,j,k)\in\mathcal{M}_0} a_k\big(|\hat{\psi}_{ijk,T}| - b_k\big) \le q_T(\alpha)\Big), \tag{2.3.6}$$

we need to choose the quantity $q_T(\alpha)$ as the $(1-\alpha)$-quantile of the statistic

$$\hat{\Psi}_T = \max_{(i,j,k)\in\mathcal{M}} a_k\big(|\hat{\psi}^0_{ijk,T}| - b_k\big)$$

in order to ensure control of the FWER at level α. Comparing (2.3.6) with (2.3.5), the current approach can be seen to differ from the traditional one in the following respect: the maximum statistic $\tilde{\Psi}_T$ is replaced by the rescaled version $\hat{\Psi}_T$ which re-weights the individual statistics $\hat{\psi}^0_{ijk,T}$ by the scale-dependent constants a_k and b_k. As demonstrated above, this translates into scale-dependent critical values $c_{ijk,T}(\alpha) = c_T(\alpha, h_k)$.

Our theory allows us to work with both the traditional choice $c_{ijk,T}(\alpha) = c_T(\alpha)$ and the more modern, scale-dependent choice $c_{ijk,T}(\alpha) = c_T(\alpha, h_k)$. Since the latter choice produces a test approach with better theoretical properties in general (see Dümbgen and Spokoiny, 2001), we restrict attention to the critical values $c_T(\alpha, h_k)$ in the sequel. There is one complication we need to deal with: As the quantiles $q_T(\alpha)$ are not known in practice, we cannot compute the critical values $c_T(\alpha, h_k)$ exactly in practice but need to approximate them. This can be achieved as follows: Under appropriate regularity conditions, it can be shown that

$$\hat{\psi}^0_{ijk,T} \approx \frac{1}{\sqrt{2Th_k}} \sum_{t=1}^{T} 1\Big(\frac{t}{T} \in \mathcal{I}_k\Big)\{\eta_{it} - \eta_{jt}\}.$$

A Gaussian version of the statistic displayed on the right-hand side above is given by

$$\phi_{ijk,T} = \frac{1}{\sqrt{2Th_k}} \sum_{t=1}^{T} 1\Big(\frac{t}{T} \in \mathcal{I}_k\Big)\big\{Z_{it} - Z_{jt}\big\},$$

where Z_{it} are independent standard normal random variables for $1 \le t \le T$ and $1 \le i \le n$. Hence, the statistic

$$\Phi_T = \max_{(i,j,k)\in\mathcal{M}} a_k\big(|\phi_{ijk,T}| - b_k\big)$$

can be regarded as a Gaussian version of the statistic $\hat{\Psi}_T$. We approximate the unknown quantile $q_T(\alpha)$ by the $(1-\alpha)$-quantile $q_{T,\mathrm{Gauss}}(\alpha)$ of Φ_T, which can be computed (approximately) by Monte Carlo simulations and can thus be treated as known.

To summarize, we propose the following procedure to simultaneously test the hypothesis $H_0^{(ijk)}$ for all $(i,j,k) \in \mathcal{M}$ at the significance level $\alpha \in (0,1)$:

$$\text{For each } (i,j,k) \in \mathcal{M}, \text{ reject } H_0^{(ijk)} \text{ if } |\hat{\psi}_{ijk,T}| > c_{T,\mathrm{Gauss}}(\alpha, h_k), \tag{2.3.7}$$

where $c_{T,\mathrm{Gauss}}(\alpha, h_k) = b_k + q_{T,\mathrm{Gauss}}(\alpha)/a_k$ with $a_k = \{\log(e/h_k)\}^{1/2}/\log\log(e^e/h_k)$ and $b_k = \sqrt{2\log(1/h_k)}$.

2.3.3 Formal properties of the test

In Theorem 2.A.1 of the Appendix, we prove that under appropriate regularity conditions, the test defined in (2.3.7) (asymptotically) controls the familywise error rate FWER(α) for each pre-specified significance level α. As shown in Corollary 3.4.1, this has the following implication:

$$\mathbb{P}\Big(\forall (i,j,k) \in \mathcal{R} : \ (i,j,k) \notin \mathcal{M}_0\Big) \geq 1 - \alpha + o(1), \tag{2.3.8}$$

where $\mathcal{R} = \{(i,j,k) \in \mathcal{M} \text{ with } |\hat{\psi}_{ijk,T}| > c_{T,\text{Gauss}}(\alpha, h_k)\}$ is the set of triples (i,j,k) for which our test rejects the null $H_0^{(ijk)}$ and $\mathcal{M}_0$ is the set of triples (i,j,k) for which $H_0^{(ijk)}$ holds true. Verbally, (2.3.8) can be expressed as follows:

With (asymptotic) probability at least $1-\alpha$*, the null hypothesis* $H_0^{(ijk)}$ *is violated for all* $(i,j,k) \in \mathcal{M}$ *for which the test rejects* $H_0^{(ijk)}$*.* (2.3.9)

In other words:

With (asymptotic) probability at least $1-\alpha$*, the functions* λ_i *and* λ_j *differ on the interval* $\mathcal{I}_k$ *for all* $(i,j,k) \in \mathcal{M}$ *for which the test rejects* $H_0^{(ijk)}$*.* (2.3.10)

Hence, the test allows us to make simultaneous confidence statements (a) about which pairs of countries (i,j) have different trend functions and (b) about where, that is, in which time intervals $\mathcal{I}_k$ the functions differ.

According to (2.3.8), our test does not produce any false positives with high probability. In addition, we would like the test not to produce any false negatives either. Put differently, we would like the test to have high power against deviations from the null. In Proposition 2.A.4 in the Appendix, we derive the power properties of the test against a certain class of local alternatives. To summarize, we show the following: Let $\lambda_i = \lambda_{i,T}$ and $\lambda_j = \lambda_{j,T}$ be functions whose difference $\lambda_{i,T} - \lambda_{j,T}$ converges to zero as $T \to \infty$. Moreover, let $\mathcal{M}_1$ be the set of triples (i,j,k) such that either

$$\lambda_{i,T}(w) - \lambda_{j,T}(w) \geq \kappa_T \sqrt{\log T/(Th_k)} \quad \text{for all } w \in \mathcal{I}_k \tag{2.3.11}$$

or

$$\lambda_{j,T}(w) - \lambda_{i,T}(w) \geq \kappa_T \sqrt{\log T/(Th_k)} \quad \text{for all } w \in \mathcal{I}_k, \tag{2.3.12}$$

where $\{\kappa_T\}$ is any sequence of positive numbers which diverges at a faster rate than $\{\sqrt{\log T}\sqrt{\log\log T}/\log\log\log T\}$. According to Proposition 2.A.4, it holds that

$$\mathbb{P}\Big(\forall (i,j,k) \in \mathcal{M}_1 : |\hat{\psi}_{ijk,T}| > c_{T,\text{Gauss}}(\alpha, h_k)\Big) = 1 - o(1). \tag{2.3.13}$$

Hence, the test detects any local deviation from the null of the form (2.3.11) or (2.3.12) with probability tending to 1.

2.3.4 Implementation of the test in practice

For a given significance level $\alpha \in (0,1)$, the test procedure defined in (2.3.7) is implemented as follows in practice:

Step 1. Compute the quantile $q_{T,\text{Gauss}}(\alpha)$ by Monte Carlo simulations. Specifically, draw a large number N (say $N = 5000$) of samples of independent standard normal random variables $\{Z_{it}^{(\ell)} : 1 \le t \le T,\, 1 \le i \le n\}$ for $1 \le \ell \le N$. Compute the value $\Phi_T^{(\ell)}$ of the Gaussian statistic Φ_T for each sample ℓ and calculate the empirical $(1-\alpha)$-quantile $\hat{q}_{T,\text{Gauss}}(\alpha)$ from the values $\{\Phi_T^{(\ell)} : 1 \le \ell \le N\}$. Use $\hat{q}_{T,\text{Gauss}}(\alpha)$ as an approximation of the quantile $q_{T,\text{Gauss}}(\alpha)$.

Step 2. Compute the critical values $c_{T,\text{Gauss}}(\alpha, h_k)$ for $1 \le k \le K$ based on the approximation $\hat{q}_{T,\text{Gauss}}(\alpha)$.

Step 3. Carry out the test for each $(i,j,k) \in \mathcal{M}$ and store the test results in the variable $r_{ijk,T} = 1(|\hat{\psi}_{ijk,T}| > c_{T,\text{Gauss}}(\alpha, h_k))$ for each $(i,j,k) \in \mathcal{M}$, that is, let $r_{ijk,T} = 1$ if the hypothesis $H_0^{(ijk)}$ is rejected and $r_{ijk,T} = 0$ otherwise.

To graphically present the test results, we produce a plot for each pair of countries $(i,j) \in \mathcal{S}$ that shows the intervals $\mathcal{I}_k$ for which the test rejects the null $H_0^{(ijk)}$, that is, the intervals in the set $\mathcal{F}_{\text{reject}}(i,j) = \{\mathcal{I}_k \in \mathcal{F} : r_{ijk,T} = 1\}$. The plot is designed such that it graphically highlights the subset of intervals $\mathcal{F}_{\text{reject}}^{\min}(i,j) = \{\mathcal{I}_k \in \mathcal{F}_{\text{reject}}(i,j) :$ there exists no $\mathcal{I}_{k'} \in \mathcal{F}_{\text{reject}}(i,j)$ with $\mathcal{I}_{k'} \subset \mathcal{I}_k\}$. The elements of $\mathcal{F}_{\text{reject}}^{\min}(i,j)$ are called minimal intervals. By definition, there is no other interval $\mathcal{I}_{k'}$ in $\mathcal{F}_{\text{reject}}(i,j)$ which is a proper subset of a minimal interval $\mathcal{I}_k$. Hence, the minimal intervals can be regarded as those intervals in $\mathcal{F}_{\text{reject}}(i,j)$ which are most informative about the precise location of the differences between the trends λ_i and λ_j. In Section 2.4, we use the graphical device just described to present the test results of our empirical application; see panels (d) in Figures 2.3–2.6.

According to (2.3.8), we can make the following simultaneous confidence statement about the intervals in $\mathcal{F}_{\text{reject}}(i,j)$ for $(i,j) \in \mathcal{S}$:

With (asymptotic) probability at least $1-\alpha$*, it holds that for every pair of countries* $(i,j) \in \mathcal{S}$*, the functions* λ_i *and* λ_j *differ on each interval in* $\mathcal{F}_{reject}(i,j)$*.* (2.3.14)

Hence, we can claim with statistical confidence at least $1-\alpha$ that the functions λ_i and λ_j differ on each time interval which is depicted in the plots of our graphical device. Since $\mathcal{F}_{\text{reject}}^{\min}(i,j) \subseteq \mathcal{F}_{\text{reject}}(i,j)$ for any $(i,j) \in \mathcal{S}$, the confidence statement (2.3.14) trivially remains to hold true when $\mathcal{F}_{\text{reject}}(i,j)$ is replaced by $\mathcal{F}_{\text{reject}}^{\min}(i,j)$.

The graphical device described above is of course not the only way to present the test results. Another object which is helpful in summarizing the test results for a given pair of countries (i,j) is the union of minimal intervals $U_{ij} = \cup_{\mathcal{I} \in \mathcal{F}_{\text{reject}}^{\min}(i,j)} \mathcal{I}$. One can formally show that the union U_{ij} is closely related to the set $U_{ij}^* = \{u \in [0,1] : \lambda_i(u) \neq$

$\lambda_j(u)\}$ of time points where the two functions λ_i and λ_j differ from each other. For a precise mathematical statement and the technical details, we refer to Lemma 2.B.9 of the Appendix.

2.4 Empirical application to COVID-19 data

We now use our test to analyze the outbreak patterns of the COVID-19 epidemic. We proceed in two steps. In Section 2.4.1, we assess the finite sample performance of our test by Monte-Carlo experiments. Specifically, we run a series of experiments which show that the test controls the FWER at level α as predicted by the theory and that it has good power properties. In Section 2.4.2, we then apply the test to a sample of COVID-19 data from different European countries. Our multiscale test is implemented in the R package `multiscale`, available on GitHub at `https://github.com/marina-khi/multiscale`.

2.4.1 Simulation experiments

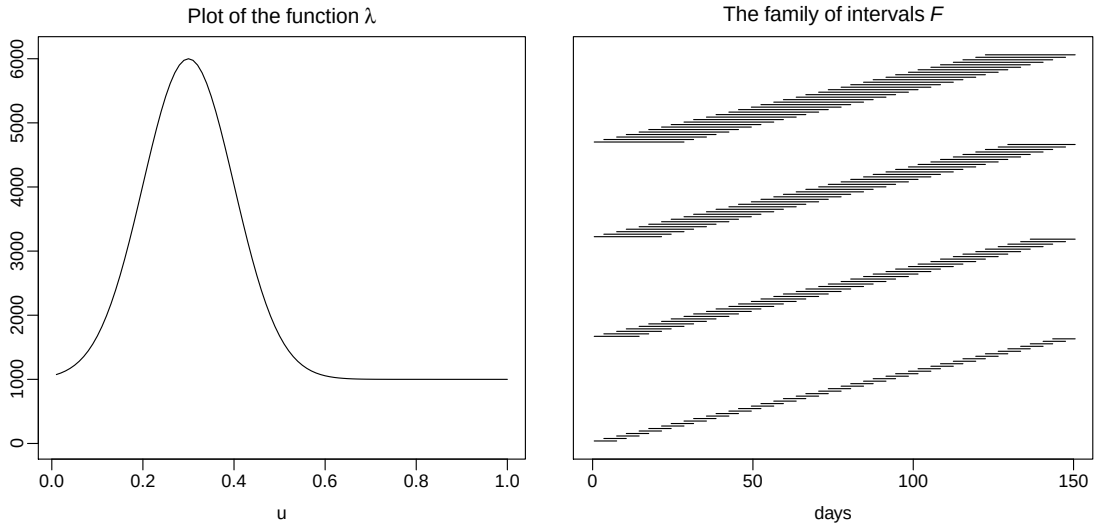
We simulate count data $\mathcal{X} = \{X_{it} : 1 \leq i \leq n, 1 \leq t \leq T\}$ by drawing the observations X_{it} independently from a negative binomial distribution with mean $\lambda_i(t/T)$ and variance $\sigma^2\lambda_i(t/T)$. By definition, X_{it} has a negative binomial distribution with parameters q and r if $\mathbb{P}(X_{it} = m) = \Gamma(m + r)/(\Gamma(r)m!)q^r(1 - q)^m$ for each $m \in \mathbb{N} \cup \{0\}$. Since $\mathbb{E}[X_{it}] = r(1 - q)/q$ and $\mathrm{Var}(X_{it}) = r(1 - q)/q^2$, we can use the parametrization $q = 1/\sigma^2$ and $r = \lambda_i(t/T)/(\sigma^2 - 1)$ to obtain that $\mathbb{E}[X_{it}] = \lambda_i(t/T)$ and $\mathrm{Var}(X_{it}) = \sigma^2\lambda_i(t/T)$. With this parametrization, the simulated data follow a nonparametric regression model of the form

$$X_{it} = \lambda_i\Big(\frac{t}{T}\Big) + \sigma\sqrt{\lambda_i\Big(\frac{t}{T}\Big)}\eta_{it},$$

where the noise variables η_{it} have zero mean and unit variance. The functions λ_i are specified below. The overdispersion parameter is set to $\sigma = 15$, which is similar to the estimate $\hat{\sigma} = 14.82$ obtained in the empirical application of Section 2.4.2. Robustness checks with $\sigma = 10$ and $\sigma = 20$ are provided in the Appendix.

We consider different values for T and n, in particular, $T \in \{100, 250, 500\}$ and $n \in \{5, 10, 50\}$. Note that in the application, we have $T = 150$ and $n = 5$. We let $\mathcal{S} = \{(i, j) : 1 \leq i < j \leq n\}$, that is, we compare all pairs of countries (i, j) with $i < j$. Moreover, we choose $\mathcal{F}$ to be a family of time intervals $\mathcal{I}_k$ with length $h_k \in \{7/T, 14/T, 21/T, 28/T\}$. Hence, the intervals in $\mathcal{F}$ have length either 7, 14, 21 or 28 days (i.e., 1, 2, 3 or 4 weeks). For each length h_k, we include all intervals that start at days $t = 1 + 7(j - 1)$ and $t = 4 + 7(j - 1)$ for $j = 1, 2, \ldots$ A graphical presentation of the family $\mathcal{F}$ for $T = 150$ (as in the application) is given in Figure 2.1b. All our simulation experiments are based on $R = 5000$ simulation runs.

In the first part of the simulation study, we examine whether our test controls the FWER as predicted by the theory. To do so, we assume that the hypothesis $H_0^{(ijk)}$ holds true for all (i, j, k) under consideration, which implies that $\lambda_i = \lambda$ for all i. We consider



(a) (b)

Figure 2.1: (a) Plot of the function λ; (b) plot of the family of intervals $\mathcal{F}$.

the function

$$\lambda(u) = 5000 \exp\Big(- \frac{(10u-3)^2}{2} \Big) + 1000, \tag{2.4.1}$$

which is similar in shape to some of the estimated trend curves in the application of Section 2.4.2. A plot of the function λ is provided in Figure 2.1a. To evaluate whether the test controls the FWER at level α, we compare the empirical size of the test with the target α. The empirical size is computed as the precentage of simulation runs in which the test falsely rejects at least one null hypothesis $H_0^{(ijk)}$.

The simulation results are reported in Table 2.1. As can be seen, the empirical size gives a reasonable approximation to the target α in all scenarios under investigation, even though the size numbers have a slight downward bias. This bias gets larger as the number of time series n increases, which reflects the fact that the test problem becomes more difficult for larger n. Already for $n = 5$, the number p of hypotheses to be tested is quite high, in particular, $p = 960, 2\,680, 5\,560$ for $T = 100, 250, 500$. This number increases to $p = 117\,600, 328\,300, 681\,100$ when $n = 50$. Hence, the dimensionality and thus the complexity of the test problem increases considerably as n gets larger. On first sight, it may seem astonishing that the downward bias does not diminish notably as the time series length T increases. This, however, has a simple explanation: The interval lengths h_k remain the same (7, 14, 21 or 28 days) as T increases, which implies that the effective sample size for computing the test statistics $\hat{\psi}_{ijk,T}$ does not change. To summarize, even though slightly conservative, the test controls the FWER quite accurately in the simulation setting at hand.

In the second part of the simulation study, we investigate the power properties of the test. To do so, we assume that $\lambda_i = \lambda$ for all $i > 1$ and that $\lambda_1 \neq \lambda$, where λ is defined in (2.4.1). Hence, only the first mean function λ_1 is different from the others. This implies

Table 2.1: Empirical size of the test for different values of n and T.

	$n = 5$			$n = 10$			$n = 50$		
	significance level α			significance level α			significance level α		
	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
$T = 100$	0.011	0.047	0.093	0.010	0.044	0.087	0.008	0.037	0.075
$T = 250$	0.009	0.047	0.091	0.009	0.046	0.087	0.008	0.035	0.069
$T = 500$	0.010	0.044	0.083	0.008	0.048	0.093	0.007	0.035	0.077

that the hypothesis $H_0^{(ijk)}$ holds true for all (i, j, k) with $i > 1$ and $j > 1$, while $H_0^{(ijk)}$ does not hold true for any pair (i, j) with either $i = 1$ or $j = 1$. We consider two different simulation scenarios. In Scenario A, the function λ_1 has the form

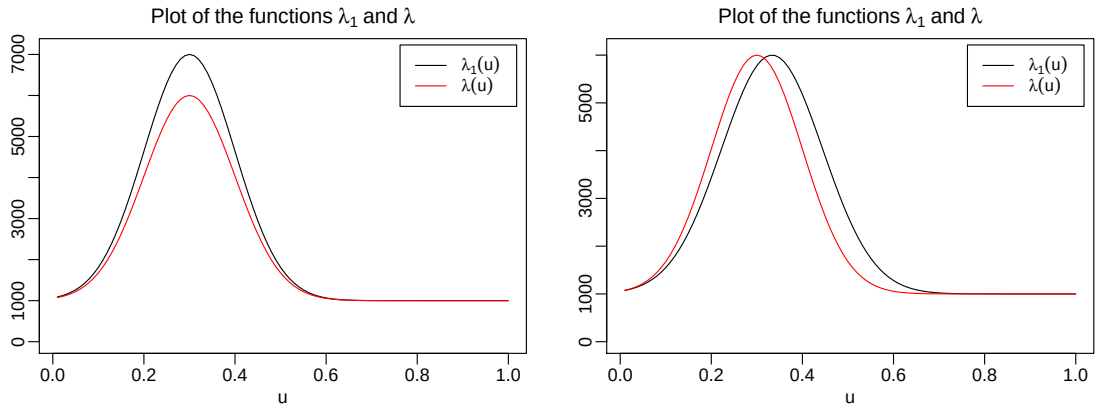
$$\lambda_1(u) = 6000 \exp\Big(- \frac{(10u - 3)^2}{2} \Big) + 1000$$

and is plotted together with λ in Figure 2.2a. As can be seen, the two functions λ_1 and λ peak at the same point in time, but the peak of λ_1 is higher than that of λ. In Scenario B, we let

$$\lambda_1(u) = 5000 \exp\Big(- \frac{(9u - 3)^2}{2} \Big) + 1000.$$

Figure 2.2b shows that the peaks of λ_1 and λ have the same height but are reached at different points in time. To evaluate the power properties of the test in Scenarios A and B, we compute the percentage of simulation runs where the test (i) correctly detects differences between λ_1 and at least one of the other mean functions and (ii) does not spuriously detect differences between the other mean functions. Put differently, we calculate the percentage of simulation runs where (i) the set $\mathcal{F}_{\text{reject}}(1, j)$ is non-empty at least for one $j \in \{2, \dots, n\}$ and (ii) all other sets $\mathcal{F}_{\text{reject}}(i, j)$ with $2 \leq i < j \leq n$ are empty. We call this percentage number the (empirical) power of the test. We thus use the term "power" a bit differently than usual.

The results for Scenario A (see Figure 2.2a) are presented in Table 2.2 and those for Scenario B (see Figure 2.2b) in Table 2.3. As can be seen, the test has substantial power in all the considered simulation settings. It is more powerful in Scenario B than in Scenario A, which is most presumably due to the fact that the differences $|\lambda_1(u) - \lambda(u)|$ are much larger in Scenario B. Moreover, it is less powerful for larger numbers of time series n, which reflects the fact that the test problem gets more high-dimensional and thus more difficult as n increases. As one would expect, the power numbers tend to become larger as the time series length T and the significance level α increase. In Scenario B (mostly for $T = 250$ and $T = 500$), however, the power numbers drop down a bit as α gets larger. This reverse dependance can be explained by the way we calculate power: we exclude simulation runs where the test spuriously detects differences between the trends in countries i and j with $i, j > 1$. The number of spurious findings increases as we make the significance level α larger, which presumably causes the slight drop in power.



(a) Scenario A

(b) Scenario B

Figure 2.2: Plot of the functions λ_1 (black) and λ (red) in the simulation scenarios A and B.

Table 2.2: Power of the test for different values of n and T in Scenario A.

	$n = 5$			$n = 10$			$n = 50$		
	significance level α			significance level α			significance level α		
	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
$T = 100$	0.335	0.518	0.597	0.306	0.474	0.545	0.212	0.352	0.418
$T = 250$	0.615	0.790	0.836	0.580	0.764	0.800	0.470	0.648	0.705
$T = 500$	0.736	0.905	0.917	0.738	0.884	0.890	0.636	0.799	0.830

Table 2.3: Power of the test for different values of n and T in Scenario B.

	$n = 5$			$n = 10$			$n = 50$		
	significance level α			significance level α			significance level α		
	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
$T = 100$	0.824	0.910	0.903	0.812	0.893	0.890	0.738	0.847	0.857
$T = 250$	0.991	0.972	0.941	0.991	0.960	0.920	0.991	0.965	0.933
$T = 500$	0.997	0.973	0.949	0.995	0.961	0.923	0.996	0.969	0.932

2.4.2 Analysis of COVID-19 data

The COVID-19 pandemic is one of the most pressing issues at present. The first outbreak occurred in Wuhan, China, in December 2019. On 30 January 2020, the World Health Organization (WHO) declared that the outbreak constitutes a Public Health Emergency of International Concern, and on 11 March 2020, the WHO characterized it as a pandemic. As of 2 February 2021, more than 102 million cases of COVID-19 infections have been reported worldwide, resulting in more than 2 million deaths.

There are many open questions surrounding the current COVID-19 pandemic. A question which is particularly relevant for governments and policy makers is whether the pandemic has developed similarly in different countries or whether there are notable differences. Identifying these differences may give some insight into which government

policies have been more effective in containing the virus than others. In what follows, we use our multiscale test to compare the development of COVID-19 in several European countries. It is important to emphasize that our test allows to identify differences in the development of the epidemic across countries in a statistically rigorous way, but it does not tell what causes these differences. By distinguishing statistically significant differences from artefacts of the sampling noise, the test provides the basis for a further investigation into the causes. Such an investigation, however, presumably goes beyond a mere statistical analysis.

2.4.2.1 Data

We analyze data from five European countries: Germany, Italy, Spain, France and the United Kingdom. For each country i, we observe a time series $\mathcal{X}_i = \{X_{it} : 1 \leq t \leq T\}$, where X_{it} is the number of newly confirmed COVID-19 cases in country i on day t. The data are freely available on the homepage of the European Center for Disease Prevention and Control (`https://www.ecdc.europa.eu`) and were downloaded on 2 February 2021.[2] As already mentioned in the Introduction, we take the first Monday after reaching 100 confirmed cases in each country as the starting date $t = 1$. Beginning the time series of each country on the day when that country reached 100 confirmed cases is a common way of "normalizing" the data (see e.g. Cohen and Kupferschmidt, 2020). Additionally aligning the data by Monday allows to take care of possible weekly cycles in the data which are produced by delays in reporting new cases over the weekend. The time series length T is taken to be equal to 150, which covers the first wave of the pandemic in all of the considered countries. The resulting dataset thus consists of $n = 5$ time series, each with $T = 150$ observations. Some of the time series contain negative values which we replaced by 0. Overall, this resulted in 4 replacements. Plots of the observed time series are presented in the upper panels (a) of Figures 2.3–2.6. As a robustness check, we have repeated the data analysis for the longer time span $T = 200$. The results are reported in Section 2.E of the Appendix.

To interpret the results produced by our multiscale test, we consider the Government Response Index (GRI) from the Oxford COVID-19 Government Response Tracker (Ox-CGRT) (Hale et al., 2020b). The GRI measures how severe the actions are that are taken by a country's government to contain the virus. It is calculated based on several common government policies such as school closures and travel restrictions. The GRI ranges from 0 to 100, with 0 corresponding to no response from the government at all and 100 corresponding to full lockdown, closure of schools and workplaces, ban on travelling, etc. Detailed information on the collection of the data for government responses and the methodology for calculating the GRI is provided in Hale et al. (2020a). Plots of the GRI time series are given in panels (c) of Figures 2.3–2.6.

[2] ECDC switched to a weekly reporting schedule for the COVID-19 situation on 17 December 2020. Hence, all daily updates have been discontinued from 14 December. The downloaded daily data set presents historical data until 14 December 2020.

2.4.2.2 Test results

We assume that the data X_{it} of each country i in our sample follow the nonparametric trend model

$$X_{it} = \lambda_i\Big(\frac{t}{T}\Big) + \sigma\sqrt{\lambda_i\Big(\frac{t}{T}\Big)}\eta_{it},$$

which was introduced in equation (3.1.1). The overdispersion parameter σ is estimated by the procedure described in Section 3.3.3, which yields the estimate $\hat{\sigma} = 14.82$. Throughout the section, we set the significance level to $\alpha = 0.05$ and implement the multiscale test in exactly the same way as in the simulation study of Section 2.4.1. In particular, we let $\mathcal{S} = \{(i,j) : 1 \leq i < j \leq 5\}$, that is, we compare all pairs of countries (i,j) with $i < j$, and we choose $\mathcal{F}$ to be the family of time intervals plotted in Figure 2.1b. Hence, all intervals in $\mathcal{F}$ have length either 7, 14, 21 or 28 days.

With the help of our multiscale method, we simultaneously test the null hypothesis $H_0^{(ijk)}$ that $\lambda_i = \lambda_j$ on the interval $\mathcal{I}_k$ for each $(i,j,k) \in \mathcal{M}$. The results are presented in Figures 2.3–2.6, each figure comparing a specific pair of countries (i,j) from our sample. For the sake of brevity, we only show the results for the pairwise comparisons of Germany with each of the four other countries. The remaining figures can be found in Section 2.D of the Appendix. Each figure splits into four panels (a)–(d). Panel (a) shows the observed time series for the two countries i and j that are compared. Panel (b) presents smoothed versions of the time series from (a), that is, it shows nonparametric kernel estimates (specifically, Nadaraya-Watson estimates) of the two trend functions λ_i and λ_j, where the bandwidth is set to 7 days and a rectangular kernel is used. Panel (c) displays the Government Response Index (GRI) of the two countries. Finally, panel (d) presents the results produced by our test: it depicts in grey the set $\mathcal{F}_{\text{reject}}(i,j)$ of all the intervals $\mathcal{I}_k$ for which the test rejects the null $H_0^{(ijk)}$. The minimal intervals in the subset $\mathcal{F}_{\text{reject}}^{\min}(i,j)$ are highlighted by a black frame. Note that according to (2.3.8), we can make the following simultaneous confidence statement about the intervals plotted in panels (d) of Figures 2.3–2.6: we can claim, with confidence of about 95%, that there is a difference between the functions λ_i and λ_j on each of these intervals.

We now have a closer look at the results in Figures 2.3–2.6. Figure 2.3 presents the comparison of Germany with Italy. The two time series of daily new cases in panel (a) can be seen to be very similar until approximately day 40. Thereafter, the German time series appears to trend downwards more strongly than the Italian one. The smoothed data in panel (b) give a similar visual impression: the kernel estimates of the German and Italian trend curves λ_i and λ_j are very close to each other until approximately day 40 but then start to differ. It is however not clear whether the differences between the two curve estimates reflect differences between the underlying trend curves or whether these are mere artefacts of sampling noise. Our test allows to clarify this issue. Inspecting panel (d), we see that the test detects significant differences between the trend curves in the time period between day 36 and 91. However, it does not find any significant differences up to day 36. Taken together, our results provide evidence that the epidemic developed

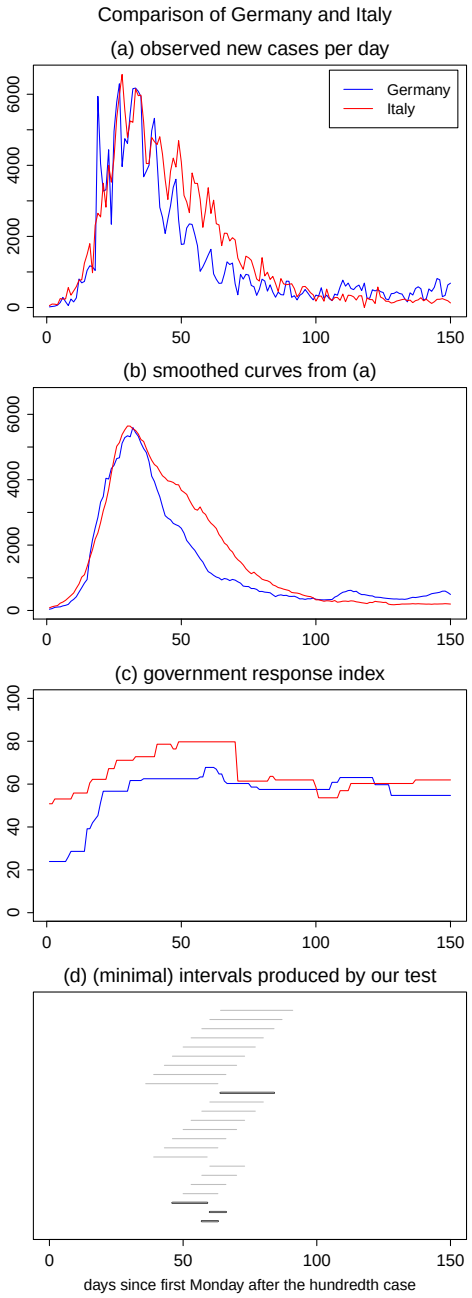


Figure 2.3: Test results for the comparison of Germany and Italy.

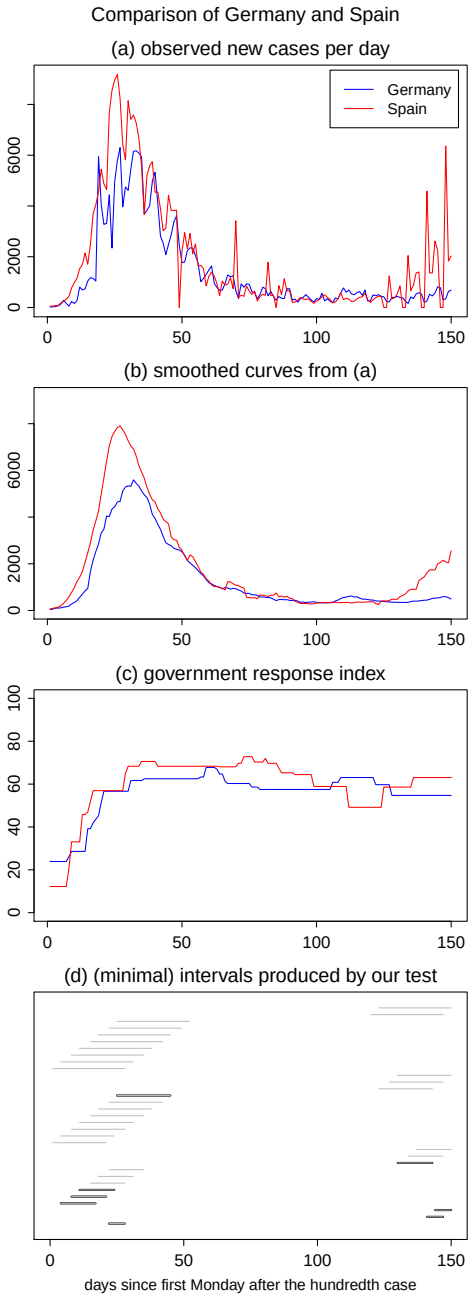


Figure 2.4: Test results for the comparison of Germany and Spain.

Note: In each figure, panel (a) shows the two observed time series, panel (b) smoothed versions of the time series, and panel (c) the corresponding Government Response Index (GRI). Panel (d) depicts the set of intervals $\mathcal{F}_{\text{reject}}(i,j)$ in grey and the subset of minimal intervals $\mathcal{F}^{\text{min}}_{\text{reject}}(i,j)$ with a black frame.

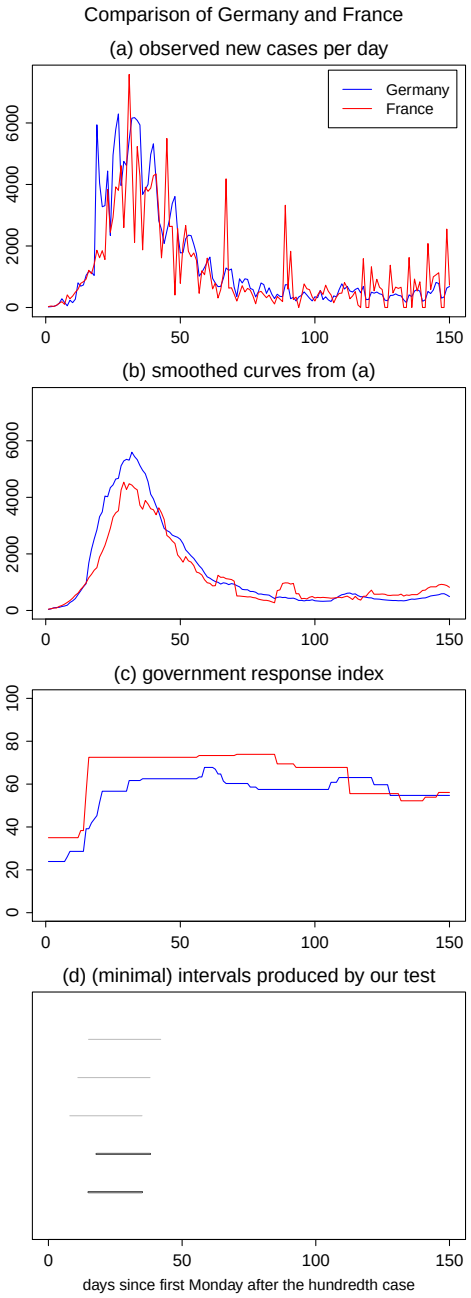


Figure 2.5: Test results for the comparison of Germany and France.

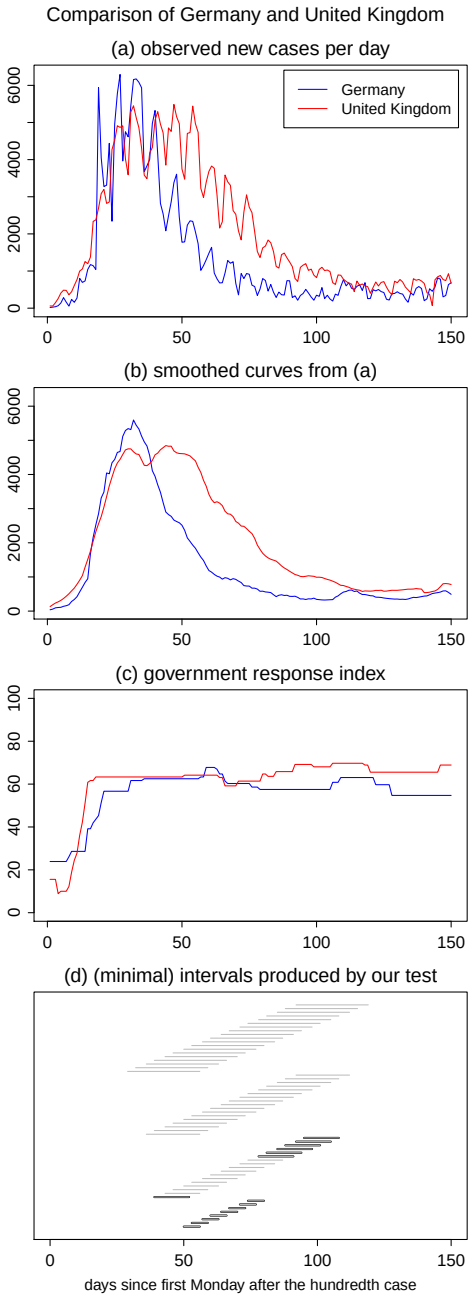


Figure 2.6: Test results for the comparison of Germany and the UK.

Note: In each figure, panel (a) shows the two observed time series, panel (b) smoothed versions of the time series, and panel (c) the corresponding Government Response Index (GRI). Panel (d) depicts the set of intervals $\mathcal{F}_{\text{reject}}(i,j)$ in grey and the subset of minimal intervals $\mathcal{F}^{\text{min}}_{\text{reject}}(i,j)$ with a black frame.

very similarly in Germany and Italy until a peak was reached around day 40. Thereafter, however, the German time series exhibits a significantly stronger downward trend than the Italian one.

A quite different picture arises when Germany is compared with Spain and France. As can be seen in Figures 2.4 and 2.5, the test detects significant differences between the German trend and the trends in Spain and France up to (approximately) day 50. This indicates that the time trends evolved differently during the outbreak of the crisis. However, the test does not find any differences in the time period between (approximately) days 50 and 120. The trends thus appear to decrease in more or less the same fashion after the first peak was reached. As can be seen in Figure 2.4, the test detects additional differences between the German and Spanish trends after day 120. This reflects the fact that the number of daily new cases in Spain picked up again after day 120, foreshadowing the second wave, whereas the numbers in Germany were still quite stable.

Finally, the comparison of Germany with the UK in Figure 2.6 reveals significant differences between the time trends in the period from (approximately) day 40 to 120. Similar to the comparison with Italy in Figure 2.3, this indicates that the trend decayed in a different fashion in Germany than in the UK after a first peak was reached. However, we do not find any significant differences between the trends during the onset of the pandemic.

2.4.2.3 Discussion

Having identified significant differences between the epidemic trends in the five countries under consideration, one may ask next what are the causes of these differences. As already mentioned at the beginning of this section, this question cannot be answered by our test. Rather, a further analysis which presumably goes beyond pure statistics is needed to shed some light on it. We here do not attempt to provide any answers. We merely discuss some observations which become apparent upon considering our test results in the light of the Government Response Index (GRI). For reasons of brevity, we focus on the comparison of Germany with Italy and Spain in Figures 2.3 and 2.4.

According to our test results in Figure 2.4, there are significant differences between the trends in Germany and Spain during the onset of the epidemic up to about day 50, with Spain having more new cases of infections than Germany on most days. After day 50, the trends become quite similar and start to decrease at approximately the same rate until around day 120. This may be due to the fact that Spain in general introduced more severe measures of lockdown than Germany (as can be seen upon inspecting the GRI in panel (c) of Figure 2.4), which may have helped to battle the spread of infection. Furthermore, around days 110–120, the measures in Spain were less strict than in Germany, which could be a reason for the detected differences between the trends towards the end of the sample. However, a much more thorough analysis is of course needed to find out whether this is indeed the case or whether other factors were mainly responsible.

Turning to the comparison of Germany and Italy, we found that the German trend drops down significantly faster than the Italian one after approximately day 40. Interestingly, the GRI of Italy almost always lies above that of Germany. Hence, even though Italy has in general taken more severe and restrictive measures against the virus than Germany, it appears that the virus could be contained better in Germany (in the sense that the trend of daily new cases went down significantly faster in Germany than in Italy). This suggests that there are indeed important factors besides the level of government response to the pandemic which substantially influence the trend of new COVID-19 cases.

This brief discussion already indicates that it is extremely difficult to determine the exact causes of the differences in epidemic trends across countries. Since even similar countries such as those in our sample differ in a variety of aspects that are relevant for the spread of the virus, it is very challenging to pin down these causes. One issue that is often discussed in the context of cross-country comparisons are country-specific strategies to test for the coronavirus. The argument is that differences between epidemic trends may be spuriously produced by country-specific test procedures.

Even though we can of course not fully exclude this possibility, our test results are presumably not driven by different test regimes in the countries under consideration. To see this, we consider again the comparison of Germany and Italy: The test regimes in these two countries are arguably quite different. Germany is often cited as the country that employed early, widespread testing with more than 100 000 tests per week even in the beginning of the pandemic (Cohen and Kupferschmidt, 2020), while testing in Italy became widespread only in the later stages of the pandemic. Nevertheless, visual inspection of the raw and smoothed data in panels (a) and (b) of Figure 2.3 suggest that the underlying time trends are very similar up to day 36. This is confirmed by our multiscale test which does not find any significant differences before that day. Hence, the different test regimes in Germany and Italy towards the beginning of the pandemic do not appear to have an overly strong effect and to produce spurious differences between the time trends. This suggests that the differences detected by our multiscale test indeed reflect differences in the way the virus spread in Germany and Italy rather than being mere artefacts of different test regimes.

APPENDICES

2.A Proofs of theoretical results

In what follows, we state and prove the main theoretical results on the multiscale test developed in Section 3.3. Throughout the Appendix, we let C be a generic positive constant that may take a different value on each occurrence. Unless stated differently, C depends neither on the time series length T nor on the dimension p of the test problem. We further use the symbols $h_{\min} := \min_{1\leq k\leq K} h_k$ and $h_{\max} := \max_{1\leq k\leq K} h_k$ to denote the smallest and largest interval length in the family $\mathcal{F}$, respectively.

The first result shows that the multiscale test asymptotically controls the FWER at level α.

Theorem 2.A.1. *Let (C1) and (C2) be satisfied. Moreover, assume that (i)* $h_{\max} = o(1/\{\log T\}^2)$*, (ii)* $h_{\min} \geq CT^{-b}$ *for some* $b \in (0,1)$*, and (iii)* $p = O(T^{(\theta/2)(1-b)-(1+\delta)})$ *for some small* $\delta > 0$*. Then for any given* $\alpha \in (0,1)$*,*

$$\mathrm{FWER}(\alpha) := \mathbb{P}\Big(\exists (i,j,k) \in \mathcal{M}_0 : |\hat{\psi}_{ijk,T}| > c_{T,\mathrm{Gauss}}(\alpha, h_k)\Big) \leq \alpha + o(1),$$

where $\mathcal{M}_0 \subseteq \mathcal{M}$ *is the set of all* $(i,j,k) \in \mathcal{M}$ *for which* $H_0^{(ijk)}$ *holds true.*

We briefly discuss the conditions (i)–(iii) on $h_{\min}$, $h_{\max}$ and p. Restriction (i) allows the maximal interval length $h_{\max}$ to converge to zero very slowly, which means that $h_{\max}$ can be picked very large in practice. According to restriction (ii), the minimal interval length $h_{\min}$ can be chosen to go to zero as any polynomial T^{-b} with some $b \in (0,1)$. Restriction (iii) allows the dimension p of the test problem to grow polynomially in T. Specifically, p may grow at most as the polynomial T^γ with $\gamma = (\theta/2)(1-b)-(1+\delta)$. As one can see, the exponent γ depends on the number of error moments θ defined in (C2) and the parameter b that specifies the minimal interval length $h_{\min}$. In particular, for any given $b \in (0,1)$, the exponent γ gets larger as θ increases. Hence, the larger the number of error moments θ, the faster p may grow in comparison to T. In the extreme case where all error moments exist, that is, where θ can be made as large as desired, p may grow as any polynomial of T, no matter how we pick $b \in (0,1)$. Thus, if the error terms have sufficiently many moments, the dimension p can be extremely large in comparison to T and the minimal interval length $h_{\min}$ can be chosen very small.

The following corollary is an immediate consequence of Theorem 2.A.1. It provides the theoretical justification needed to make simultaneous confidence statements of the form (2.3.9), (2.3.10) and (2.3.14).

Corollary 1.A.1. *Under the conditions of Theorem 2.A.1,*

$$\mathbb{P}\Big(\forall (i,j,k) \in \mathcal{R} :\ (i,j,k) \notin \mathcal{M}_0\Big) \geq 1 - \alpha + o(1),$$

where $\mathcal{R} = \{(i,j,k) \in \mathcal{M} \text{ with } |\hat{\psi}_{ijk,T}| > c_{T,\text{Gauss}}(\alpha, h_k)\}$ *is the set of triples* (i,j,k) *for which the test rejects the null* $H_0^{(ijk)}$.

The next result specifies the power of the multiscale test against a certain class of local alternatives. To formulate it, we allow the functions $\lambda_{i,T}$ and $\lambda_{j,T}$ to depend on T, that is, we consider sequences of functions $\{\lambda_{i,T}\}$ and $\{\lambda_{j,T}\}$ rather than fixed functions λ_i and λ_j.

Proposition 2.A.4. *Let the conditions of Theorem 2.A.1 be satisfied and let* $\mathcal{M}_1$ *be the set of triples* $(i,j,k) \in \mathcal{M}$ *such that either*

$$\lambda_{i,T}(w) - \lambda_{j,T}(w) \geq \kappa_T \sqrt{\log T/(Th_k)} \quad \text{for all } w \in \mathcal{I}_k \tag{2.A.1}$$

or

$$\lambda_{j,T}(w) - \lambda_{i,T}(w) \geq \kappa_T \sqrt{\log T/(Th_k)} \quad \text{for all } w \in \mathcal{I}_k, \tag{2.A.2}$$

where $\{\kappa_T\}$ *is any sequence of positive numbers for which* $\kappa_T/\ell_T \to \infty$ *with* $\ell_T = \sqrt{\log T}\sqrt{\log\log T}/\log\log\log T$. *Then*

$$\mathbb{P}\Big(\forall (i,j,k) \in \mathcal{M}_1 : |\hat{\psi}_{ijk,T}| > c_{T,\text{Gauss}}(\alpha, h_k)\Big) = 1 - o(1)$$

for any given $\alpha \in (0,1)$.

Proof of Theorem 2.A.1. The proof proceeds in several steps.

Step 1. Let $\hat{\Psi}_T = \max_{(i,j,k)\in\mathcal{M}} a_k(|\hat{\psi}^0_{ijk,T}| - b_k)$ with $\hat{\psi}^0_{ijk,T}$ introduced in (2.3.4) and define $\Psi_T = \max_{(i,j,k)\in\mathcal{M}} a_k(|\psi^0_{ijk,T}| - b_k)$ with

$$\psi^0_{ijk,T} = \frac{1}{\sqrt{2Th_k}} \sum_{t=1}^{T} 1\Big(\frac{t}{T} \in \mathcal{I}_k\Big)(\eta_{it} - \eta_{jt}).$$

To start with, we prove that

$$\big|\hat{\Psi}_T - \Psi_T\big| = o_p(r_T), \tag{2.A.3}$$

where $\{r_T\}$ is any null sequence that converges more slowly to zero than $\rho_T = \sqrt{\log T}\,\{\log p/\sqrt{Th_{\min}} + h_{\max}\sqrt{\log p}\}$, that is, $\rho_T/r_T \to 0$ as $T \to \infty$. Since the proof of (2.A.3) is rather technical and lengthy, the details are provided in Section 2.B.

Step 2. We next prove that

$$\sup_{q\in\mathbb{R}} \Big|\mathbb{P}\big(\Psi_T \leq q\big) - \mathbb{P}\big(\Phi_T \leq q\big)\Big| = o(1). \tag{2.A.4}$$

To do so, we rewrite the statistics Ψ_T and Φ_T as follows: Define

$$V_t^{(ijk)} = V_{t,T}^{(ijk)} := \sqrt{\frac{T}{2Th_k}} 1\Big(\frac{t}{T} \in \mathcal{I}_k\Big)(\eta_{it} - \eta_{jt})$$

for $(i,j,k) \in \mathcal{M}$ and let $\boldsymbol{V}_t = (V_t^{(ijk)} : (i,j,k) \in \mathcal{M})$ be the p-dimensional random vector with the entries $V_t^{(ijk)}$. With this notation, we get that $\psi_{ijk,T}^0 = T^{-1/2} \sum_{t=1}^T V_t^{(ijk)}$ and thus

$$\begin{aligned}\Psi_T &= \max_{(i,j,k)\in\mathcal{M}} a_k\big(|\psi_{ijk,T}^0| - b_k\big) \\ &= \max_{(i,j,k)\in\mathcal{M}} a_k\Big\{\Big|\frac{1}{\sqrt{T}} \sum_{t=1}^T V_t^{(ijk)}\Big| - b_k\Big\}.\end{aligned}$$

Analogously, we define

$$W_t^{(ijk)} = W_{t,T}^{(ijk)} := \sqrt{\frac{T}{2Th_k}} 1\Big(\frac{t}{T} \in \mathcal{I}_k\Big)(Z_{it} - Z_{jt})$$

with Z_{it} i.i.d. standard normal and let $\boldsymbol{W}_t = (W_t^{(ijk)} : (i,j,k) \in \mathcal{M})$. The vector $\boldsymbol{W}_t$ is a Gaussian version of $\boldsymbol{V}_t$ with the same mean and variance. In particular, $\mathbb{E}[\boldsymbol{W}_t] = \mathbb{E}[\boldsymbol{V}_t] = 0$ and $\mathbb{E}[\boldsymbol{W}_t \boldsymbol{W}_t^\top] = \mathbb{E}[\boldsymbol{V}_t \boldsymbol{V}_t^\top]$. Similarly as before, we can write $\phi_{ijk,T} = T^{-1/2} \sum_{t=1}^T W_t^{(ijk)}$ and

$$\begin{aligned}\Phi_T &= \max_{(i,j,k)\in\mathcal{M}} a_k\big(|\phi_{ijk,T}| - b_k\big) \\ &= \max_{(i,j,k)\in\mathcal{M}} a_k\Big\{\Big|\frac{1}{\sqrt{T}} \sum_{t=1}^T W_t^{(ijk)}\Big| - b_k\Big\}.\end{aligned}$$

For any $q \in \mathbb{R}$, it holds that

$$\begin{aligned}\mathbb{P}\big(\Psi_T \le q\big) &= \mathbb{P}\Big(\max_{(i,j,k)\in\mathcal{M}} a_k\Big\{\Big|\frac{1}{\sqrt{T}} \sum_{t=1}^T V_t^{(ijk)}\Big| - b_k\Big\} \le q\Big) \\ &= \mathbb{P}\Big(\Big|\frac{1}{\sqrt{T}} \sum_{t=1}^T V_t^{(ijk)}\Big| \le c_{ijk}(q) \text{ for all } (i,j,k) \in \mathcal{M}\Big) \\ &= \mathbb{P}\Big(\Big|\frac{1}{\sqrt{T}} \sum_{t=1}^T \boldsymbol{V}_t\Big| \le \boldsymbol{c}(q)\Big),\end{aligned}$$

where $\boldsymbol{c}(q) = (c_{ijk}(q) : (i,j,k) \in \mathcal{M})$ is the $\mathbb{R}^p$-vector with the entries $c_{ijk}(q) = q/a_k + b_k$, we use the notation $|v| = (|v_1|, \ldots, |v_p|)^\top$ for vectors $v \in \mathbb{R}^p$ and the inequality $v \le w$ is to be understood componentwise for $v, w \in \mathbb{R}^p$. Analogously, we have

$$\mathbb{P}\big(\Phi_T \le q\big) = \mathbb{P}\Big(\Big|\frac{1}{\sqrt{T}} \sum_{t=1}^T \boldsymbol{W}_t\Big| \le \boldsymbol{c}(q)\Big).$$

With this notation at hand, we can make use of Proposition 2.1 from Chernozhukov et al. (2017). In our context, this proposition can be stated as follows:

Proposition 2.A.5. *Assume that*

(a) $T^{-1}\sum_{t=1}^{T}\mathbb{E}(V_t^{(ijk)})^2 \geq \delta > 0$ *for all* $(i,j,k) \in \mathcal{M}$.

(b) $T^{-1}\sum_{t=1}^{T}\mathbb{E}[|V_t^{(ijk)}|^{2+r}] \leq B_T^r$ *for all* $(i,j,k) \in \mathcal{M}$ *and* $r = 1, 2$, *where* $B_T \geq 1$ *are constants that may tend to infinity as* $T \to \infty$.

(c) $\mathbb{E}[\{\max_{(i,j,k)\in\mathcal{M}} |V_t^{(ijk)}|/B_T\}^\theta] \leq 2$ *for all* t *and some* $\theta > 4$.

Then

$$\sup_{\boldsymbol{c}\in\mathbb{R}^p}\Big|\mathbb{P}\Big(\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{V}_t\Big| \leq \boldsymbol{c}\Big) - \mathbb{P}\Big(\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{W}_t\Big| \leq \boldsymbol{c}\Big)\Big| \leq C\Big\{\Big(\frac{B_T^2\log^7(pT)}{T}\Big)^{1/6} + \Big(\frac{B_T^2\log^3(pT)}{T^{1-2/\theta}}\Big)^{1/3}\Big\}, \tag{2.A.5}$$

where C *depends only on* δ *and* θ.

It is straightforward to verify that assumptions (a)–(c) are satisfied under the conditions of Theorem 2.A.1 for sufficiently large T, where B_T can be chosen as $B_T = Cp^{1/\theta}h_{\min}^{-1/2}$ with C sufficiently large. Moreover, it can be shown that the right-hand side of (2.A.5) is $o(1)$ for this choice of B_T. Hence, Proposition 2.A.5 yields that

$$\sup_{\boldsymbol{c}\in\mathbb{R}^p}\Big|\mathbb{P}\Big(\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{V}_t\Big| \leq \boldsymbol{c}\Big) - \mathbb{P}\Big(\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{W}_t\Big| \leq \boldsymbol{c}\Big)\Big| = o(1),$$

which in turn implies (2.A.4).

Step 3. With the help of (2.A.3) and (2.A.4), we now show that

$$\sup_{q\in\mathbb{R}}\Big|\mathbb{P}\big(\hat{\Psi}_T \leq q\big) - \mathbb{P}\big(\Phi_T \leq q\big)\Big| = o(1). \tag{2.A.6}$$

To start with, the above supremum can be bounded by

$$\begin{aligned}
&\sup_{q\in\mathbb{R}}\Big|\mathbb{P}\big(\hat{\Psi}_T \leq q\big) - \mathbb{P}\big(\Phi_T \leq q\big)\Big| \\
&= \sup_{q\in\mathbb{R}}\Big|\mathbb{P}\Big(\Psi_T \leq q + \big\{\Psi_T - \hat{\Psi}_T\big\}\Big) - \mathbb{P}\big(\Phi_T \leq q\big)\Big| \\
&\leq \sup_{q\in\mathbb{R}}\max\Big\{\Big|\mathbb{P}\Big(\Psi_T \leq q + \big|\Psi_T - \hat{\Psi}_T\big|\Big) - \mathbb{P}\big(\Phi_T \leq q\big)\Big|, \\
&\qquad\qquad \Big|\mathbb{P}\Big(\Psi_T \leq q - \big|\Psi_T - \hat{\Psi}_T\big|\Big) - \mathbb{P}\big(\Phi_T \leq q\big)\Big|\Big\} \\
&\leq \sup_{q\in\mathbb{R}}\max\Big\{\Big|\mathbb{P}\Big(\Psi_T \leq q + r_T\Big) - \mathbb{P}\big(\Phi_T \leq q\big)\Big| + \mathbb{P}\Big(\big|\Psi_T - \hat{\Psi}_T\big| > r_T\Big), \\
&\qquad\qquad \Big|\mathbb{P}\Big(\Psi_T \leq q - r_T\Big) - \mathbb{P}\big(\Phi_T \leq q\big)\Big| + \mathbb{P}\Big(\big|\Psi_T - \hat{\Psi}_T\big| > r_T\Big)\Big\} \\
&\leq \max_{\ell=0,1}\sup_{q\in\mathbb{R}}\Big|\mathbb{P}\Big(\Psi_T \leq q + (-1)^\ell r_T\Big) - \mathbb{P}\big(\Phi_T \leq q\big)\Big| + \mathbb{P}\Big(\big|\Psi_T - \hat{\Psi}_T\big| > r_T\Big)
\end{aligned}$$

$$= \max_{\ell=0,1} \sup_{q\in\mathbb{R}} \Big|\mathbb{P}\Big(\Psi_T \le q + (-1)^\ell r_T\Big) - \mathbb{P}\big(\Phi_T \le q\big)\Big| + o(1), \tag{2.A.7}$$

where the last line is by (2.A.3). Moreover, for $\ell = 0, 1$,

$$\begin{aligned}
&\sup_{q\in\mathbb{R}} \Big|\mathbb{P}\Big(\Psi_T \le q + (-1)^\ell r_T\Big) - \mathbb{P}\big(\Phi_T \le q\big)\Big| \\
&\le \sup_{q\in\mathbb{R}} \Big|\mathbb{P}\Big(\Psi_T \le q + (-1)^\ell r_T\Big) - \mathbb{P}\Big(\Phi_T \le q + (-1)^\ell r_T\Big)\Big| \\
&\quad + \sup_{q\in\mathbb{R}} \Big|\mathbb{P}\Big(\Phi_T \le q + (-1)^\ell r_T\Big) - \mathbb{P}\big(\Phi_T \le q\big)\Big| \\
&= \sup_{q\in\mathbb{R}} \Big|\mathbb{P}\Big(\Phi_T \le q + (-1)^\ell r_T\Big) - \mathbb{P}\big(\Phi_T \le q\big)\Big| + o(1),
\end{aligned} \tag{2.A.8}$$

the last line following from (2.A.4). Finally, by Nazarov's inequality (Nazarov, 2003 and Lemma A.1 in Chernozhukov et al., 2017), we have that for $\ell = 0, 1$,

$$\begin{aligned}
&\sup_{q\in\mathbb{R}} \Big|\mathbb{P}\Big(\Phi_T \le q + (-1)^\ell r_T\Big) - \mathbb{P}\big(\Phi_T \le q\big)\Big| \\
&= \sup_{q\in\mathbb{R}} \Big|\mathbb{P}\Big(\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \boldsymbol{W}_t\Big| \le \boldsymbol{c}(q + (-1)^\ell r_T)\Big) - \mathbb{P}\Big(\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \boldsymbol{W}_t\Big| \le \boldsymbol{c}(q)\Big)\Big| \\
&\le C\frac{r_T\sqrt{\log(2p)}}{\min_{1\le k\le K} a_k} \le C r_T \sqrt{\log\log T}\sqrt{\log(2p)},
\end{aligned} \tag{2.A.9}$$

where C is a constant that depends only on the parameter δ defined in condition (a) of Proposition 2.A.5 and we have used the fact that $\min_k a_k \ge c/\sqrt{\log\log T}$ for some $c > 0$. Inserting (2.A.8) and (2.A.9) into equation (2.A.7) completes the proof of (2.A.6).

Step 4. By definition of the quantile $q_{T,\mathrm{Gauss}}(\alpha)$, it holds that $\mathbb{P}(\Phi_T \le q_{T,\mathrm{Gauss}}(\alpha)) \ge 1-\alpha$. As shown in Section 2.B, we even have that

$$\mathbb{P}(\Phi_T \le q_{T,\mathrm{Gauss}}(\alpha)) = 1 - \alpha \tag{2.A.10}$$

for any $\alpha \in (0, 1)$. From this and (2.A.6), it immediately follows that

$$\mathbb{P}\big(\hat{\Psi}_T \le q_{T,\mathrm{Gauss}}(\alpha)\big) = 1 - \alpha + o(1), \tag{2.A.11}$$

which in turn implies that

$$\begin{aligned}
\mathrm{FWER}(\alpha) &= \mathbb{P}\Big(\exists (i,j,k) \in \mathcal{M}_0 : |\hat{\psi}_{ijk,T}| > c_{T,\mathrm{Gauss}}(\alpha, h_k)\Big) \\
&= \mathbb{P}\Big(\max_{(i,j,k)\in\mathcal{M}_0} a_k\big(|\hat{\psi}_{ijk,T}| - b_k\big) > q_{T,\mathrm{Gauss}}(\alpha)\Big) \\
&= \mathbb{P}\Big(\max_{(i,j,k)\in\mathcal{M}_0} a_k\big(|\hat{\psi}^0_{ijk,T}| - b_k\big) > q_{T,\mathrm{Gauss}}(\alpha)\Big) \\
&\le \mathbb{P}\Big(\max_{(i,j,k)\in\mathcal{M}} a_k\big(|\hat{\psi}^0_{ijk,T}| - b_k\big) > q_{T,\mathrm{Gauss}}(\alpha)\Big)
\end{aligned}$$

$$= \mathbb{P}\big(\hat{\Psi}_T > q_{T,\text{Gauss}}(\alpha)\big) = \alpha + o(1).$$

This completes the proof of Theorem 2.A.1. □

Proof of Corollary 3.4.1. With the help of Theorem 2.A.1, we obtain that

$$\begin{aligned} 1 - \alpha + o(1) &\leq 1 - \text{FWER}(\alpha) \\ &= \mathbb{P}\Big(\forall (i,j,k) \in \mathcal{M}_0 : |\hat{\psi}_{ijk,T}| \leq c_{T,\text{Gauss}}(\alpha, h_k)\Big) \\ &\leq \mathbb{P}\Big(\forall (i,j,k) \in \mathcal{R} : (i,j,k) \notin \mathcal{M}_0\Big), \end{aligned}$$

which is the statement of Corollary 3.4.1. □

Proof of Proposition 2.A.4. To start with, note that

$$c\frac{1}{\sqrt{\log\log T}} \leq a_k \leq C\frac{\sqrt{\log T}}{\log\log\log T} \quad \text{and} \quad b_k \leq C\sqrt{\log T} \tag{2.A.12}$$

with appropriately chosen constants c and C. We decompose the statistics $\hat{\psi}_{ijk,T}$ into two parts. In particular, we write $\hat{\psi}_{ijk,T} = \hat{\psi}^A_{ijk,T} + \hat{\psi}^B_{ijk,T}$ with

$$\begin{aligned} \hat{\psi}^A_{ijk,T} &= \frac{\sigma \sum_{t=1}^T 1(\frac{t}{T} \in \mathcal{I}_k)\Big(\sqrt{\lambda_i\big(\frac{t}{T}\big)}\eta_{it} - \sqrt{\lambda_j\big(\frac{t}{T}\big)}\eta_{jt}\Big)}{\hat{\sigma}\{\sum_{t=1}^T 1(\frac{t}{T} \in \mathcal{I}_k)(X_{it} + X_{jt})\}^{1/2}} \\ \hat{\psi}^B_{ijk,T} &= \frac{\sum_{t=1}^T 1(\frac{t}{T} \in \mathcal{I}_k)\big(\lambda_i\big(\frac{t}{T}\big) - \lambda_j\big(\frac{t}{T}\big)\big)}{\hat{\sigma}\{\sum_{t=1}^T 1(\frac{t}{T} \in \mathcal{I}_k)(X_{it} + X_{jt})\}^{1/2}}. \end{aligned}$$

As we will prove below, it holds that

$$\min_{(i,j,k)\in\mathcal{M}_1} |\hat{\psi}^B_{ijk,T}| \geq C\kappa_T\sqrt{\log T} \quad \text{with prob. approaching 1} \tag{2.A.13}$$

for some sufficiently small constant $C > 0$ and

$$\max_{(i,j,k)\in\mathcal{M}} |\hat{\psi}^A_{ijk,T}| = O_p(\sqrt{\log T}). \tag{2.A.14}$$

From (2.A.13) and (2.A.14), it follows that

$$\begin{aligned} \min_{(i,j,k)\in\mathcal{M}_1} a_k(|\hat{\psi}_{ijk,T}| - b_k) \geq \min_{(i,j,k)\in\mathcal{M}_1} a_k|\hat{\psi}^B_{ijk,T}| \\ - \max_{(i,j,k)\in\mathcal{M}} a_k(|\hat{\psi}^A_{ijk,T}| + b_k) \geq C\frac{\kappa_T\sqrt{\log T}}{\sqrt{\log\log T}} \end{aligned} \tag{2.A.15}$$

with probability tending to 1, where we have used the bounds on a_k and b_k from (2.A.12) and the assumption that $\kappa_T/\ell_T \to \infty$ with ℓ_T defined in Proposition 2.A.4. It further holds that

$$q_{T,\text{Gauss}}(\alpha) \leq C\frac{\log T}{\log\log\log T} \tag{2.A.16}$$

with a sufficiently large constant C, since

$$\begin{aligned}
&\mathbb{P}\Big(\max_{(i,j,k)\in\mathcal{M}} a_k\big(|\phi_{ijk,T}| - b_k\big) \le C\frac{\log T}{\log\log\log T}\Big)\\
&\ge \mathbb{P}\Big(\max_{(i,j,k)\in\mathcal{M}} a_k \max_{(i,j,k)\in\mathcal{M}} \big(|\phi_{ijk,T}| - b_k\big) \le C\frac{\log T}{\log\log\log T}\Big)\\
&\ge \mathbb{P}\Big(\frac{\sqrt{\log T}}{\log\log\log T}\max_{(i,j,k)\in\mathcal{M}} |\phi_{ijk,T}| \le C\frac{\log T}{\log\log\log T}\Big)\\
&= \mathbb{P}\Big(\max_{(i,j,k)\in\mathcal{M}} |\phi_{ijk,T}| \le C\sqrt{\log T}\Big) \ge 1-\alpha,
\end{aligned}$$

where the last inequality is a consequence of the fact that the terms $\phi_{ijk,T}$ are normally distributed random variables and $|\mathcal{M}| = p \le CT^\gamma$. From (2.A.15), (2.A.16) and the assumption that $\kappa_T/\ell_T \to \infty$, we can finally conclude that

$$\begin{aligned}
&\mathbb{P}\Big(\forall(i,j,k)\in\mathcal{M}_1 : |\hat\psi_{ijk,T}| > c_{T,\mathrm{Gauss}}(\alpha,h_k)\Big)\\
&\quad= \mathbb{P}\Big(\min_{(i,j,k)\in\mathcal{M}_1} a_k\big(|\hat\psi_{ijk,T}| - b_k\big) > q_{T,\mathrm{Gauss}}(\alpha)\Big) = 1-o(1),
\end{aligned}$$

which is the statement of Proposition 2.A.4.

It remains to prove (2.A.13) and (2.A.14). From (2.B.10) in Section 2.B together with some straightforward arguments, it follows that for any fixed $\delta > 0$,

$$\min_{(i,j,k)\in\mathcal{M}} \Big\{\frac{1}{Th_k}\sum_{t=1}^{T} 1\Big(\frac{t}{T}\in\mathcal{I}_k\Big)(X_{it}+X_{jt})\Big\} \ge (2-\delta)\lambda_{\min} \tag{2.A.17}$$

$$\max_{(i,j,k)\in\mathcal{M}} \Big\{\frac{1}{Th_k}\sum_{t=1}^{T} 1\Big(\frac{t}{T}\in\mathcal{I}_k\Big)(X_{it}+X_{jt})\Big\} \le (2+\delta)\lambda_{\max} \tag{2.A.18}$$

with probability tending to 1. Since $\hat\sigma^2 = \sigma^2 + O_p(\sqrt{\log p/T})$ by Lemma 2.B.8, it further holds that with probability tending to 1,

$$(1-\delta)\sigma \le \hat\sigma \le (1+\delta)\sigma \tag{2.A.19}$$

for any fixed $\delta > 0$. Taking into account that for any $(i,j,k) \in \mathcal{M}_1$, either $\lambda_{i,T}(w) - \lambda_{j,T}(w) \ge \kappa_T\sqrt{\log T/(Th_k)}$ or $\lambda_{j,T}(w) - \lambda_{i,T}(w) \ge \kappa_T\sqrt{\log T/(Th_k)}$ for all $w \in \mathcal{I}_k$, we can use (2.A.18) and (2.A.19) to obtain that

$$\min_{(i,j,k)\in\mathcal{M}_1} \big|\hat\psi^B_{ijk,T}\big| \ge \frac{\kappa_T\sqrt{\log T}}{(1+\delta)\sigma\sqrt{(2+\delta)\lambda_{\max}}} = C\kappa_T\sqrt{\log T}$$

with probability tending to 1. Moreover, with the help of (2.A.17), (2.A.19) and analogous arguments as for the proof of (2.B.6) in Section 2.B, we can show that

$$\max_{(i,j,k)\in\mathcal{M}} \big|\hat\psi^A_{ijk,T}\big| = O_p\big(\sqrt{\log p}\big) = O_p\big(\sqrt{\log T}\big),$$

where the last equation is due to the fact that $p = O(T^\gamma)$ for a fixed $\gamma > 0$. □

2.B Technical details

In what follows, we provide the technical details omitted in the Section 2.A. To start with, we prove the following auxiliary lemma.

Lemma 2.B.8. *Under the conditions of Theorem 2.A.1, it holds that*

$$\big|\hat{\sigma}^2 - \sigma^2\big| = O_p\Big(\sqrt{\frac{\log p}{T}}\Big).$$

Proof of Lemma 2.B.8. By definition, $\hat{\sigma}^2 = |\mathcal{C}|^{-1}\sum_{i\in\mathcal{C}}\hat{\sigma}_i^2$ and $\hat{\sigma}_i^2 = \{\sum_{t=2}^T (X_{it} - X_{it-1})^2\}$
$\{2\sum_{t=1}^T X_{it}\}^{-1}$. It holds that

$$\frac{1}{T}\sum_{t=2}^T (X_{it} - X_{it-1})^2 = \frac{\sigma^2}{T}\sum_{t=2}^T \lambda_i\Big(\frac{t}{T}\Big)(\eta_{it} - \eta_{it-1})^2 + \big\{R_{i,T}^{(1)} + \ldots + R_{i,T}^{(5)}\big\}, \tag{2.B.1}$$

where

$$\begin{aligned}
R_{i,T}^{(1)} &= \frac{2\sigma}{T}\sum_{t=2}^T \Big(\lambda_i\Big(\frac{t}{T}\Big) - \lambda_i\Big(\frac{t-1}{T}\Big)\Big)\sqrt{\lambda_i\Big(\frac{t}{T}\Big)}(\eta_{it} - \eta_{it-1}) \\
R_{i,T}^{(2)} &= \frac{2\sigma^2}{T}\sum_{t=2}^T \Big(\sqrt{\lambda_i\Big(\frac{t}{T}\Big)} - \sqrt{\lambda_i\Big(\frac{t-1}{T}\Big)}\Big)\sqrt{\lambda_i\Big(\frac{t}{T}\Big)}\eta_{it-1}(\eta_{it} - \eta_{it-1}) \\
R_{i,T}^{(3)} &= \frac{1}{T}\sum_{t=2}^T \Big(\lambda_i\Big(\frac{t}{T}\Big) - \lambda_i\Big(\frac{t-1}{T}\Big)\Big)^2 \\
R_{i,T}^{(4)} &= \frac{2\sigma}{T}\sum_{t=2}^T \Big(\lambda_i\Big(\frac{t}{T}\Big) - \lambda_i\Big(\frac{t-1}{T}\Big)\Big)\Big(\sqrt{\lambda_i\Big(\frac{t}{T}\Big)} - \sqrt{\lambda_i\Big(\frac{t-1}{T}\Big)}\Big)\eta_{it-1} \\
R_{i,T}^{(5)} &= \frac{\sigma^2}{T}\sum_{t=2}^T \Big(\sqrt{\lambda_i\Big(\frac{t}{T}\Big)} - \sqrt{\lambda_i\Big(\frac{t-1}{T}\Big)}\Big)^2\eta_{it-1}^2.
\end{aligned}$$

With the help of an exponential inequality and standard arguments, it can be shown that

$$\max_{i\in\mathcal{C}}\Big|\frac{1}{T}\sum_{t=2}^T w_i\Big(\frac{t}{T}\Big)\big\{g(\eta_{it},\eta_{it-1}) - \mathbb{E}g(\eta_{it},\eta_{it-1})\big\}\Big| = O_p\Big(\sqrt{\frac{\log p}{T}}\Big),$$

where we let $g(x,y) = x$, $g(x,y) = y$, $g(x,y) = |x|$, $g(x,y) = |y|$, $g(x,y) = x^2$, $g(x,y) = y^2$ or $g(x,y) = xy$, and $w_i(t/T)$ are deterministic weights with the property that $|w_i(t/T)| \leq w_{\max} < \infty$ for all i, t and T and some positive constant $w_{\max}$. Using this uniform

convergence result along with conditions (C1) and (C2), we obtain that

$$\max_{i\in\mathcal{C}}\Big|\frac{1}{T}\sum_{t=2}^{T}\lambda_i\Big(\frac{t}{T}\Big)(\eta_{it}-\eta_{it-1})^2-\frac{2}{T}\sum_{t=1}^{T}\lambda_i\Big(\frac{t}{T}\Big)\Big|=O_p\Big(\sqrt{\frac{\log p}{T}}\Big)$$

and

$$\max_{1\le\ell\le 5}\max_{i\in\mathcal{C}}|R_{i,T}^{(\ell)}|=O_p(T^{-1}).$$

Applying these two statements to (2.B.1), we can infer that

$$\max_{i\in\mathcal{C}}\Big|\frac{1}{T}\sum_{t=2}^{T}(X_{it}-X_{it-1})^2-\frac{2\sigma^2}{T}\sum_{t=1}^{T}\lambda_i\Big(\frac{t}{T}\Big)\Big|=O_p\Big(\sqrt{\frac{\log p}{T}}\Big).\tag{2.B.2}$$

By similar but simpler arguments, we additionally get that

$$\max_{i\in\mathcal{C}}\Big|\frac{1}{T}\sum_{t=1}^{T}X_{it}-\frac{1}{T}\sum_{t=1}^{T}\lambda_i\Big(\frac{t}{T}\Big)\Big|=O_p\Big(\sqrt{\frac{\log p}{T}}\Big).\tag{2.B.3}$$

From (2.B.2) and (2.B.3), it follows that $\max_{i\in\mathcal{C}}|\hat{\sigma}_i^2-\sigma^2|=O_p(\sqrt{\log p/T})$, which in turn implies that $|\hat{\sigma}^2-\sigma^2|=O_p(\sqrt{\log p/T})$ as well. □

Proof of (2.A.3). Since

$$\begin{aligned}\big|\hat{\Psi}_T-\Psi_T\big|&\le\max_{(i,j,k)\in\mathcal{M}}a_k\big|\hat{\psi}_{ijk,T}^0-\psi_{ijk,T}^0\big|\\&\le\max_{1\le k\le K}a_k\max_{(i,j,k)\in\mathcal{M}}\big|\hat{\psi}_{ijk,T}^0-\psi_{ijk,T}^0\big|\\&\le C\sqrt{\log T}\max_{(i,j,k)\in\mathcal{M}}\big|\hat{\psi}_{ijk,T}^0-\psi_{ijk,T}^0\big|,\end{aligned}$$

it suffices to prove that

$$\max_{(i,j,k)\in\mathcal{M}}\big|\hat{\psi}_{ijk,T}^0-\psi_{ijk,T}^0\big|=o_p\Big(\frac{r_T}{\sqrt{\log T}}\Big).\tag{2.B.4}$$

To start with, we reformulate $\hat{\psi}_{ijk,T}^0$ as

$$\hat{\psi}_{ijk,T}^0=\hat{\psi}_{ijk,T}^*+\Big(\frac{\sigma}{\hat{\sigma}}-1\Big)\hat{\psi}_{ijk,T}^*,$$

where

$$\hat{\psi}_{ijk,T}^*=\frac{\sum_{t=1}^{T}1(\frac{t}{T}\in\mathcal{I}_k)\overline{\lambda}_{ij}^{1/2}(\frac{t}{T})(\eta_{it}-\eta_{jt})}{\{\sum_{t=1}^{T}1(\frac{t}{T}\in\mathcal{I}_k)(X_{it}+X_{jt})\}^{1/2}}.$$

With this notation, we can establish the bound

$$\max_{(i,j,k)\in\mathcal{M}}\big|\hat{\psi}_{ijk,T}^0-\psi_{ijk,T}^0\big|\le\max_{(i,j,k)\in\mathcal{M}}\big|\hat{\psi}_{ijk,T}^*-\psi_{ijk,T}^0\big|$$

$$+ \Big|\frac{\sigma}{\hat{\sigma}} - 1\Big| \max_{(i,j,k)\in\mathcal{M}} \big|\hat{\psi}^*_{ijk,T} - \psi^0_{ijk,T}\big|$$
$$+ \Big|\frac{\sigma}{\hat{\sigma}} - 1\Big| \max_{(i,j,k)\in\mathcal{M}} \big|\psi^0_{ijk,T}\big|,$$

which shows that (2.B.4) is implied by the three statements

$$\max_{(i,j,k)\in\mathcal{M}} \big|\hat{\psi}^*_{ijk,T} - \psi^0_{ijk,T}\big| = O_p\Big(\frac{\log p}{\sqrt{T h_{\min}}} + h_{\max}\sqrt{\log p}\Big) \tag{2.B.5}$$
$$\max_{(i,j,k)\in\mathcal{M}} \big|\psi^0_{ijk,T}\big| = O_p\big(\sqrt{\log p}\big) \tag{2.B.6}$$
$$\big|\hat{\sigma}^2 - \sigma^2\big| = O_p\Big(\sqrt{\frac{\log p}{T}}\Big). \tag{2.B.7}$$

Since (2.B.7) has already been verified in Lemma 2.B.8, it remains to prove the statements (2.B.5) and (2.B.6).

We start with the proof of (2.B.6). Applying an exponential inequality along with standard arguments yields that

$$\max_{i\in\mathcal{C}} \max_{1\le k\le K} \Big|\frac{1}{\sqrt{T h_k}} \sum_{t=1}^{T} 1\Big(\frac{t}{T} \in \mathcal{I}_k\Big) w_i\Big(\frac{t}{T}\Big)\eta_{it}\Big| = O_p\big(\sqrt{\log p}\big), \tag{2.B.8}$$

where $w_i(t/T)$ are general deterministic weights with the property that $|w_i(t/T)| \le w_{\max} < \infty$ for all i, t and T and some positive constant $w_{\max}$. This immediately implies (2.B.6).

We next turn to the proof of (2.B.5). As the functions λ_i are uniformly Lipschitz continuous by (C1), it can be shown that

$$\max_{i\in\mathcal{C}} \max_{1\le k\le K} \Big|\frac{1}{T h_k} \sum_{t=1}^{T} 1\Big(\frac{t}{T} \in \mathcal{I}_k\Big)\lambda_i\Big(\frac{t}{T}\Big) - \frac{1}{h_k}\int_{w\in\mathcal{I}_k} \lambda_i(w)dw\Big| \le \frac{C}{T h_{\min}}. \tag{2.B.9}$$

From this, the uniform convergence result (2.B.8) and condition (C1), we can infer that

$$\max_{(i,j,k)\in\mathcal{M}} \Big|\frac{1}{T h_k} \sum_{t=1}^{T} 1\Big(\frac{t}{T} \in \mathcal{I}_k\Big)(X_{it} + X_{jt}) - \frac{1}{h_k}\int_{w\in\mathcal{I}_k} \big\{\lambda_i(w) + \lambda_j(w)\big\}dw\Big| = O_p\Big(\sqrt{\frac{\log p}{T h_{\min}}}\Big) \tag{2.B.10}$$

and

$$\max_{(i,j,k)\in\mathcal{M}} \Big|\frac{1}{\sqrt{T h_k}} \sum_{t=1}^{T} 1\Big(\frac{t}{T} \in \mathcal{I}_k\Big)\overline{\lambda}_{ij}^{1/2}\Big(\frac{t}{T}\Big)(\eta_{it} - \eta_{jt}) - \Big\{\frac{\int_{w\in\mathcal{I}_k} \overline{\lambda}_{ij}(w)dw}{h_k}\Big\}^{1/2} \frac{1}{\sqrt{T h_k}} \sum_{t=1}^{T} 1\Big(\frac{t}{T} \in \mathcal{I}_k\Big)(\eta_{it} - \eta_{jt})\Big| = O_p\Big(h_{\max}\sqrt{\log p}\Big). \tag{2.B.11}$$

The claim (2.B.5) follows from (2.B.10) and (2.B.11) along with straightforward calculations. □

Proof of (2.A.10). The proof is by contradiction. Suppose that (2.A.10) does not hold true, that is, $\mathbb{P}(\Phi_T \leq q_{T,\text{Gauss}}(\alpha)) = 1 - \alpha + \xi$ for some $\xi > 0$. By Nazarov's inequality,

$$\begin{aligned}\mathbb{P}\big(\Phi_T \leq q_{T,\text{Gauss}}(\alpha)\big) - \mathbb{P}\big(\Phi_T \leq q_{T,\text{Gauss}}(\alpha) - \eta\big) &\leq C\frac{\eta\sqrt{\log(2p)}}{\min_{1\leq k\leq K} a_k} \\ &\leq C\eta\sqrt{\log\log T}\sqrt{\log(2p)}\end{aligned}$$

for any $\eta > 0$, where the last inequality uses the fact that $\min_k a_k \geq c/\sqrt{\log\log T}$ for some $c > 0$. Hence,

$$\begin{aligned}\mathbb{P}\big(\Phi_T \leq q_{T,\text{Gauss}}(\alpha) - \eta\big) &\geq \mathbb{P}\big(\Phi_T \leq q_{T,\text{Gauss}}(\alpha)\big) - C\eta\sqrt{\log\log T}\sqrt{\log(2p)} \\ &= 1 - \alpha + \xi - C\eta\sqrt{\log\log T}\sqrt{\log(2p)} > 1 - \alpha\end{aligned}$$

for $\eta > 0$ sufficiently small. This contradicts the definition of the quantile $q_{T,\text{Gauss}}(\alpha)$ according to which $q_{T,\text{Gauss}}(\alpha) = \inf_{q\in\mathbb{R}}\{\mathbb{P}(\Phi_T \leq q) \geq 1 - \alpha\}$. □

Finally, as announced in Section 2.3.4, we formally show that the union of minimal intervals $U_{ij} = \cup_{\mathcal{I}\in\mathcal{F}^{\min}_{\text{reject}}(i,j)}\mathcal{I}$ is closely related to the set of time points $U^*_{ij} = \{u \in [0,1] : \lambda_i(u) \neq \lambda_j(u)\}$ where λ_i and λ_j differ. We consider the following scenario:

(a) We let (i, j) be a fixed pair of countries whose trend functions λ_i and λ_j do not depend on T. Hence, unlike in Proposition 2.A.4, we consider fixed rather than local alternatives.

(b) We suppose that the family of intervals $\mathcal{F}$ has the structure

$$\begin{aligned}\mathcal{F} = \big\{[u, u+h] \subseteq [0,1] : u = kh_{\min} \text{ for some } k = 0, \ldots, h_{\min}^{-1} - 1 \text{ and} \\ h = 2^\ell h_{\min} \text{ for some } \ell = 1, \ldots, L\big\},\end{aligned} \tag{2.B.12}$$

where $h_{\min}$ is chosen such that $h_{\min}^{-1}$ is a natural number. The family $\mathcal{F}$ consists of the intervals $[0, h_{\min}]$, $[h_{\min}, 2h_{\min}]$, $[2h_{\min}, 3h_{\min}]$, ... and unions thereof. Note that we could allow for more general families $\mathcal{F}$. We nevertheless work with the specific structure (2.B.12) to keep the technical arguments as simple as possible.

We can prove the following lemma in this scenario.

Lemma 2.B.9. *Let the conditions of Theorem 2.A.1 be satisfied and let $\alpha \in (0,1)$ be given. Then*

$$\mathbb{P}\big(\Delta(U_{ij}, U^*_{ij}) \leq C\rho_T\big) \geq 1 - \alpha + o(1),$$

where $\Delta(U_{ij}, U^*_{ij}) = \mathcal{L}\{(U_{ij} \setminus U^*_{ij}) \cup (U^*_{ij} \setminus U_{ij})\}$ is the Lebesgue measure $\mathcal{L}$ of the symmetric difference between the two sets U_{ij} and U^*_{ij} and $\{\rho_T\}$ is a null sequence of positive numbers.

Since ρ_T converges to 0 as $T \to \infty$, Lemma 2.B.9 essentially says that the difference between U_{ij} and U_{ij}^* is small ($\leq C\rho_T = o(1)$) with high probability ($\geq 1 - \alpha + o(1)$). In this sense, U_{ij} can be regarded as an approximation of U_{ij}^*.

Proof of Lemma 2.B.9. Let

$$U_{ij}^{*,\text{int}} = \Big\{ u \in U_{ij}^* : |\lambda_i(u) - \lambda_j(u)| \geq \kappa_T \sqrt{\log T/(Th_{\min})} \Big\}$$

and define the two collections of intervals

$$\begin{aligned}
\mathcal{C}^< &= \big\{ \mathcal{I} \in \mathcal{F} : \ell(\mathcal{I}) = h_{\min} \text{ and } \mathcal{I} \subseteq U_{ij}^{*,\text{int}} \big\} \\
\mathcal{C}^> &= \big\{ \mathcal{I} \in \mathcal{F} : \ell(\mathcal{I}) = \bar{h} \text{ and } \mathcal{J} \subseteq \mathcal{I} \text{ for some } \mathcal{J} \in \mathcal{F}_{\text{reject}}^{\min}(i,j) \big\},
\end{aligned}$$

where $\ell(\mathcal{I})$ denotes the length of the interval $\mathcal{I} \subseteq [0,1]$. In what follows, we examine the two sets of time points

$$U_{ij}^< = \bigcup_{\mathcal{I} \in \mathcal{C}^<} \mathcal{I} \quad \text{and} \quad U_{ij}^> = \bigcup_{\mathcal{I} \in \mathcal{C}^>} \mathcal{I}.$$

(i) By Proposition 2.A.4, it holds that $\mathbb{P}(\mathcal{C}^< \subseteq \mathcal{F}_{\text{reject}}^{\min}(i,j)) \to 1$, which implies that $U_{ij}^< \subseteq U_{ij}$ with probability tending to 1. Moreover, straightforward arguments yield that $\Delta(U_{ij}^<, U_{ij}^*) \leq C\xi_T$ with some null sequence $\{\xi_T\}$ and some sufficiently large constant C.

(ii) By construction, it holds that $U_{ij} \subseteq U_{ij}^>$. Moreover, $\mathcal{L}(U_{ij}^> \setminus U_{ij}^*) \leq C\bar{h}$ with probability at least $1 - \alpha + o(1)$, since $\mathcal{F}_{\text{reject}}^{\min}(i,j) \cap \{\mathcal{I} \in \mathcal{F} : \mathcal{I} \cap U_{ij}^* = \emptyset\} = \emptyset$ with probability at least $1 - \alpha + o(1)$ by Theorem 2.A.1. In addition, $\mathcal{L}(U_{ij}^* \setminus U_{ij}^>) \leq \mathcal{L}(U_{ij}^* \setminus U_{ij}^<) \leq \Delta(U_{ij}^<, U_{ij}^*) \leq C\xi_T$ with probability tending to 1 by (i). As a consequence, $\Delta(U_{ij}^>, U_{ij}^*) \leq C \max\{\xi_T, \bar{h}\}$ with probability at least $1 - \alpha + o(1)$.

To summarize, we have found that

$$\mathbb{P}\big(U_{ij}^< \subseteq U_{ij} \subseteq U_{ij}^>\big) = 1 - o(1)$$

and

$$\mathbb{P}\Big(\max\big\{\Delta(U_{ij}^<, U_{ij}^*), \Delta(U_{ij}^>, U_{ij}^*)\big\} \leq C\rho_T \Big) \geq 1 - \alpha + o(1),$$

where we set $\rho_T = \max\{\xi_T, \bar{h}\} = o(1)$. Taken together, these two statements imply that

$$\mathbb{P}\big(\Delta(U_{ij}, U_{ij}^*) \leq C\rho_T\big) \geq 1 - \alpha + o(1). \qquad \square$$

2.C Robustness checks for Section 2.4.1

In this section, we supplement the simulation experiments of Section 2.4.1 by some robustness checks. Specifically, we repeat the experiments with different values of the overdispersion parameter σ. The larger we choose σ, the more noise we put on top of the time trend,

that is, on top of the underlying signal. Hence, by varying σ, we can assess how sensitive our test is to changes in the noise-to-signal ratio. We first repeat the size simulations for $\sigma = 10$ and $\sigma = 20$. The results are presented in Tables 2.C.1 and 2.C.4, respectively. As can be seen, the empirical size numbers are very similar to those for $\sigma = 15$ in Table 2.1. We next rerun the power simulations for $\sigma = 10$ and $\sigma = 20$, where we consider the two Scenarios A and B as in Section 2.4.1. The results can be found in Tables 2.C.2, 2.C.3, 2.C.5 and 2.C.6. They show that the test is much more powerful for $\sigma = 10$ than for $\sigma = 20$. This is what one would expect, since a higher value of σ corresponds to a higher noise-to-signal ratio. In particular, the higher σ, the more noisy the data, and thus the more difficult it is to identify differences between the trend curves. Nevertheless, even in the very noisy case with $\sigma = 20$, our test has quite some power, which tends to increase swiftly as T gets larger.

Table 2.C.1: Empirical size of the test for $\sigma = 10$.

	$n = 5$			$n = 10$			$n = 50$		
	significance level α			significance level α			significance level α		
	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
$T = 100$	0.009	0.043	0.085	0.008	0.039	0.075	0.005	0.023	0.055
$T = 250$	0.011	0.047	0.095	0.010	0.050	0.094	0.009	0.039	0.079
$T = 500$	0.009	0.052	0.101	0.013	0.049	0.101	0.010	0.039	0.084

Table 2.C.2: Power of the test in Scenario A for $\sigma = 10$.

	$n = 5$			$n = 10$			$n = 50$		
	significance level α			significance level α			significance level α		
	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
$T = 100$	0.836	0.915	0.911	0.833	0.903	0.898	0.777	0.874	0.882
$T = 250$	0.986	0.971	0.938	0.984	0.956	0.918	0.980	0.961	0.924
$T = 500$	0.996	0.975	0.946	0.994	0.965	0.927	0.992	0.963	0.918

Table 2.C.3: Power of the test in Scenario B for $\sigma = 10$.

	$n = 5$			$n = 10$			$n = 50$		
	significance level α			significance level α			significance level α		
	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
$T = 100$	0.991	0.973	0.946	0.994	0.970	0.935	0.994	0.971	0.940
$T = 250$	0.993	0.969	0.941	0.993	0.959	0.919	0.991	0.960	0.925
$T = 500$	0.996	0.976	0.948	0.993	0.966	0.928	0.993	0.962	0.917

Table 2.C.4: Empirical size of the test for $\sigma = 20$.

	$n = 5$			$n = 10$			$n = 50$		
	significance level α			significance level α			significance level α		
	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
$T = 100$	0.011	0.050	0.094	0.010	0.047	0.092	0.009	0.034	0.070
$T = 250$	0.009	0.047	0.088	0.008	0.044	0.085	0.006	0.032	0.062
$T = 500$	0.008	0.038	0.081	0.006	0.039	0.079	0.006	0.025	0.060

Table 2.C.5: Power of the test in Scenario A for $\sigma = 20$.

	$n = 5$			$n = 10$			$n = 50$		
	significance level α			significance level α			significance level α		
	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
$T = 100$	0.144	0.275	0.352	0.115	0.231	0.304	0.048	0.120	0.163
$T = 250$	0.244	0.434	0.538	0.204	0.403	0.486	0.133	0.247	0.305
$T = 500$	0.296	0.563	0.662	0.273	0.511	0.603	0.175	0.338	0.433

Table 2.C.6: Power of the test in Scenario B for $\sigma = 20$.

	$n = 5$			$n = 10$			$n = 50$		
	significance level α			significance level α			significance level α		
	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
$T = 100$	0.438	0.636	0.704	0.404	0.598	0.669	0.277	0.449	0.526
$T = 250$	0.864	0.934	0.927	0.850	0.923	0.915	0.811	0.891	0.898
$T = 500$	0.960	0.968	0.949	0.961	0.964	0.935	0.945	0.961	0.941

2.D Additional graphs for Section 2.4.2

Here, we provide the pairwise comparisons between Italy, France, Spain and the UK that were omitted in Section 2.4.2. The plots have the same format as Figures 2.3–2.6.

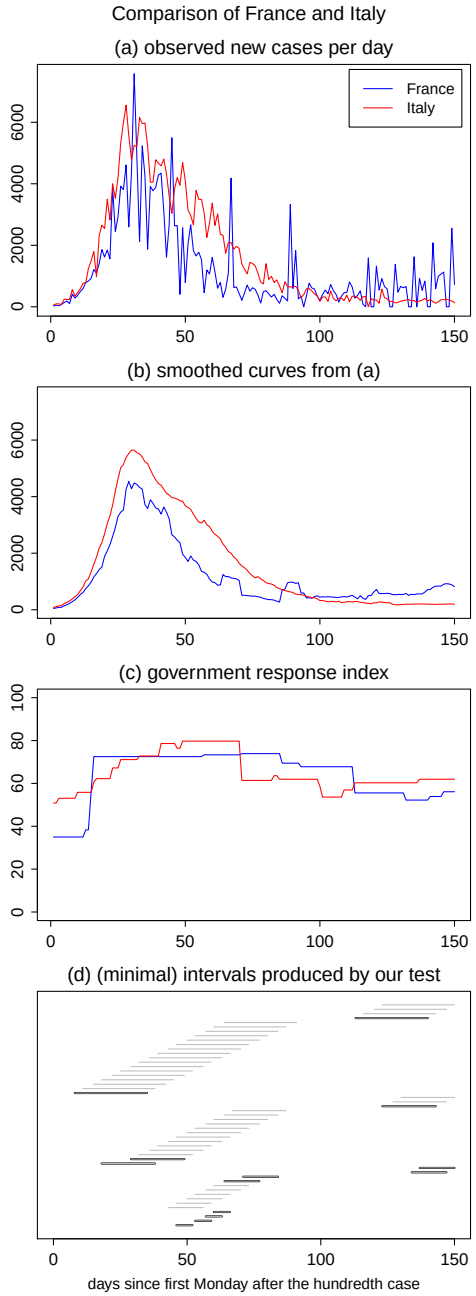


Figure 2.D.1: Test results for the comparison of France and Italy.

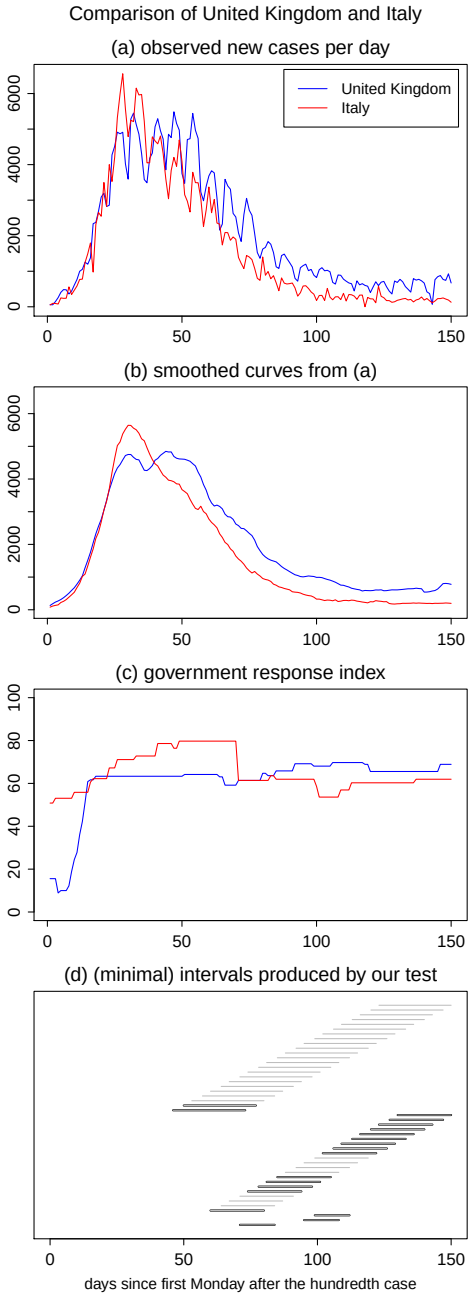


Figure 2.D.2: Test results for the comparison of the UK and Italy.

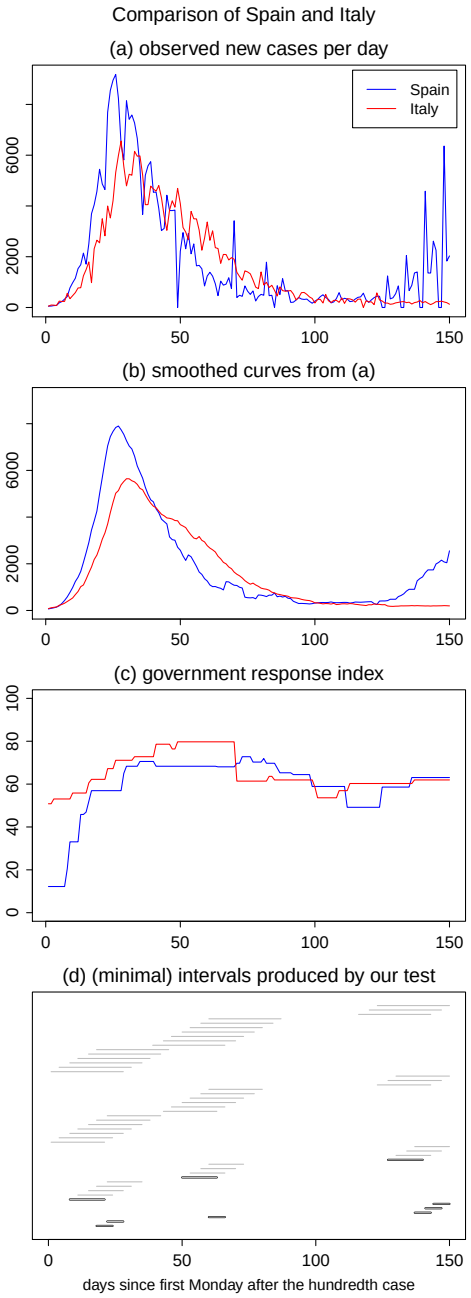


Figure 2.D.3: Test results for the comparison of Spain and Italy.

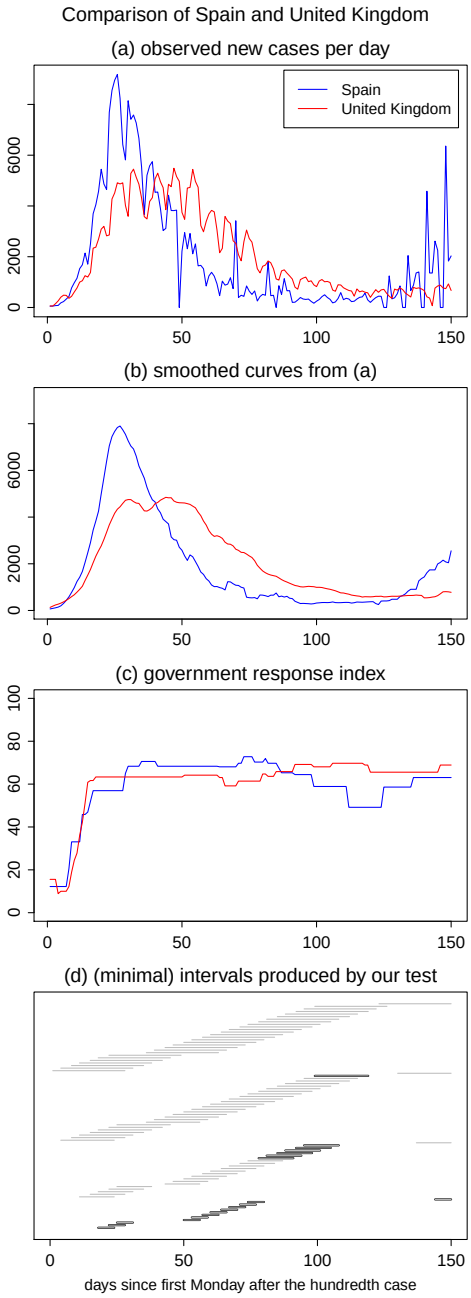


Figure 2.D.4: Test results for the comparison of Spain and the UK.

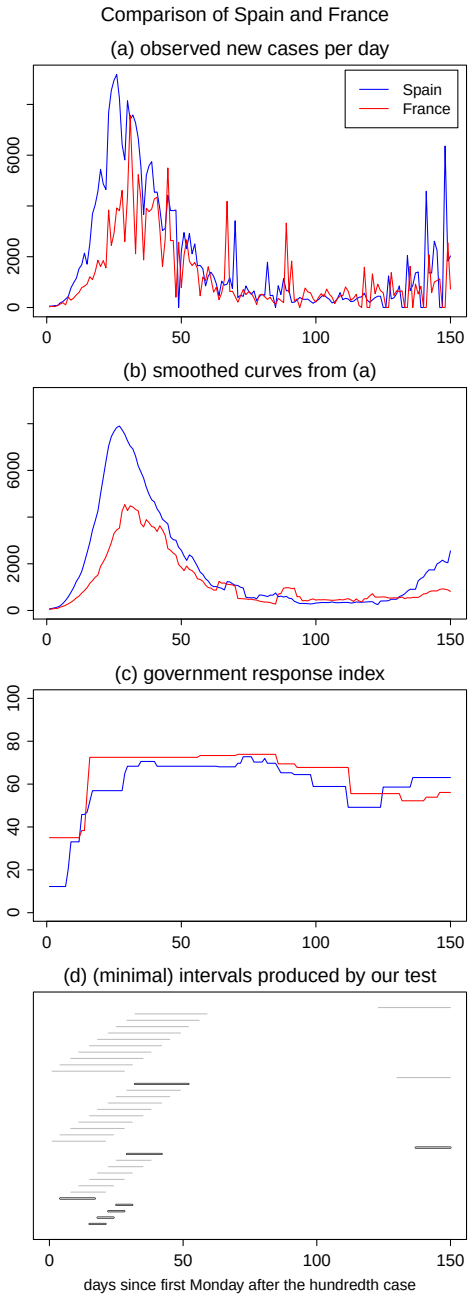


Figure 2.D.5: Test results for the comparison of Spain and France.

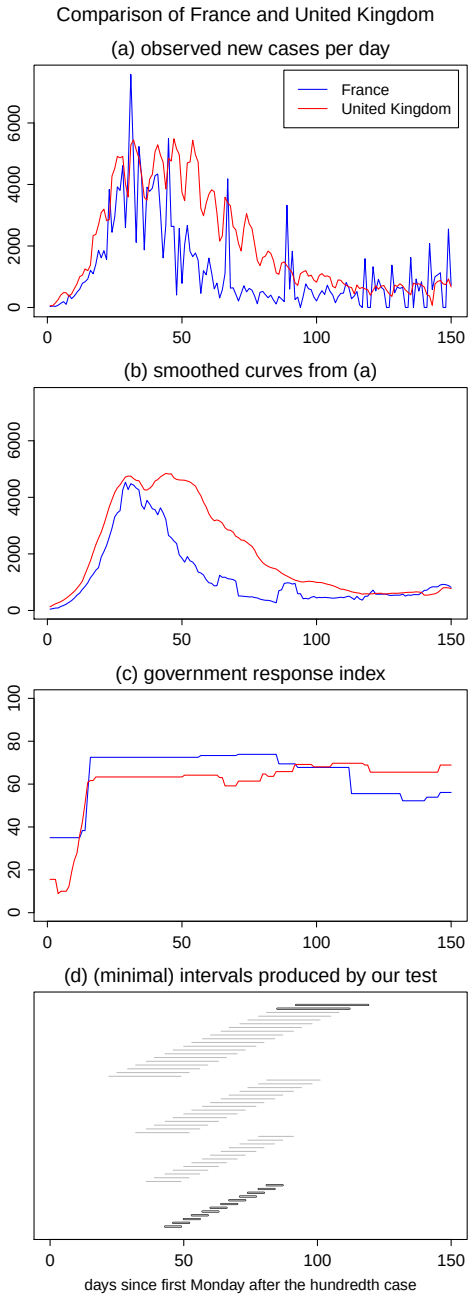


Figure 2.D.6: Test results for the comparison of France and the UK.

2.E Robustness checks for Section 2.4.2

To check the robustness of the empirical results from Section 2.4.2, we compare the time trends of the same five European countries (Germany, Italy, France, Spain and the UK) over a longer time span. Specifically, we take $T = 200$ as opposed to $T = 150$ in Section 2.4.2. The analysis is performed in the same way as in Section 2.4.2. The results are displayed in Figures 2.E.1–2.E.10, which have the same format as Figures 2.3–2.6.

As is clearly visible in the figures, the data for Spain have extremely high volatility towards the end of the observation period between days 150 and 200. At least partly, this is presumably due to repeated delays in reporting. Such delays explain why the number of daily new cases is 0 on multiple occasions followed by a sharp increase of the numbers on the next day. This spuriously high day-to-day volatility of the data heavily influences the precision of the estimator $\hat{\sigma}$. Specifically, we can expect $\hat{\sigma}$ to strongly overestimate the true underlying σ. Indeed, we get $\hat{\sigma} = 29.7$ when estimating σ with $T = 200$ compared to $\hat{\sigma} = 14.82$ with $T = 150$. Higher values of $\hat{\sigma}$ result in lower values of the test statistics, which leads to more conservative results of the test procedure. As a consequence, Figures 2.E.1–2.E.10 differ drastically from Figures 2.3–2.6 and 2.D.1–2.D.6.

Since the Spanish data between days 150 and 200 are most probably inaccurate and can thus not be taken at face value[3], we exclude Spain from our analysis and repeat the multiscale test with $T = 200$ for only four European countries: Germany, Italy, France and the UK. The results are presented in Figures 2.E.11–2.E.16. As can be seen, the test is much less conservative and the results are much more consistent with those in Section 2.4.2. Specifically, for each pair of countries (i, j), the set $\mathcal{F}_{\text{reject}}(i, j)$ constructed for the longer time period $T = 200$ contains very similar intervals as the set $\mathcal{F}_{\text{reject}}(i, j)$ constructed for the shorter time period $T = 150$. Hence, the test finds devations from the null hypothesis up to day $T = 150$ on essentially the same intervals, no matter whether the overall time series length is $T = 150$ or $T = 200$. This shows that the test produces robust results that are barely influenced by the overall length of the time series.

[3]Note that the high day-to-day volatility is present not only in the data set of the European Center for Disease Prevention and Control, which underlies our analysis, but also in the data sets provided by the World Health Organization and by John Hopkins University.

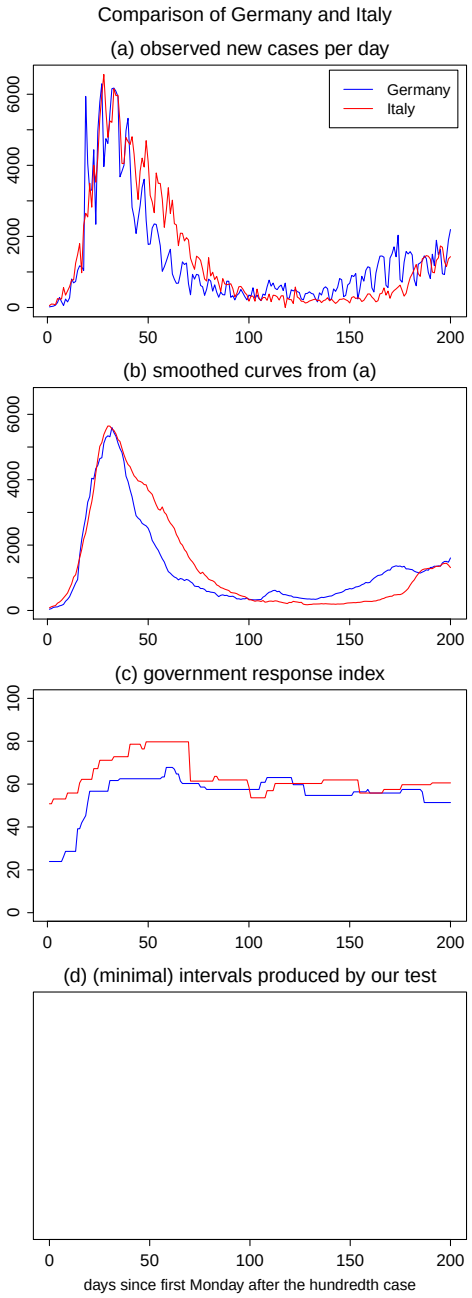


Figure 2.E.1: Test results for the comparison of Germany and Italy ($T = 200$).

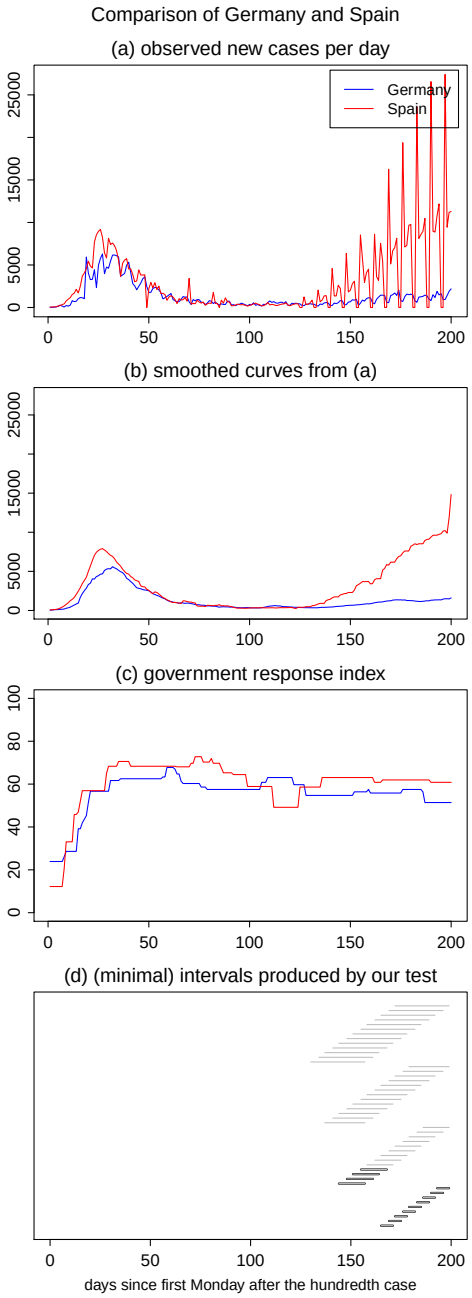


Figure 2.E.2: Test results for the comparison of Germany and Spain ($T = 200$).

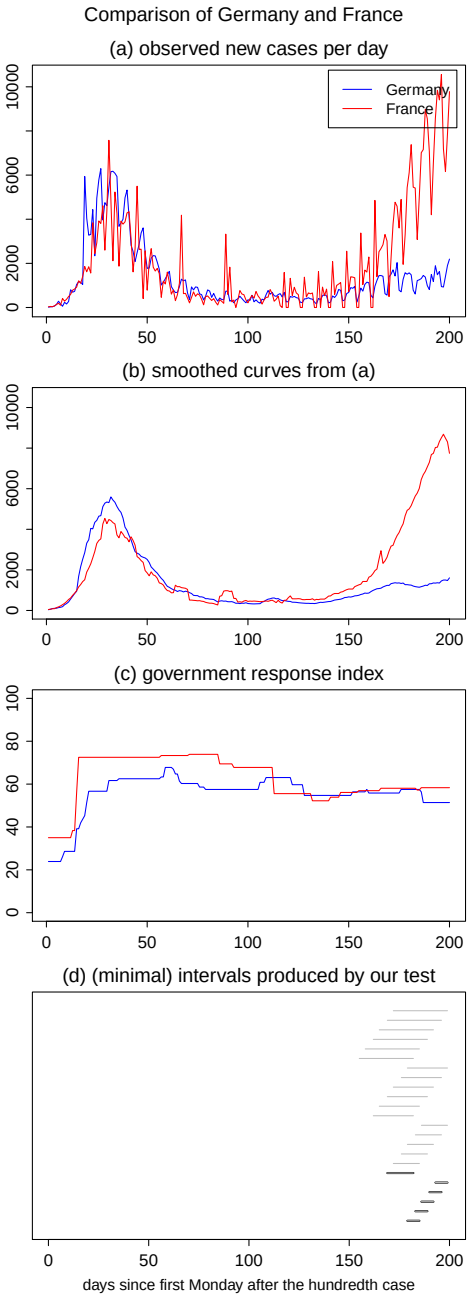


Figure 2.E.3: Test results for the comparison of Germany and France ($T = 200$).

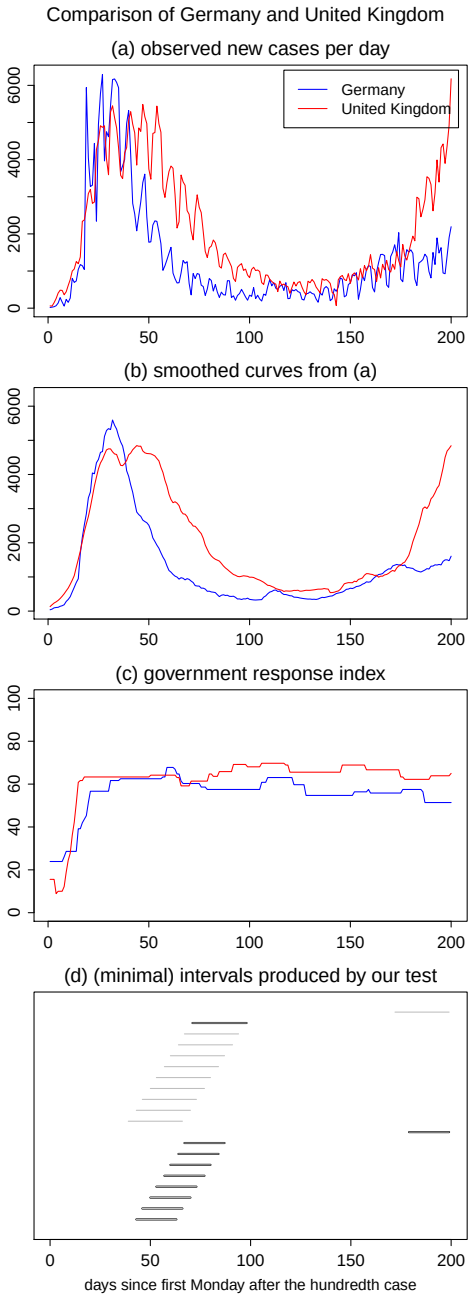


Figure 2.E.4: Test results for the comparison of Germany and the UK ($T = 200$).

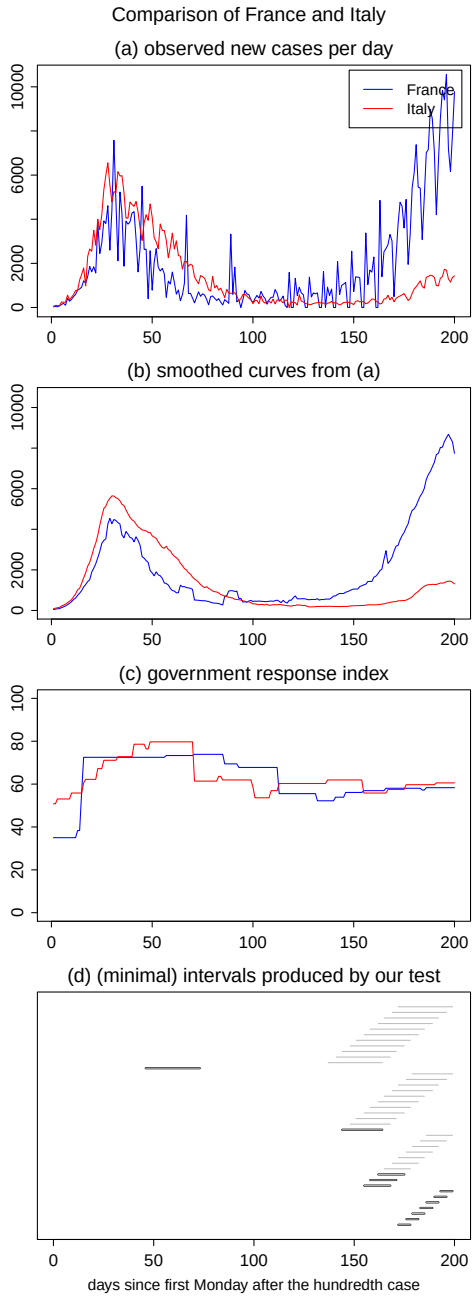


Figure 2.E.5: Test results for the comparison of France and Italy ($T = 200$).

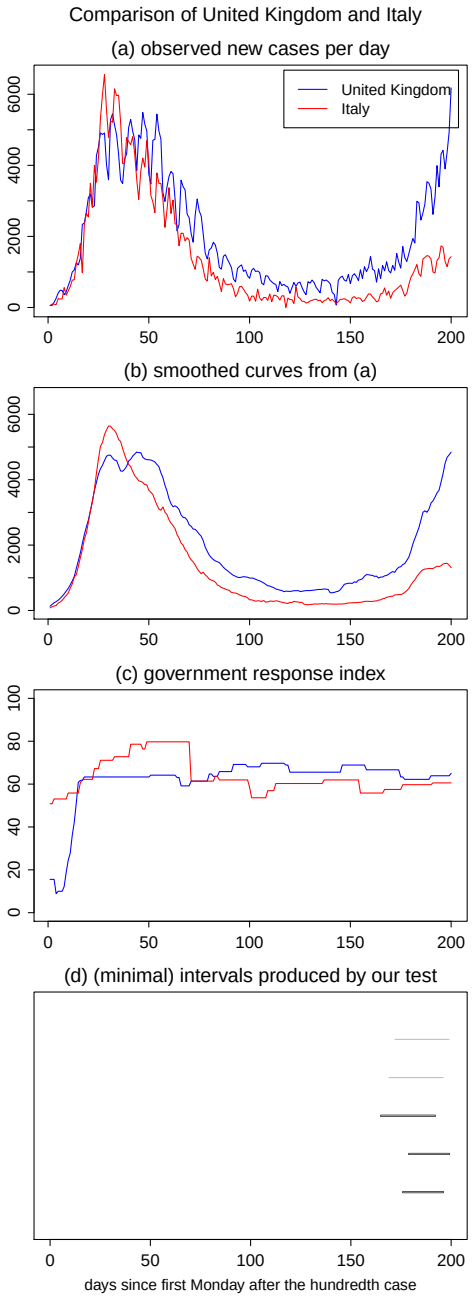


Figure 2.E.6: Test results for the comparison of the UK and Italy ($T = 200$).

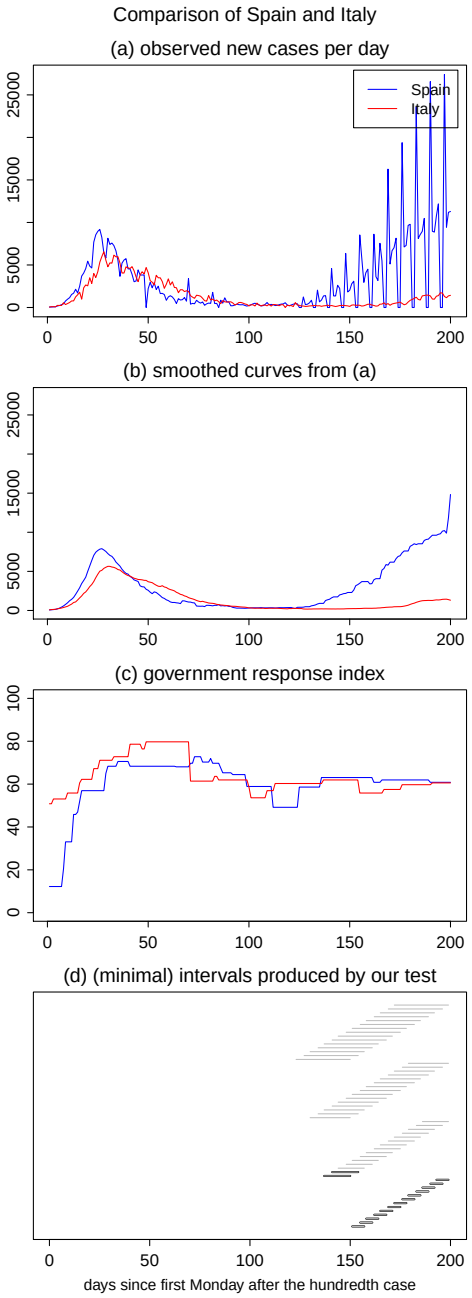


Figure 2.E.7: Test results for the comparison of Spain and Italy ($T = 200$).

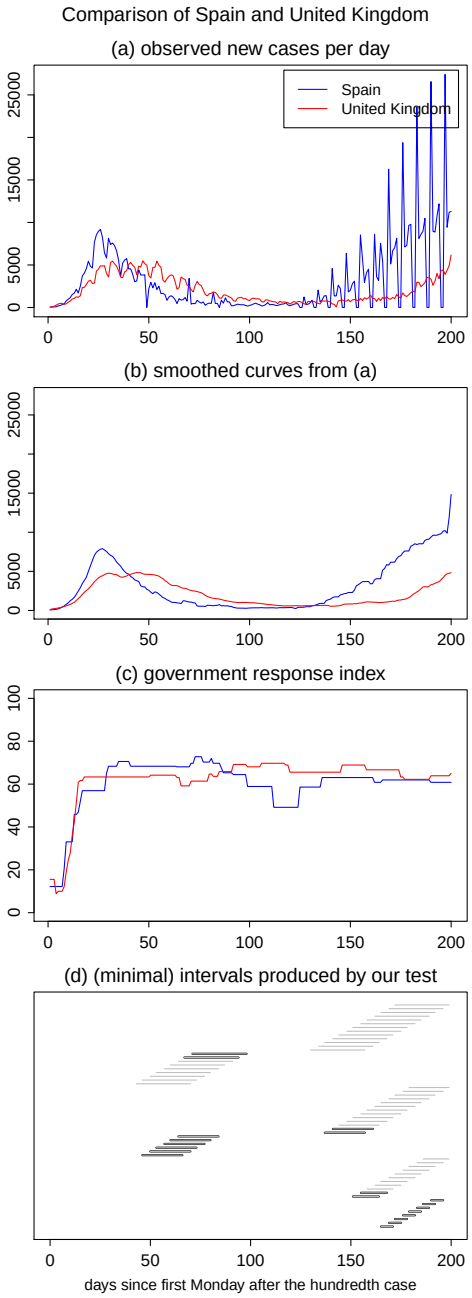


Figure 2.E.8: Test results for the comparison of Spain and the UK ($T = 200$).

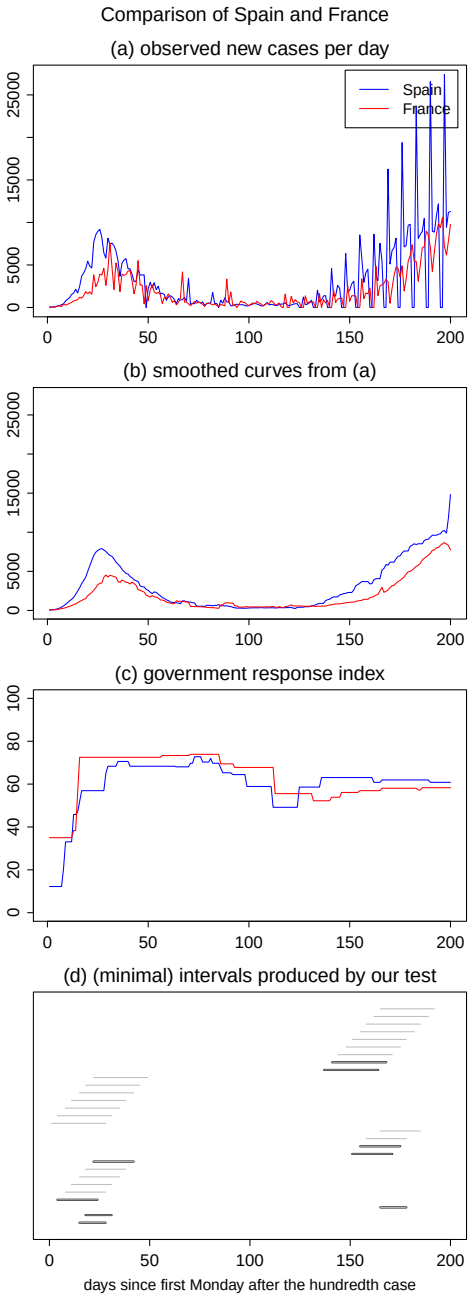


Figure 2.E.9: Test results for the comparison of Spain and France ($T = 200$).

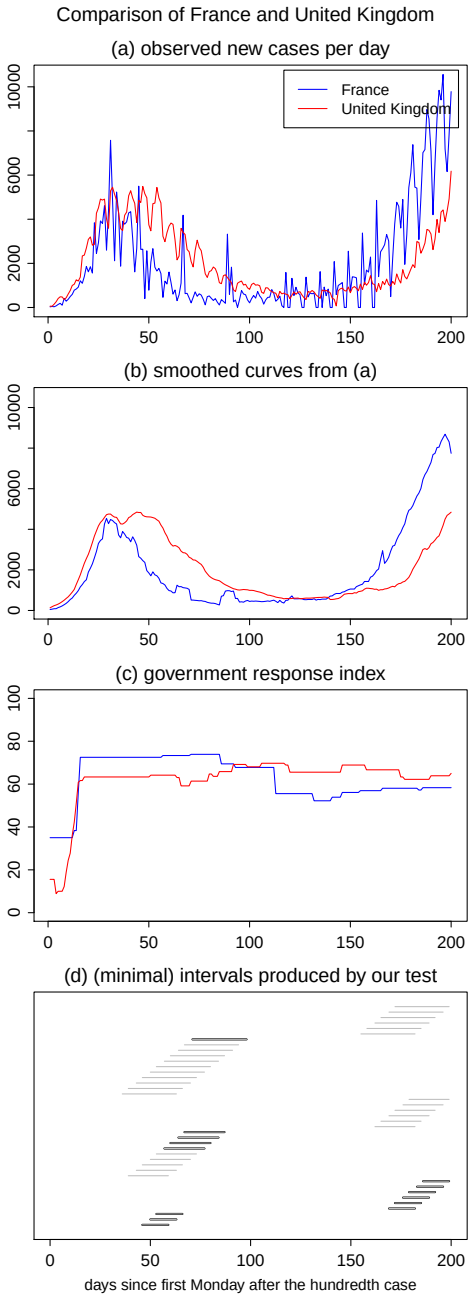


Figure 2.E.10: Test results for the comparison of France and the UK ($T = 200$).

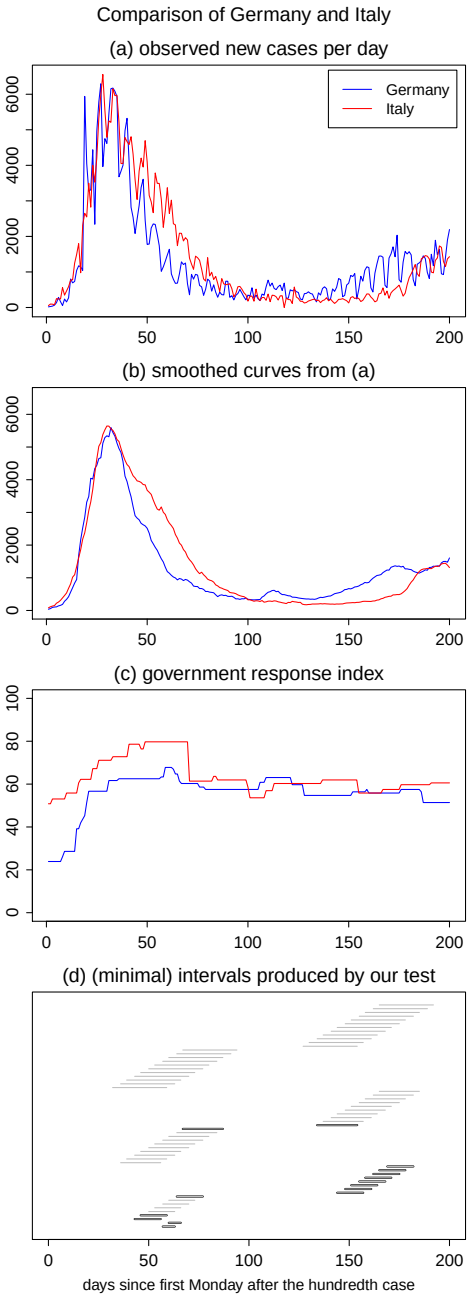


Figure 2.E.11: Test results for the comparison of Germany and Italy excluding Spain ($T = 200$).

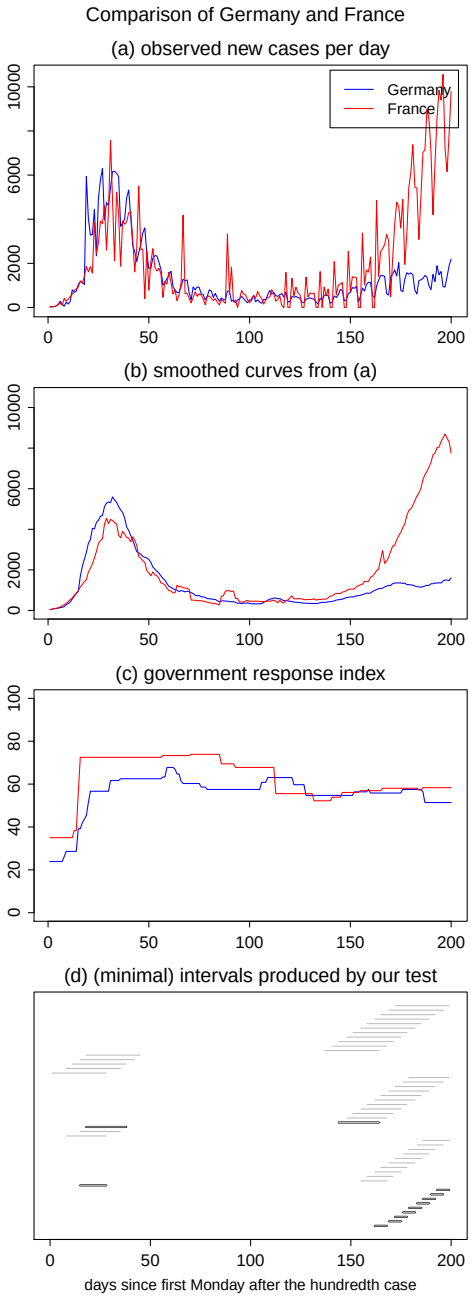


Figure 2.E.12: Test results for the comparison of Germany and France excluding Spain ($T = 200$).

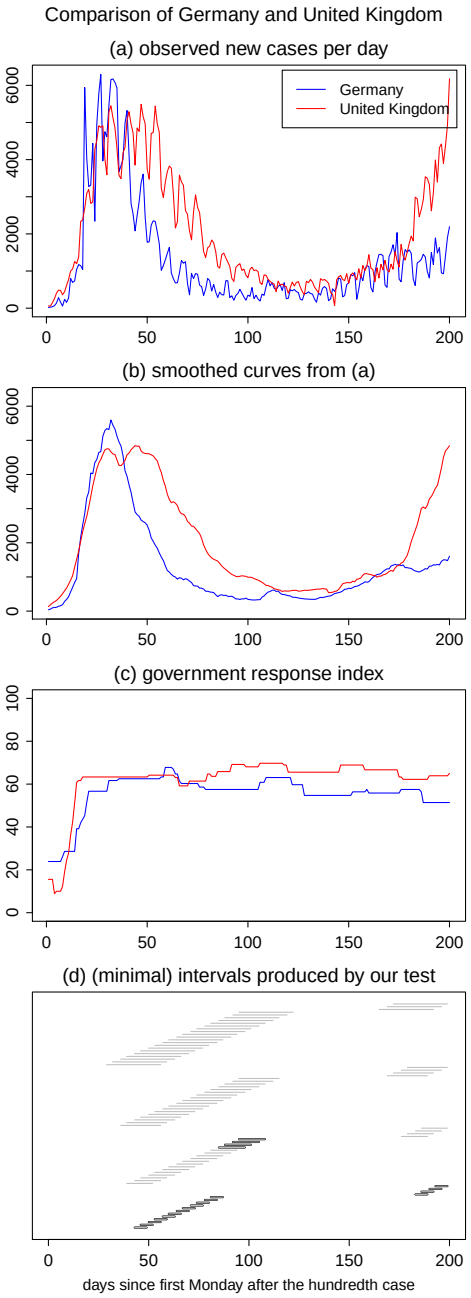


Figure 2.E.13: Test results for the comparison of Germany and the UK excluding Spain ($T = 200$).

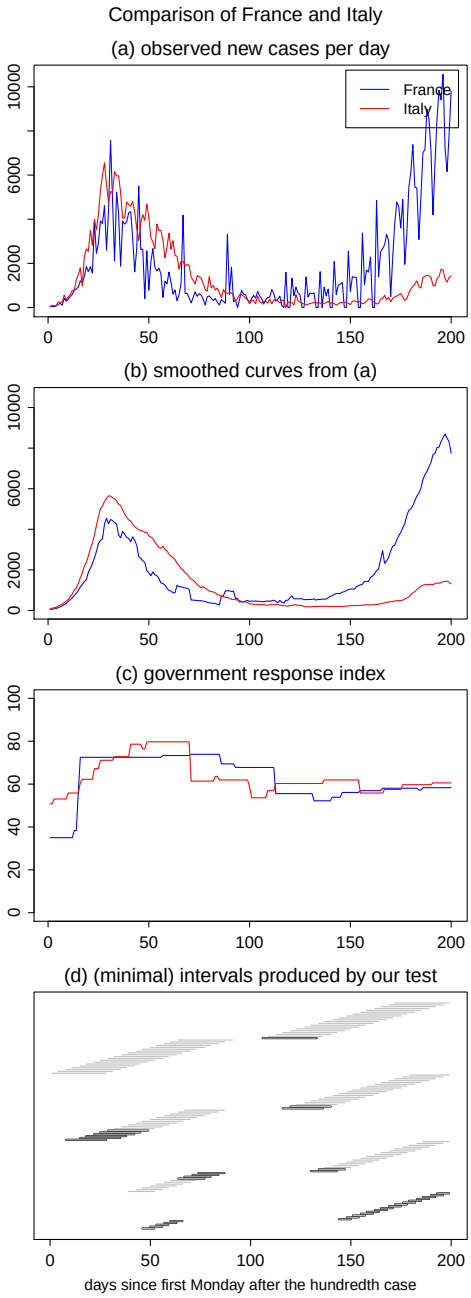


Figure 2.E.14: Test results for the comparison of France and Italy excluding Spain ($T = 200$).

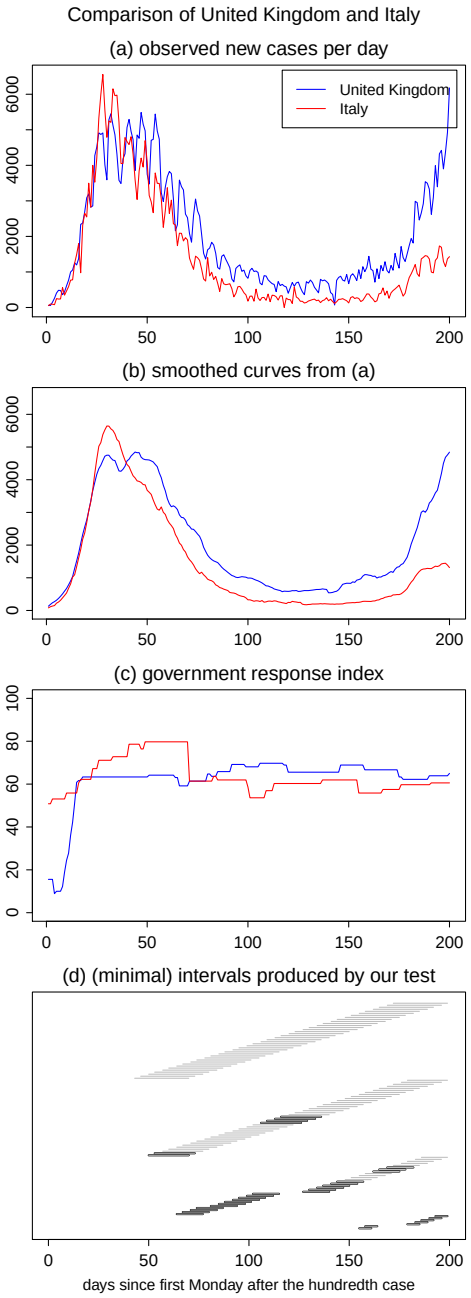


Figure 2.E.15: Test results for the comparison of the UK and Italy excluding Spain ($T = 200$).

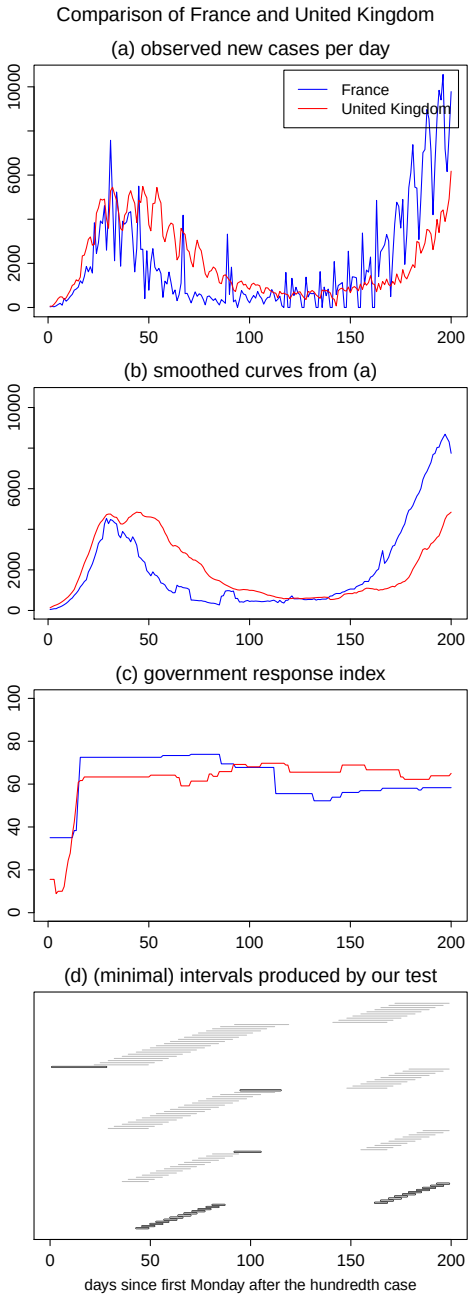


Figure 2.E.16: Test results for the comparison of France and the UK excluding Spain ($T = 200$).

Chapter 3

Multiscale Testing for Equality of Nonparametric Trend Curves

Joint with Michael Vogt

3.1 Introduction

Comparison of several regression curves is a classical topic in econometrics and statistics. In many cases of practical interest, the functional forms of the objective regression curves are unknown, hence, the parametric approach is not applicable. In this paper, we propose a novel approach that addresses this particular problem in a nonparametric context. Specifically, we present a new testing procedure for detecting differences between the nonparametric trends curves.

In what follows, we consider a general panel framework with heterogeneous trends. Suppose we observe a panel of n time series $\mathcal{Z}_i = \{(Y_{it}, \mathbf{X}_{it}) : 1 \leq t \leq T\}$ for $1 \leq i \leq n$, where Y_{it} are real-valued random variables and $\mathbf{X}_{it} = (X_{it,1}, \ldots, X_{it,d})^\top$ are d-dimensional random vectors. Each time series $\mathcal{Z}_i$ is modelled by the equation

$$Y_{it} = m_i\Big(\frac{t}{T}\Big) + \boldsymbol{\beta}_i^\top \mathbf{X}_{it} + \alpha_i + \varepsilon_{it} \tag{3.1.1}$$

for $1 \leq t \leq T$, where $\boldsymbol{\beta}_i$ is a $d \times 1$ vector of unknown parameters, $\mathbf{X}_{it}$ is a $d \times 1$ vector of individual covariates or controls, m_i is an unknown nonparametric (deterministic) trend function defined on $[0, 1]$, α_i are so-called fixed effect error terms and $\mathcal{E}_i = \{\varepsilon_{it} : 1 \leq t \leq T\}$ is a zero-mean stationary error process.

An important question in many applications is whether the observed time series have the common trend. In other words, the researchers would like to know if m_i are the same for all i. Moreover, when there is evidence that this is not the case, one of the major related statistical problems is to determine which of the trends are different and whether we can group the time series with the similar trends together. 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 this paper, we introduce new statistical methods to approach these questions. In particular, we develop a test of the hypothesis that all time trends in

model (3.1.1) are the same. In this setting, the null hypothesis is formulated as

$$H_0 : m_1 = m_2 = \ldots = m_n, \tag{3.1.2}$$

whereas the alternative hypothesis is

$$H_1 : \text{ there exists } x \in [0,1] \text{ such that } m_i(x) \neq m_j(x) \text{ for some } 1 \leq i < j \leq n.$$

The method that we propose does not only allow to test whether the null hypothesis is violated. It also allows to detect, with a given statistical confidence, which time trends are different and in which time regions they differ. More specifically, for any given interval $[u-h, u+h] \subseteq [0,1]$, consider the hypothesis

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

Here, we can regard h as a bandwidth, a common tuning parameter in nonparametric estimation. The given interval $\mathcal{I}_{(u,h)} = [u-h, u+h] \subseteq [0,1]$ is then fully characterized by u, its center (a location parameter), and h, the bandwidth. In order to determine the regions where the time trends are different, we consider a broad range of pairs (u,h) with the property that they fully cover the unit interval $[0,1]$. Formally, let $\mathcal{G} := \{(u,h) : \mathcal{I}_{(u,h)} = [u-h, u+h] \subseteq [0,1]\}$ be a grid of location-bandwidth points such that

$$\bigcup_{(u,h)\in\mathcal{G}} \mathcal{I}_{(u,h)} = [0,1].$$

We then reformulate our null hypothesis (3.1.2) as

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

$H_0^{[i,j]}(u,h)$ can thus be viewed as a local null hypothesis that characterizes the behavior of two trend function only locally, whereas H_0 specified in (3.1.2) is the global null hypothesis that is concerned with the comparison of all of the trends on the whole unit interval.

In this paper, we introduce a method that allows us to test the hypotheses $H_0^{[i,j]}(u,h)$ simultaneously for all pairs (i,j) and for all intervals $\mathcal{I}_{(u,h)}$ under consideration. Specifically, we develop a multiscale test for the model (3.1.1). The underlying idea of any multiscale test is to consider a number of test statistics (each corresponding to a different set of values of some tuning parameters) all at once rather than to perform a separate test for each single test statistics. In our case, this means testing many local null hypotheses $H_0^{[i,j]}(u,h)$ simultaneously which leads to a well-known multiple testing problem. Our method accounts for this problem by using appropriate critical values that depend on the scale of the problem, i.e. on the number of hypotheses tested simultaneously and

the relationship between them. In the paper, we show that the suggested procedure for obtaining critical values leads to good theoretical properties of the proposed test: it has the correct (asymptotic) level and an (asymptotic) power of one against a certain class of local alternatives.

Trend comparison is a common statistical problem that arises in various contexts. For example, in economics the researchers compare trends in real gross domestic product across several countries (Grier and Tullock, 1989), in yield over time of US Treasury bills at different maturities (Park et al., 2009b), or the evolution of long-term interest rates in a number of countries (Christiansen and Pigott, 1997). In finance, comparison and subsequent classification of the trends of market fragmentation can be used to assess the market quality in the European stock market (Vogt and Linton, 2017, 2020). In climatology, the temperature time series in different geographical areas are investigated in the context of the regional and global warming trends (Karoly and Wu, 2005). Finally, in industry, mobile phone providers are interested in finding the differences between the cell phone download activity in various locations (Degras et al., 2012).

In the statistical literature, the problem of testing whether the observed time series all have the same trend has been widely studied, and tests for equality of trends or regression curves have been developed in Härdle and Marron (1990), Hall and Hart (1990), Delgado (1993) and Degras et al. (2012) among many others. Versions of model (3.1.1) with a parametric trend are considered in Vogelsang and Franses (2005), Sun (2011) and Xu (2012) among others. In the nonparametric context, Li et al. (2010), Atak et al. (2011), Robinson (2012) and Chen et al. (2012) studied panel models under the assumption that the observed time series have a common time trend. However, in many applications the restriction of including a common time trend in the model is questionable at best. For instance, when we observe a large number of time series it is reasonable to expect that at least some of the trends are different from the others. Consequently, it often makes sense to relax the assumption of a common trend, which leads to more flexible panel settings with heterogeneous trends. Such models have been studied, for example, in Degras et al. (2012), Zhang et al. (2012) and Hidalgo and Lee (2014). Degras et al. (2012) consider the problem of testing H_0 in a model that is a special case of (3.1.1) and does not include additional regressors. Chen and Wu (2018) develop theory for a very similar model framework but under more general conditions on the error terms. Zhang et al. (2012) investigate the problem of testing the hypothesis H_0 in a slightly restricted version of model (3.1.1), where $\boldsymbol{\beta}_i = \boldsymbol{\beta}$ for all i. All of these tests have an important drawback: they involve classical nonparametric estimation of the trend functions that depends on one or several bandwidth parameters, which imposes a certain limit on the applicability of such tests since in most cases it is far from clear how to choose bandwidth parameters in an appropriate way. Contrary to the aforementioned methods, our multiscale testing procedure allows us to consider a large collection of bandwidths simultaneously avoiding the problem of choosing only one bandwidth altogether.

Recently, Khismatullina and Vogt (2021) proposed a new inference method that al-

lows researchers to detect differences between epidemic time trends in the context of the COVID-19 pandemic. In their paper, the authors present a statistically rigorous procedure that, similarly to ours, not only allows to compare trends across different countries, but to pinpoint the time intervals where the differences occur as well. Moreover, they also circumvent the need to pick a bandwidth parameter by using a multiscale testing approach. However, the model in Khismatullina and Vogt (2021) is only a special case of the model (3.1.1) which includes neither the covariates $\mathbf{X}_{it}$, nor the fixed effects α_i. Furthermore, the authors place major restriction on the error terms: in their model, ε_{it} are independent across t. In contrast, our model (3.1.1) can be regarded as a generalized version of theirs that allows for a wider range of economic and financial applications.

To sum up, the main theoretical contribution of the current paper is the multiscale testing method that allows to make simultaneous confidence statements about which of the time trends are distinct and the regions where they differ. We believe that currently there are no equivalent statistical methods. Even though tests for equality of the trends have been developed already for a while, most existing procedures allow only to test whether the trend curves are all the same or not, but they almost never allow to infer which curves are different and where. To the best of our knowledge, the only two exceptions are Khismatullina and Vogt (2021), whose contribution is briefly discussed above, and Park et al. (2009b) who developed SiZer methods for the comparison of nonparametric trend curves in a significantly simplified version of the model (3.1.1). In addition to restricted model, Park et al. (2009b) derive theoretical results for their analysis only for the special case of observing only two time series, whereas in other cases, the algorithm is provided without detailed proof.

The structure of the paper is as follows. Section 3.2 introduces the model setting and the necessary technical assumptions that are required for the theory. The multiscale test is developed step by step in Section 3.3. The main theoretical results are presented in Section 3.4. Section 3.5 deals with estimating the unknown parameters necessary for construction of the test statistics. To keep the discussion as clear as possible, we include in the main text of the paper only the essential parts of the theoretical arguments, and the technical details and extended proofs are deferred to the Appendix. Section 3.6 concludes.

3.2 The model

Throughout the paper, we adopt the following notation. For a real-valued vector $\mathbf{v} = (v_1, \dots, v_m) \in \mathbb{R}^m$, we write $|\mathbf{v}| = \big(\sum_{i=1}^m v_i^2\big)^{1/2}$ and $|\mathbf{v}|_q = \big(\sum_{i=1}^m v_i^q\big)^{1/q}$ respectively. For a random vector $\mathbf{V}$, we define it's $\mathcal{L}^q, q > 1$ norm as $||\mathbf{V}||_q = \big(\mathbb{E}|\mathbf{V}|^q\big)^{1/q}$. For the particular case $q = 2$, we write $||\mathbf{V}|| := ||\mathbf{V}||_2$.

Following Wu (2005), we define the *physical dependence measure* for the process $\mathbf{L}(\mathcal{F}_t)$ as the following:

$$\delta_q(\mathbf{L}, t) = ||\mathbf{L}(\mathcal{F}_t) - \mathbf{L}(\mathcal{F}_t')||_q, \tag{3.2.1}$$

where $\mathcal{F}_t = (\ldots, \epsilon_{-1}, \epsilon_0, \epsilon_1, \ldots, \epsilon_{t-1}, \epsilon_t)$ and $\mathcal{F}'_t = (\ldots, \epsilon_{-1}, \epsilon'_0, \epsilon_1, \ldots, \epsilon_{t-1}, \epsilon_t)$ is a coupled process of $\mathcal{F}_t$ with ϵ'_0 being an i.i.d. copy of ϵ_0. Intuitively, $\delta_q(\mathbf{L}, t)$ measures the dependency of $\mathbf{L}(\mathcal{F}_t)$ on ϵ_0, i.e., how replacing ϵ_0 by an i.i.d. copy while keeping all other innovations in place affects the output $\mathbf{L}(\mathcal{F}_t)$.

3.2.1 Setting

As was already briefly discussed in Section 3.1, the model setting is as follows. We observe a panel of n time series $\mathcal{Z}_i = \{(Y_{it}, \mathbf{X}_{it}) : 1 \leq t \leq T\}$ of length T for $1 \leq i \leq n$. Each time series $\mathcal{Z}_i$ satisfies the model equation

$$Y_{it} = \boldsymbol{\beta}_i^\top \mathbf{X}_{it} + m_i\Big(\frac{t}{T}\Big) + \alpha_i + \varepsilon_{it} \tag{3.2.2}$$

for $1 \leq t \leq T$, where $\boldsymbol{\beta}_i$ is a $d \times 1$ vector of unknown parameters, $\mathbf{X}_{it}$ is a $d \times 1$ vector of individual covariates, m_i is an unknown nonparametric trend function defined on $[0, 1]$ with $\int_0^1 m_i(u)du = 0$ for all i, α_i is a (deterministic or random) intercept term and $\mathcal{E}_i = \{\varepsilon_{it} : 1 \leq t \leq T\}$ is a zero-mean stationary error process. As common in nonparametric regression, the trend functions m_i in model (3.2.2) depend on rescaled time t/T rather than on real time t. Using rescaled time is equivalent to restricting the domain of the functions to the unit interval which in turn allows us to apply the usual asymptotic arguments. Discussion about the application of the rescaled time in the context of nonparametric estimation can be found in Robinson (1989), Dahlhaus (1997) and Vogt and Linton (2014). The condition $\int_0^1 m_i(u)du = 0$ for all i is a necessary identification condition due the presence of α_i. Without imposing this condition, we can freely increase the functions m_i by any (positive or negative) constant c_i while simultaneously subtract the same constant from the intercept term α_i:

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

The term α_i can be regarded as an additional error component. In the econometrics literature, it is commonly called a fixed effect and is often interpreted as the term which captures unobserved characteristics of the time series $\mathcal{Z}_i$ that 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 (3.2.2), we allow the n time series $\mathcal{Z}_i$ in our panel 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. As usual in nonparametric estimation, we also assume that all the trend functions $m_i(\cdot)$ are continuously differentiable on $[0, 1]$. Technical conditions regarding the model are discussed further in this section.

Finally, throughout the paper we restrict attention to the case where the number of time series n in model (3.2.2) is fixed. Extending our theoretical results to the case where n slowly grows with the sample size T is a possible topic for further research.

3.2.2 Assumptions

Each process $\mathcal{E}_i$ is supposed to satisfy the following conditions:

(C1) For each i the variables ε_{it} allow for the representation $\varepsilon_{it} = G_i(\ldots, \eta_{it-1}, \eta_{it})$, where η_{it} are i.i.d. random variables across t and $G_i : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ is a measurable function. Denote $\mathcal{J}_{it} = (\ldots, \eta_{it-2}, \eta_{it-1}, \eta_{it})$.

(C2) For all i it holds that $\mathbb{E}[\varepsilon_{it}] = 0$ and $\|\varepsilon_{it}\|_q < \infty$ for some $q > 4$.

Assumption (C1) can be translated as the restriction on the error process $\mathcal{E}_i$ to be stationary and causal (in a sense that ε_{it} does not depend on the future innovations η_{is}, $s > t$). The class of error processes that satisfies the condition (C1) is massive, and includes linear processes, their nonlinear transformation, as well as a large variety of nonlinear processes such as Markov chain models and nonlinear autoregressive models (Wu and Wu, 2016). Assumption (C2) is a standard moment condition.

Following Wu (2005), we impose conditions on the dependence structure of the error processes $\mathcal{E}_i$ in terms of the physical dependence measure $\delta_q(G_i, t)$ defined in (3.2.1). In particular, we assume the following:

(C3) Define $\Theta_{i,t,q} = \sum_{s \geq t} \delta_q(G_i, s)$ for $t \geq 0$. For each i it holds that $\Theta_{i,t,q} = O(t^{-\tau_q}(\log t)^{-A})$, where $A > \frac{2}{3}(1/q + 1 + \tau_q)$ and $\tau_q = \{q^2 - 4 + (q-2)\sqrt{q^2 + 20q + 4}\}/8q$.

For fixed i and t, $\Theta_{i,t,q}$ measures the cumulative effect of η_0 on $(\varepsilon_{is})_{s \geq t}$ in terms of $\mathcal{L}^q$-norm. Condition (C3) assumes that the overall cumulative effect is finite and puts some restrictions on the rate of decay of $\Theta_{i,t,q}$. Assumption (C3) is fulfilled by a wide range of stationary processes $\mathcal{E}_i$. For a detailed discussion of an assumption (C3), as well as assumptions (C1)–(C2) and some examples of the error processes that satisfy these conditions, see Khismatullina and Vogt (2020).

Regarding the independent variables $\mathbf{X}_{it}$, we assume the following for each i:

(C4) The covariates $\mathbf{X}_{it}$ allow for the representation $\mathbf{X}_{it} = \mathbf{H}_i(\ldots, u_{it-1}, u_{it})$ with u_{it} being i.i.d. random variables and $\mathbf{H}_i := (H_{i1}, H_{i2}, \ldots, H_{id})^\top : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^d$ being a measurable function such that $\mathbf{H}_i(\mathcal{U}_{it})$ is well defined. We denote $\mathcal{U}_{it} = (\ldots, u_{it-1}, u_{it})$.

(C5) Let $\mathbf{N}_i$ be the $d \times d$ matrix with kl-th entry $n_{i,kl} = \mathbb{E}[H_{ik}(\mathcal{U}_{i0})H_{il}(\mathcal{U}_{i0})]$. We assume that the smallest eigenvalue of $\mathbf{N}_i$ is strictly bigger than 0.

(C6) Let $\mathbb{E}[\mathbf{H}_i(\mathcal{U}_{i0})] = \mathbf{0}$ and $||\mathbf{H}_i(\mathcal{U}_{it})||_{q'} < \infty$ for some $q' > \max\{2\theta, 4\}$, where θ will be introduced further in Assumption (C12).

(C7) $\sum_{s=0}^{\infty} \delta_{q'}(\mathbf{H}_i, s) < \infty$ for q' from Assumption (C6).

(C8) For each i it holds that $\sum_{s=t}^{\infty} \delta_{q'}(\mathbf{H}_i, s) = O(t^{-\alpha})$ for q' from Assumption (C6) and for some $\alpha > 1/2 - 1/q'$.

As with the error processes $\mathcal{E}_i$, $\mathbf{X}_i$ is guaranteed to be stationary and causal by Assumption (C4). Assumptions (C5) and (C6) are technical conditions that prevents asymptotic multicollinearity and ensures that all the necessary moments exist, respectively. Moreover, similarly to the restriction on the error processes, we also employ the definition of the physical dependence measure $\delta_q(\cdot, \cdot)$ in Assumptions (C7) - (C8), thus, making certain that the cumulative effect of the innovation u_0 on $(\mathbf{X}_{it})_{t \geq 0}$ is finite.

To be able to prove the main theorems in Section 3.3, we need additional assumptions on the relationship between the covariates and the error process.

(C9) $\mathbf{X}_{it}$ (elementwise) and ε_{is} are uncorrelated for each $t, s \in \{1, \ldots, T\}$.

(C10) Let $\zeta_{i,t} = (u_{it}, \eta_{it})^\top$. Define $\mathcal{I}_{it} = (\ldots, \zeta_{i,t-1}, \zeta_{i,t})$ and $\mathbf{U}_i(\mathcal{I}_{it}) = \mathbf{H}_i(\mathcal{U}_{it}) G_i(\mathcal{J}_{it})$. With this notation at hand, we assume that $\sum_{s=0}^{\infty} \delta_2(\mathbf{U}_i, s) < \infty$.

Assumption (C9) is a slightly relaxed independence assumption: even though we do not require the covariates $\mathbf{X}_{it}$ to be completely independent with the error terms ε_{it}, our theoretical results depend upon them being uncorrelated. We in particular need this restriction in order to prove asymptotic consistency for the differencing estimator $\widehat{\boldsymbol{\beta}}_i$ of $\boldsymbol{\beta}_i$ proposed in Section 3.5.1. In principle, it would be possible to relax this assumption even further, but that would involve much more complicated estimation procedure of $\boldsymbol{\beta}_i$ and more arduous technical arguments. Assumption (C10) ensures short-range dependence among the variables in our model. Again, we can interpret this as the fact that the cumulative effect of a single error on all future values is bounded.

We employ these assumptions to prove the main theoretical results in our paper. For detailed proofs, we refer the reader to the Appendix.

Remark 3.2.1. *The conditions (C4)–(C10) can be relaxed to cover nonstationary regressors as well as stationary ones. For example, (C4) may then be replaced by*

(C4) The covariates $\mathbf{X}_{it}$ allow for the representation $\mathbf{X}_{it} = \mathbf{H}_i(t; \ldots, u_{it-1}, u_{it})$ with u_{it} being i.i.d. random variables and $\mathbf{H}_i := (H_{i1}, H_{i2}, \ldots, H_{id})^\top : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^d$ is a measurable function such that $\mathbf{H}_i(t; \mathcal{U}_{it})$ is well defined.*

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

3.3 Testing procedure

In this section, we develop a multiscale testing procedure for the problem of comparison of the trend curves m_i in model (3.2.2). 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.

3.3.1 Preliminary steps

Testing the null hypothesis $H_0 : m_1 = m_2 = \ldots = m_n$ in model (3.2.2) is a challenging task not only because it involves nonparametric estimation of the functions $m_i(\cdot)$, but also due to the presence an unknown fixed term α_i and a vector of unknown parameters $\boldsymbol{\beta}_i$. It is clear that if α_i and $\boldsymbol{\beta}_i$ are known, the problem of testing for the common time trend would be greatly simplified. That is, we would test $H_0 : m_1 = m_2 = \ldots = m_n$ in the model

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

which is a standard nonparametric regression equation. However, in reality the variables Y_{it}° are not observed since the intercept α_i and the coefficients $\boldsymbol{\beta}_i$ are not known. Nevertheless, given the appropriate estimators $\widehat{\alpha}_i$ and $\widehat{\boldsymbol{\beta}}_i$, we can consider

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

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

But before we proceed further, we fshow how to construct consitent estimates $\widehat{\alpha}_i$ and $\widehat{\boldsymbol{\beta}}_i$. To begin with, we focus on the estimation of the vector of unknown parameters $\boldsymbol{\beta}_i$. We construct the estimator $\widehat{\boldsymbol{\beta}}_i$ in the following way.

For each i, we consider the time series $\{\Delta Y_{it} : 2 \leq t \leq T\}$ of the differences $\Delta Y_{it} = Y_{it} - Y_{it-1}$. We can write

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

where $\Delta \mathbf{X}_{it} = \mathbf{X}_{it} - \mathbf{X}_{it-1}$ and $\Delta\varepsilon_{it} = \varepsilon_{it} - \varepsilon_{it-1}$. Since $m_i(\cdot)$ is Lipschitz (by our assumption that $m_i(\cdot)$ is continuously differentiable on $[0,1]$), we can use the fact that $\big|m_i\big(\frac{t}{T}\big) - m_i\big(\frac{t-1}{T}\big)\big| = O\big(\frac{1}{T}\big)$ and rewrite

$$\Delta Y_{it} = \boldsymbol{\beta}_i^\top \Delta \mathbf{X}_{it} + \Delta\varepsilon_{it} + O\Big(\frac{1}{T}\Big). \tag{3.3.1}$$

Now, for each i we employ the least squares estimation method to estimate $\boldsymbol{\beta}_i$ in (3.3.1), treating $\Delta \mathbf{X}_{it}$ as the regressors and ΔY_{it} as the response variable. That is, we

propose the following differencing estimator:

$$\widehat{\boldsymbol{\beta}}_i = \Big(\sum_{t=2}^{T} \Delta \mathbf{X}_{it} \Delta \mathbf{X}_{it}^\top \Big)^{-1} \sum_{t=2}^{T} \Delta \mathbf{X}_{it} \Delta Y_{it} \tag{3.3.2}$$

We will show in Section 3.5.1 that $\widehat{\boldsymbol{\beta}}_i$ is a consistent estimator of $\boldsymbol{\beta}_i$ with the property $\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i = O_P(T^{-1/2})$.

Next, given $\widehat{\boldsymbol{\beta}}_i$, consider an appropriate estimator $\widehat{\alpha}_i$ for the intercept α_i calculated by

$$\begin{aligned} \widehat{\alpha}_i &= \frac{1}{T} \sum_{t=1}^{T} \big(Y_{it} - \widehat{\boldsymbol{\beta}}_i^\top \mathbf{X}_{it} \big) = \frac{1}{T} \sum_{t=1}^{T} \big(\boldsymbol{\beta}_i^\top \mathbf{X}_{it} - \widehat{\boldsymbol{\beta}}_i^\top \mathbf{X}_{it} + \alpha_i + m_i(t/T) + \varepsilon_{it} \big) = \\ &= \big(\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i \big)^\top \frac{1}{T} \sum_{t=1}^{T} \mathbf{X}_{it} + \alpha_i + \frac{1}{T} \sum_{i=1}^{T} m_i(t/T) + \frac{1}{T} \sum_{i=1}^{T} \varepsilon_{it}. \end{aligned} \tag{3.3.3}$$

Note that $\frac{1}{T}\sum_{i=1}^{T} \varepsilon_{it} = O_P(T^{-1/2})$ and $\frac{1}{T}\sum_{i=1}^{T} m_i(t/T) = O(T^{-1})$ due to Lipschitz continuity of m_i and normalization $\int_0^1 m_i(u)du = 0$. Furthermore, $\frac{1}{T}\sum_{t=1}^{T} \mathbf{X}_{it} = O_P(1)$ by Chebyshev's inequality and $\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i = O_P(T^{-1/2})$. Plugging all these results together in (3.3.3), we get that $\widehat{\alpha}_i - \alpha_i = O_P(T^{-1/2})$. Thus, the unobserved variables $Y_{it}^\circ := Y_{it} - \boldsymbol{\beta}_i^\top \mathbf{X}_{it} - \alpha_i = m_i(t/T) + \varepsilon_{it}$ can be well approximated by $\widehat{Y}_{it}$ since $\widehat{Y}_{it} = Y_{it} - \widehat{\alpha}_i - \widehat{\boldsymbol{\beta}}_i^\top \mathbf{X}_{it} = Y_{it}^\circ + O_P(T^{-1/2})$.

We now turn to the estimator of the long-run error variance $\sigma_i^2 = \sum_{\ell=-\infty}^{\infty} \mathrm{Cov}(\varepsilon_{i0}, \varepsilon_{i\ell})$ which is necessary for the construction of the test statistics later on. For the moment, we assume that the long-run variance does not depend on i, that is $\sigma_i^2 = \sigma^2$ for all i. We will need this further for conducting the testing procedure properly. Nevertheless, we keep the indices throughout the paper in order to be congruous in notation. We further let $\widehat{\sigma}_i^2$ be an estimator of σ_i^2 which is computed from the constructed sample $\{\widehat{Y}_{it} : 1 \le t \le T\}$. We thus regard $\widehat{\sigma}_i^2 = \widehat{\sigma}_i^2(\widehat{Y}_{i1}, \ldots, \widehat{Y}_{iT})$ as a function of the variables $\widehat{Y}_{it}$ for $1 \le t \le T$. Hence, whereas the true long-run variance is the same for all time series, the estimators are different. Throughout the section, we assume that $\widehat{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$ where the conditions on ρ_T will be provided further in Section 3.4. Details on how to construct $\widehat{\sigma}_i^2$ are deferred to Section 3.5.2.

3.3.2 Construction of the test statistics

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

$$\widehat{\psi}_{ij,T}(u,h) = \sum_{t=1}^{T} w_{t,T}(u,h)(\widehat{Y}_{it} - \widehat{Y}_{jt}), \tag{3.3.4}$$

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

$$w_{t,T}(u,h) = \frac{\Lambda_{t,T}(u,h)}{\{\sum_{t=1}^{T} \Lambda_{t,T}(u,h)^2\}^{1/2}}, \tag{3.3.5}$$

where

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

$S_{T,\ell}(u,h) = (Th)^{-1}\sum_{t=1}^{T} K(\frac{\frac{t}{T}-u}{h})(\frac{\frac{t}{T}-u}{h})^{\ell}$ for $\ell = 1,2$ and K is a kernel function. As common in the nonparametric estimation, we assume that K has the following properties:

(C11) The kernel K is non-negative, symmetric about zero and integrates to one. Moreover, it has compact support $[-1,1]$ and is Lipschitz continuous, that is, $|K(v) - K(w)| \leq C|v-w|$ for any $v, w \in \mathbb{R}$ and some constant $C > 0$.

Assumption (C11) allows us to use the usual kernel functions such as rectangular, Epanechnikov and Gaussian kernels.

We regard the kernel average $\widehat{\psi}_{ij,T}(u,h)$ as a measure of the distance between the two trend curves m_i and m_j on the interval $\mathcal{I}_{(u,h)} = [u-h, u+h]$. However, instead with working directly with the kernel averages $\widehat{\psi}_{ij,T}(u,h)$, we replace them by their normalized and corrected version:

$$\hat{\psi}^0_{ij,T}(u,h) = \Big|\frac{\widehat{\psi}_{ij,T}(u,h)}{(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2)^{1/2}}\Big| - \lambda(h). \tag{3.3.6}$$

Here, $\lambda(h) = \sqrt{2\log\{1/(2h)\}}$ is an additive correction term that balances the significance of many test statistics that correspond to different values of bandwidth parameters (see the discussion on this topic and comparison between multiscale testing procedures with and without this correction term in Khismatullina and Vogt (2020)).

We now aggregate the test statistics $\hat{\psi}^0_{ij,T}(u,h)$ for all i and j and a wide range of different locations u and bandwidths (scales) h:

$$\widehat{\Psi}_{n,T} = \max_{1\leq i<j\leq n} \max_{(u,h)\in\mathcal{G}_T} \hat{\psi}^0_{ij,T}(u,h), \tag{3.3.7}$$

In (3.3.7), $\mathcal{G}_T$ stands for the set of location-bandwidth pairs (u,h) that was mentioned in Section 3.1. We use the subscript T in $\mathcal{G}_T$ to point out that the choice of the grid depends on the sample size T. Specifically, throughout the paper, we suppose that $\mathcal{G}_T$ is some subset of $\mathcal{G}_T^{\text{full}} = \{(u,h) : u = t/T \text{ for some } 1 \leq t \leq T \text{ and } h \in [h_{\min}, h_{\max}]\}$, where $h_{\min}$ and $h_{\max}$ denote some minimal and maximal bandwidth value, respectively. As was already discussed in Section 3.1, we assume that the set of intervals $\{\mathcal{I}_{(u,h)} = [u-h, u+h] : (u,h) \in \mathcal{G}_T\}$ covers the whole unit interval. Furthermore, for our theoretical results, we require the following additional conditions to hold:

(C12) $|\mathcal{G}_T| = O(T^{\theta})$ for some arbitrarily large but fixed constant $\theta > 0$, where $|\mathcal{G}_T|$ denotes the cardinality of $\mathcal{G}_T$.

(C13) $h_{\min} \gg T^{-(1-\frac{2}{q})} \log T$, that is, $h_{\min}/\{T^{-(1-\frac{2}{q})} \log T\} \to \infty$ with $q > 4$ defined in (C2) and $h_{\max} < 1/2$.

Assumption (C12) places relatively mild restrictions on the grid $\mathcal{G}_T$: we allow the grid to grow with the sample size but only at a polynomial rate T^θ with fixed θ. This is not a severe constraint because under this limitation, we can still work with the full set of location-bandwidth points $\mathcal{G}_T = \mathcal{G}_T^{\text{full}}$ which is more than enough for most applied problems. Assumption (C13) is concerned with the minimal and the maximal bandwidths that we use for our analysis. Specifically, according to Assumption (C13), we can choose the minimal bandwidth $h_{\min}$ that converges to zero slower than $T^{-(1-\frac{2}{q})} \log T$ as the sample size T goes to infinity. The maximal bandwidth $h_{\max}$ can be picked very large.

Note that the value $\max_{(u,h)\in\mathcal{G}_T} \hat{\psi}^0_{ij,T}(u,h)$ simultaneously takes into account all intervals $\mathcal{I}_{(u,h)} = [u-h, u+h]$ with $(u,h) \in \mathcal{G}_T$. Thus, it can be interpreted as a global distance measure between the two curves m_i and m_j, and the test statistics $\widehat{\Psi}_{n,T}$ is then defined as the maximal distance between any pair of curves m_i and m_j with $i \neq j$.

In Section 3.3.3, we show how to test the null hypothesis $H_0 : m_1 = m_2 = \ldots = m_n$ using the mutliscale test statistics $\widehat{\Psi}_{n,T}$.

3.3.3 The testing procedure

Let Z_{it} for $1 \leq t \leq T$ and $1 \leq i \leq n$ be independent standard normal random variables which are independent of the error terms ε_{js} and the covariates $\mathbf{X}_{js}$ for all $1 \leq s \leq T$ and $1 \leq j \leq n$. Denote the empirical average of the variables $Z_{i1}, \ldots, Z_{iT}$ by $\bar{Z}_{i,T} = T^{-1} \sum_{t=1}^T Z_{it}$. To simplify the notation, we will omit the subscript T in $\bar{Z}_{i,T}$ in what follows. Similarly as with $\hat{\psi}^0_{ij,T}(u,h)$, for each i and j, we introduce the normalized and corrected Gaussian kernel averages

$$\phi^0_{ij,T}(u,h) = \left| \frac{\phi_{ij,T}(u,h)}{(\sigma_i^2 + \sigma_j^2)^{1/2}} \right| - \lambda(h), \tag{3.3.8}$$

where

$$\phi_{ij,T}(u,h) = \sum\nolimits_{t=1}^T w_{t,T}(u,h) \left\{ \sigma_i (Z_{it} - \bar{Z}_i) - \sigma_j (Z_{jt} - \bar{Z}_j) \right\} \tag{3.3.9}$$

with $w_{t,T}(u,h)$ defined in (3.3.5).

Next, in the same way as in (3.3.7), we define the global Gaussian test statistics

$$\Phi_{n,T} = \max_{1\leq i<j\leq n} \max_{(u,h)\in\mathcal{G}_T} \phi^0_{ij,T}(u,h) \tag{3.3.10}$$

and denote its $(1-\alpha)$-quantile by $q_{n,T}(\alpha)$.

Our multiscale test of the hypothesis $H_0 : m_1 = m_2 = \ldots = m_n$ is defined as follows:

For a given significance level $\alpha \in (0,1)$, we reject H_0 if $\widehat{\Psi}_{n,T} > q_{n,T}(\alpha)$.

Remark 3.3.1. *To prove the theoretical results in Section 3.4, we will use the following fact. By our assumption that the long-run variance σ_i^2 does not depend on i (i.e. $\sigma_i^2 = \sigma_j^2 = \sigma^2$), we can rewrite the Gaussian normalized kernel averages* (3.3.8) *as*

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

which means that the distribution of the Gaussian test statistics does not depend neither on the data $\mathcal{Z}_i = \{(Y_i, \mathbf{X}_i) : 1 \le t \le T)\}, \mathcal{Z}_j = \{(Y_j, \mathbf{X}_j) : 1 \le t \le T)\}$, nor on any unknown quantities such as σ_i^2 or σ_j^2, and thus can be regarded as known. In addition to exploiting this fact while proving the theoretical results, we will also use it for calculating (approximately) the quantiles of $\Phi_{n,T}$ by the Monte Carlo simulations in Section 3.3.5. However, for the sake of similarity to $\hat{\psi}^0_{ij,T}(u,h)$, in what follows, we will stick to the definition (3.3.8)*, which involves the long-run variances σ_i and σ_j.*

Remark 3.3.2. *By construction, the $(1-\alpha)$ Gaussian quantile $q_{n,T}(\alpha)$ depends not only on the number of times series considered n and the sample size T, but on the choice of the set of location-bandwidth pairs $\mathcal{G}_T$ as well. However, we do not explicitly include this dependence since we believe it will only lead to the unnecessary complication of the notation.*

3.3.4 Locating the differences

Suppose we reject the null hypothesis H_0. This fact does not provide us with a lot of information about the behaviour of the trend functions $m_i(\cdot)$. After performing the test described in Section 3.3.3, we can only make confidence statements that some of the trend functions are not equal somewhere on $[0,1]$ (with a given statistical confidence), but we can not tell which of the functions are different and where they differ. Hence, we need an additional step in the testing procedure in order to locate those differences.

Formally, for a given pair of time series (i,j) and for any given interval $\mathcal{I}_{(u,h)} = [u-h, u+h]$ such that $(u,h) \in \mathcal{G}_T$ we consider the hypothesis

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

We view $H_0^{[i,j]}(u,h)$ as the 'local' null hypothesis because it is concerned with only two trend functions $m_i(\cdot)$ and $m_j(\cdot)$ and their equality on a small, 'local', interval $\mathcal{I}_{(u,h)} = [u-h, u+h]$. In contrast, we refer to H_0 introduced in (3.1.2) as the global null hypothesis.

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

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

For each pair of time series (i,j), denote the set of intervals $\mathcal{I}_{(u,h)}$ that consists of the intervals where we reject $H_0^{[i,j]}(u,h)$ at a significance level α by $\mathcal{S}^{[i,j]}(\alpha)$. We will prove later in Section 3.4, that we can make the following confidence statements:

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

3.3.5 Implementation of the test in practice

In practice, we implement the test procedure described in Sections 3.3.3 and 3.3.4 in the following way.

Step 1. Fix a significance level $\alpha \in (0,1)$.

Step 2. Compute the (approximated) quantile $q_{n,T}(\alpha)$ by Monte Carlo simulations. Specifically, draw a large number N (say $N = 5000$) of samples of independent standard normal random variables $\{Z_{it}^{(\ell)} : 1 \leq t \leq T, 1 \leq i \leq n\}$ for $1 \leq \ell \leq N$. For each sample ℓ, compute the value $\Phi_{n,T}^{(\ell)}$ of the Gaussian test statistics $\Phi_{n,T}$ and store them. Calculate the empirical $(1-\alpha)$-quantile $\hat{q}_{n,T}(\alpha)$ from the stored values $\{\Phi_T^{(\ell)} : 1 \leq \ell \leq N\}$. Use $\hat{q}_{n,T}(\alpha)$ as an approximated value of the quantile $q_{n,T}(\alpha)$.

Step 3. Carry out the test for the global hypothesis H_0 by calculating $\widehat{\Psi}_{n,T}$ and checking if $\widehat{\Psi}_{n,T} > q_{n,T}(\alpha)$. Reject the null if it is true.

Step 4. For each i, j, $1 \leq i < j \leq n$, and each $(u,h) \in \mathcal{G}_T$, carry out the test for the local null hypothesis $H_0^{[i,j]}(u,h)$ by checking if $\hat{\psi}_{ij,T}^0(u,h) > q_{n,T}(\alpha)$. For each pair of time series (i,j), find the set of intervals $\mathcal{S}^{[i,j]}(\alpha)$ that consists of the intervals where we reject $H_0^{[i,j]}(u,h)$.

Step 5. Display the results. One of the possible ways to do that is to produce a separate plot for each of the pairwise comparisons and draw only the intervals where we reject the corresponding local null. Formally, on each of the plots that present the results of the comparison of time series i and j, we display the intervals $\mathcal{I}_{(u,h)} = [u-h, u+h] \in \mathcal{S}^{[i,j]}(\alpha)$, i.e. the (rescaled) time intervals where we reject $H_0^{[i,j]}(u,h)$.

3.4 Theoretical properties of the test

In order to invesitgate the theoretical properties of our multiscale test, we introduce two auxiliary test statistics. First auxiliary test statistics $\widehat{\Phi}_{n,T}$ can be regarded as a version of $\widehat{\Psi}_{n,t}$ which is exactly equal to it under the global null:

$$\widehat{\Phi}_{n,T} = \max_{1 \leq i < j \leq n} \max_{(u,h) \in \mathcal{G}_T} \widehat{\phi}_{ij,T}^0(u,h), \tag{3.4.1}$$

where

$$\widehat{\phi}^0_{ij,T}(u,h) = \left| \frac{\widehat{\phi}_{ij,T}(u,h)}{\{\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\}^{1/2}} \right| - \lambda(h) \tag{3.4.2}$$

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

Here we denote $\bar{\varepsilon}_i = \bar{\varepsilon}_{i,T} := T^{-1}\sum_{t=1}^T \varepsilon_{it}$ and $\bar{\mathbf{X}}_i = \bar{\mathbf{X}}_{i,T} := T^{-1}\sum_{t=1}^T \mathbf{X}_{it}$. Note that under the global null, we have $\widehat{\phi}_{ij,T}(u,h) = \widehat{\psi}_{ij,T}(u,h)$, $\widehat{\phi}^0_{ij,T}(u,h) = \widehat{\psi}^0_{ij,T}(u,h)$ and $\widehat{\Phi}_{n,T} = \widehat{\Psi}_{n,T}$, where the first two equalities hold true even under the corresponding local null $H_0^{[i,j]}(u,h)$. Hence, in order to determine the distribution of our main test statistic $\widehat{\Psi}_{n,T}$ under the null, we can simply study the behaviour of $\widehat{\Phi}_{n,T}$.

However, $\widehat{\Phi}_{n,T}$ depends on the covariates $\mathbf{X}_{it}$ whereas the Gaussian version $\Phi_{n,T}$ that is used to calculate critical values for our test (defined in (3.3.10)) is independent of them. This is the reason why we need to introduce additional intermediate test statistic that does not include the covariates, therefore, connecting $\widehat{\Phi}_{n,T}$ and $\Phi_{n,T}$. This intermediate test statistics will play an important role in the proof of our main theoretical result.

Formally, for each i, j we construct the kernel averages as

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

We can view these kernel averages as constructed under the null from the unobserved variables $\widehat{\widehat{Y}}_{it}$ and $\widehat{\widehat{Y}}_{jt}$ given by the following formula:

$$\begin{aligned}\widehat{\widehat{Y}}_{it} :&= Y_{it} - \boldsymbol{\beta}_i^\top \mathbf{X}_{it} - \frac{1}{T}\sum_{t=1}^{T}\big(Y_{it} - \boldsymbol{\beta}_i^\top \mathbf{X}_{it}\big) = \\ &= m_i\Big(\frac{t}{T}\Big) - \frac{1}{T}\sum_{t=1}^{T} m_i\Big(\frac{t}{T}\Big) + \varepsilon_{it} - \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}.\end{aligned}$$

The intermediate statistic $\widehat{\widehat{\Phi}}_{n,T}$ is then defined as

$$\widehat{\widehat{\Phi}}_{n,T} = \max_{1\le i<j\le n}\max_{(u,h)\in\mathcal{G}_T}\left\{\left|\frac{\widehat{\widehat{\phi}}_{ij,T}(u,h)}{\big(\widehat{\widehat{\sigma}}_i^2 + \widehat{\widehat{\sigma}}_j^2\big)^{1/2}}\right| - \lambda(h)\right\} \tag{3.4.3}$$

with $\widehat{\widehat{\sigma}}_i^2$ being 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 unobserved sample $\{\widehat{\widehat{Y}}_{it} : 1 \le t \le T\}$. We thus regard $\widehat{\widehat{\sigma}}_i^2 = \widehat{\widehat{\sigma}}_i^2(\widehat{\widehat{Y}}_{i1}, \ldots, \widehat{\widehat{Y}}_{iT})$ as a function of the variables $\widehat{\widehat{Y}}_{it}$ for $1 \le t \le T$. As with the estimator $\widehat{\sigma}_i^2$, we assume that $\widehat{\widehat{\sigma}}_i^2 = \sigma_i^2 + o_p(\rho_T)$ with $\rho_T = o(\sqrt{h_{\min}}/\log T)$.

The statistics $\widehat{\widehat{\Phi}}_{n,T}$ can thus be viewed as a version of the statistic $\widehat{\Phi}_{n,T}$ without the

covariates. We formally prove that these two statistics are close in Proposition 3.C.12.

Now we can formally state our main theoretical result which characterizes the asymptotic behaviour of the statistic $\widehat{\Phi}_{n,T}$.

Theorem 3.4.1. *Suppose that the error processes $\mathcal{E}_i = \{\varepsilon_{it} : 1 \leq t \leq T\}$ are independent across i and satisfy (C1)–(C3) for each i. Moreover, let (C4)–(C13) be fulfilled and assume that for all i, $m_i(\cdot)$ is a continuously differentiable function on $[0,1]$ satisfying the property $\int_0^1 m_i(u)du = 0$. Furthermore, for all i, $i \in \{1,\dots,n\}$ assume that we have $\sigma_i^2 = \sigma^2$, $\widehat{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$ and $\widehat{\widehat{\sigma}}_i^2 = \sigma_i^2 + o_p(\rho_T)$ with $\rho_T = o(\sqrt{h_{\min}}/\log T)$. Then*

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

Theorem 3.4.1 is the principal instrument for deriving theoretical properties of our multiscale test. The full proof of the theorem is provided in the Appendix. Here, we briefly present the main arguments.

First, we show that the distribution of the intermediate statistics $\widehat{\widehat{\Phi}}_{n,T}$ introduced in (3.4.3) is indeed close to the distribution of $\widehat{\Phi}_{n,T}$, and therefore, we can approximate the distribution of $\widehat{\Phi}_{n,T}$ with the help of $\widehat{\widehat{\Phi}}_{n,T}$.

Second, we show that we can replace $\widehat{\widehat{\Phi}}_{n,T}$ by an identically distributed version $\widetilde{\Phi}_{n,T}$ which is close to the Gaussian statistics $\Phi_{n,T}$ defined in (3.3.10). Formally, by the means of strong approximation theory derived in Berkes et al. (2014) we prove that there exist statistics $\widetilde{\Phi}_{n,T}$ which are distributed as $\widehat{\widehat{\Phi}}_{n,T}$ for any $T \geq 1$ and which have the property that

$$\big|\widetilde{\Phi}_{n,T} - \Phi_{n,T}\big| = o_p(\delta_T), \tag{3.4.4}$$

where $\delta_T = o(1)$.

Then, we employ the anti-concentration results derived in Chernozhukov et al. (2015) in order to show that $\Phi_{n,T}$ does not concentrate too strongly in small regions of the form $[x - \delta_T, x + \delta_T]$. Or, in other words, it holds that

$$\sup_{x\in\mathbb{R}} \mathbb{P}\big(|\Phi_{n,T} - x| \leq \delta_T\big) = o(1) \tag{3.4.5}$$

Taking (3.4.4) together with (3.4.5) and the fact that $\widetilde{\Phi}_{n,T}$ has the same distribution as $\widehat{\widehat{\Phi}}_{n,T}$ yileds that

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

And finally, by the fact mentioned in the beginning of this proof that the distribution of the intermediate statistics $\widehat{\widehat{\Phi}}_{n,T}$ is close to the distribution of $\widehat{\Phi}_{n,T}$, we conclude that

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

which immediately implies the statement of Theorem 3.4.1.

Remark 3.4.1. *The proof of Theorem 3.4.1 builds on two important theoretical results: strong approximation theory developed in Berkes et al. (2014) and anti-concentration results proved in Chernozhukov et al. (2015). These results were already combined together for the purpose of developing the multiscale test for dependent data in Khismatullina and Vogt (2020). We can say that our proof can be regarded as a further development of the proof strategy in Khismatullina and Vogt (2020) where they proposed a similar testing procedure for investigating properties of the trend function in one time series. We extend their theoretical result not only by working with multiple time series, but also by including the covariate terms in the model* (3.1.1)*. Hence, our proof strategy builds on the similar stones but is much more technically involved.*

Now we examine the theoretical properties of the testing procedure proposed in Sections 3.3.3 and 3.3.4 with the help of Theorem 3.4.1. The following proposition (which is a direct consequence of Theorem 3.4.1) states that our test has correct (asymptotical) size.

Proposition 3.4.1. *Suppose that the conditions of Theorem 3.4.1 are satisfied. Then under the null H_0, we have*

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

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

Proposition 3.4.2. *Let the conditions of Theorem 3.4.1 be satisfied. Moreover, assume that for some pair of indices i and j, the functions $m_i = m_{i,T}$ and $m_j = m_{j,T}$ have the following property: There exists $(u,h) \in \mathcal{G}_T$ with $[u-h, u+h] \subseteq [0,1]$ such that $m_{i,T}(w) - m_{j,T}(w) \geq c_T\sqrt{\log T/(Th)}$ for all $w \in [u-h, u+h]$ or $m_{j,T}(w) - m_{i,T}(w) \geq c_T\sqrt{\log T/(Th)}$ for all $w \in [u-h, u+h]$, where $\{c_T\}$ is any sequence of positive numbers with $c_T \to \infty$. Then*

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

Proof of Proposition 3.4.2 is provided in the Appendix.

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

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

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

Proposition 3.4.3. *Suppose that the conditions of Theorem 3.4.1 are satisfied. Then*

$$FWER(\alpha) \leq \alpha.$$

Proposition 3.4.3 is a direct consequence of Theorem 3.4.1. Nevertheless, the detailed proof of the proposition is provided in the Appendix.

The following corollary is an immediate consequence of Proposition 3.4.3 and gives the theoretical justification necessary for making simultaneous confidence statements about the locations of the differences between the trends.

Corollary 3.4.1. *Under the conditions of Theorem 3.4.1, for any given $\alpha \in (0,1)$ we have*

$$\mathbb{P}\Big(\forall\, i,j \in \{1,\ldots,n\}, (u,h) \in \mathcal{G}_T \text{ such that } H_0^{[i,j]}(u,h) \text{ is true} : |\hat{\psi}^0_{ij,T}(u,h)| \leq q_{n,T}(\alpha)\Big) \geq 1-\alpha+o(1).$$

With the help of Corollary 3.4.1, we are able to make simultaneous confidence statements about which of the trends are different and where:

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

3.5 Estimation of the parameters

3.5.1 Estimation of $\boldsymbol{\beta}_i$

As was already mentioned in Section 3.3.1, for each i, we construct a differencing estimator $\widehat{\boldsymbol{\beta}}_i$ of the vector of unknown parameters $\boldsymbol{\beta}_i$ using the first differences:

$$\widehat{\boldsymbol{\beta}}_i = \Big(\sum_{t=2}^{T} \Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^\top\Big)^{-1} \sum_{t=2}^{T} \Delta\mathbf{X}_{it}\Delta Y_{it} \tag{3.5.1}$$

where $\Delta\mathbf{X}_{it} = \mathbf{X}_{it} - \mathbf{X}_{it-1}$ and $\Delta Y_{it} = Y_{it} - Y_{it-1}$. The asymptotic consistency for this differencing estimator is given by the following theorem:

Theorem 3.5.1. *Under the conditions of Theorem 3.4.1, we have*

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

where $\widehat{\boldsymbol{\beta}}_i$ is the differencing estimator given by (3.5.1).

Detailed proof of the Theorem 3.5.1 is provided in the Appendix. Here we briefly outline the main steps of the proof.

After rearranging the terms, we can write

$$\begin{aligned}\sqrt{T}(\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i) =& \Big(\frac{1}{T}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^{\top}\Big)^{-1}\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta m_{it} \\ &+ \Big(\frac{1}{T}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^{\top}\Big)^{-1}\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta\varepsilon_{it},\end{aligned} \tag{3.5.2}$$

where $\Delta m_{it} = m_i\left(\frac{t}{T}\right) - m_i\left(\frac{t-1}{T}\right)$ and $\Delta\varepsilon_{it} = \varepsilon_{it} - \varepsilon_{it-1}$.

We look at each part of (3.5.2) separately. First, by Assumption (C6) and applying Chebyshev's and Cauchy-Schwarz inequalities we show that

$$\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta m_{it} = O_P\Big(\frac{1}{\sqrt{T}}\Big).$$

Then, by similar arguments and applying Proposition 3.D.14, we have that

$$\left|\Big(\frac{1}{T}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^{\top}\Big)^{-1}\right| = O_P(1),$$

where $|A|$ with A being a matrix is any matrix norm.

These two facts together lead to the fact that the first summand in (3.5.2) is $O_P(1/\sqrt{T})$.

Finally, we turn our attention to the second summand in (3.5.2). We already know that $\left|\left(\frac{1}{T}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^{\top}\right)^{-1}\right| = O_P(1)$. Moreover, by Proposition 3.D.17,

$$\left|\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}\right| = O_P(1).$$

Hence, we have that

$$\Big(\frac{1}{T}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^{\top}\Big)^{-1}\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta\varepsilon_{it} = O_P(1). \tag{3.5.3}$$

The statement of the theorem follows.

3.5.2 Estimation of σ_i^2

Following Kim (2016), we estimate the long-run variance σ_i for each of the time series i using the variant of the subseries variance estimator proposed first by

Carlstein (1986) and then extended by Wu and Zhao (2007). Formally, we set

$$\widehat{\sigma}_i^2 = \frac{1}{2(M-1)s_T} \sum_{m=1}^{M} \left[\sum_{t=1}^{s_T} \left(Y_{i(t+ms_T)} - Y_{i(t+(m-1)s_T)} - \widehat{\boldsymbol{\beta}}_i^\top \left(\mathbf{X}_{i(t+ms_T)} - \mathbf{X}_{i(t+(m-1)s_T)} \right) \right) \right]^2, \tag{3.5.4}$$

where s_T is the length of subseries and $M = \lfloor T/s_T \rfloor$ is the largest integer not exceeding T/s_T. As per the optimality result in Carlstein (1986), we set $s_T \asymp T^{1/3}$. For a finite sample, we choose $s_T = \lfloor T^{1/3} \rfloor$. According to Lemma 3.D.11 in Appendix, $\widehat{\sigma}_i^2$ is an asymptotically consistent estimator of σ_i^2 with the rate of convergence $O_P(T^{-2/3})$. Recall that the rate of convergence of $\widehat{\sigma}_i^2$ necessary for proving our theoretical results is $o_P(\rho_T)$ with $\rho_T = o(\sqrt{h_{\min}}/\log T)$. Under Assumption (C13), we have that $h_{\min} \gg T^{-\left(1-\frac{2}{q}\right)} \log T$, hence, $\sqrt{h_{\min}}/\log T \gg T^{-\left(\frac{1}{2}-\frac{1}{q}\right)}/\sqrt{\log T}$. This means that we can for example take $\rho_T = T^{-\left(\frac{1}{2}-\frac{1}{q}\right)}/\sqrt{\log T}$, which is still a slower rate of convergence than $T^{-2/3}$. To sum up, the subseries variance estimator provided by (3.5.4) satisfies the necessary conditions for Theorem 3.4.1, and thus can be used for the construction of our multiscale statistics $\widehat{\Psi}_{n,T}$.

3.6 Conclusion

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

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

For the proof of the theoretical results, the main tools that are used are strong approximation theory developed in Berkes et al. (2014) and the anti-concentration bounds for Gaussian random vectors verified in Chernozhukov et al. (2015). The proof strategy that

we employ in our paper has already been used in Khismatullina and Vogt (2020), however, in that paper the authors proposed a multiscale method for testing qualitative hypotheses only about one time series. Our method can be regarded as a generalized version of the test developed in Khismatullina and Vogt (2020) where we not only consider comparison between various time series, but also add the covariates to the model and propose an estimation procedure for the unknown parameters.

Regarding future research, this project suggests some interesting issues and topics for consideration. First, consider the situation that the null hypothesis $H_0 : m_1 = \ldots = m_n$ is violated in the general panel data model (3.1.1). Even though some of the trend functions m_i are different in this case, there may still be groups of time series with the same time trend. An interesting statistical problem to investigate in the future is how to estimate the unknown groups (and their unknown number) from the data. Second, as was already mentioned, it should be possible to extend our theoretical results to the case where the number of time series slowly grows with the sample size. Further insight can be gained by broadening the current work in these and possibly other directions.

APPENDICES

In the Appendix, we provide detailed proofs for the theoretical results from Sections 3.4 and 3.5. We use the following notation: The symbol C denotes a universal real constant which may take a different value on each occurrence. For $a, b \in \mathbb{R}$, we write $a \vee b = \max\{a, b\}$. For $x \in \mathbb{R}, x \geq 0$, we write $\lfloor x \rfloor$ to denote the integer value of x and $\lceil x \rceil$ to denote the smallest integer greater than or equal to x. 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 the distribution function of the standard normal distribution, respectively.

3.A Statistics used in the Appendix

In the proof of Theorem 3.4.1, we use a number of different test statistics, either already defined in Section 3.3 or first introduced below. Each of these statistics plays an important role in one or more steps of the proof. In the following list, we present these statistics, describe how they are constructed and explain in which parts of the proof they are used.

- Our main multiscale test statistic (defined in (3.3.7)):

$$\widehat{\Psi}_{n,T} = \max_{1 \leq i < j \leq n} \max_{(u,h) \in \mathcal{G}_T} \left\{ \left| \frac{\widehat{\psi}_{ij,T}(u,h)}{(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2)^{1/2}} \right| - \lambda(h) \right\},$$
$$\text{with} \quad \widehat{\psi}_{ij,T}(u,h) = \sum_{t=1}^{T} w_{t,T}(u,h)(\widehat{Y}_{it} - \widehat{Y}_{jt}).$$

  This statistic is our main quantity of interest because the kernel average $\widehat{\psi}_{ij,T}(u,h)$ measures the approximate distance between the trends m_i and m_j on an interval $\mathcal{I}_{(u,h)} = [u-h, u+h]$.

- The Gaussian statistic that is used for calculating the critical values for our test procedure (defined in (3.3.10)):

$$\Phi_{n,T} = \max_{1 \leq i < j \leq n} \max_{(u,h) \in \mathcal{G}_T} \left\{ \left| \frac{\phi_{ij,T}(u,h)}{(\sigma_i^2 + \sigma_j^2)^{1/2}} \right| - \lambda(h) \right\},$$
$$\text{with} \quad \phi_{ij,T}(u,h) = \sum\nolimits_{t=1}^{T} w_{t,T}(u,h) \left\{ \sigma_i(Z_{it} - \bar{Z}_i) - \sigma_j(Z_{jt} - \bar{Z}_j) \right\}.$$

- Auxiliary test statistic (defined in (3.4.1)) that can be regarded as the version of our multiscale statistic under the null.

$$\widehat{\Phi}_{n,T} = \max_{1 \leq i < j \leq n} \max_{(u,h) \in \mathcal{G}_T} \left\{ \left| \frac{\widehat{\phi}_{ij,T}(u,h)}{\{\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\}^{1/2}} \right| - \lambda(h) \right\},$$
$$\text{with} \quad \widehat{\phi}_{ij,T}(u,h) = \sum_{t=1}^{T} w_{t,T}(u,h) \big\{ (\varepsilon_{it} - \bar{\varepsilon}_i) + (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top (\mathbf{X}_{it} - \bar{\mathbf{X}}_i)$$
$$- (\varepsilon_{jt} - \bar{\varepsilon}_j) - (\boldsymbol{\beta}_j - \widehat{\boldsymbol{\beta}}_j)^\top (\mathbf{X}_{jt} - \bar{\mathbf{X}}_j) \big\}.$$

Our main theoretical result (Theorem 3.4.1) investigates the distribution of $\widehat{\Phi}_{n,T}$.

- Intermediate statistic that is close to $\widehat{\Phi}_{n,T}$ but is constructed from the kernel averages $\widehat{\widehat{\phi}}_{ij,T}(u,h)$ that are different from $\widehat{\phi}_{ij,T}(u,h)$ only by the fact that they do not include the covariates X_{it}:

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

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

We can view these kernel averages as constructed (under the null) from the unobserved variables $\widehat{\widehat{Y}}_{it}$ that are defined by

$$\widehat{\widehat{Y}}_{it} := Y_{it} - \boldsymbol{\beta}_i^\top \mathbf{X}_{it} - \frac{1}{T}\sum_{t=1}^{T}\big(Y_{it} - \boldsymbol{\beta}_i^\top \mathbf{X}_{it}\big) =$$

$$= m_i\Big(\frac{t}{T}\Big) - \frac{1}{T}\sum_{t=1}^{T} m_i\Big(\frac{t}{T}\Big) + \varepsilon_{it} - \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}.$$

The definition of $\widehat{\widehat{\phi}}^0_{ij,T}(u,h)$ also includes the auxiliary estimator $\widehat{\widehat{\sigma}}_i^2$ of the long-run error variance σ_i^2 which is computed from the augmented sample $\{\widehat{\widehat{Y}}_{it} : 1 \le t \le T\}$. We thus regard $\widehat{\widehat{\sigma}}_i^2 = \widehat{\widehat{\sigma}}_i^2(\widehat{\widehat{Y}}_{i1}, \ldots, \widehat{\widehat{Y}}_{iT})$ as a function of the variables $\widehat{\widehat{Y}}_{it}$ for $1 \le t \le T$. As with $\widehat{\sigma}_i^2$, we assume that $\widehat{\widehat{\sigma}}_i^2 = \sigma_i^2 + o_p(\rho_T)$ with $\rho_T = o(\sqrt{h_{\min}}/\log T)$.

- Auxiliary statistic that has the same distribution as $\widehat{\widehat{\Phi}}_{n,T}$ for each $T = 1, 2, \ldots$:

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

$$\text{with} \quad \widetilde{\phi}_{ij,T}(u,h) = \sum\nolimits_{t=1}^{T} w_{t,T}(u,h)\big\{(\widetilde{\varepsilon}_{it} - \bar{\widetilde{\varepsilon}}_i) - (\widetilde{\varepsilon}_{jt} - \bar{\widetilde{\varepsilon}}_j)\big\},$$

where $[\widetilde{\varepsilon}_{i1}, \ldots, \widetilde{\varepsilon}_{iT}] \overset{\mathcal{D}}{=} [\varepsilon_{i1}, \ldots, \varepsilon_{iT}]$ for each i and T. In Proposition 3.C.8, using the strong approximation theory by Berkes et al. (2014), we formally prove that such statistic exists and has the property of being close to the Gaussian statistic $\Phi_{n,T}$.

3.B Auxiliary results

Here, we state some auxiliary results that will be used further in the proof of Theorem 3.4.1.

Definition 3.B.1. *For a given $q > 0$ and $\alpha > 0$, we define dependence adjusted norm as* $||X_\cdot||^q_{q,\alpha} = \sup_{m\ge 0}(m+1)^\alpha \sum_{t=m}^{\infty} \delta_q(X,t)$.

Theorem 3.B.2. *Wu and Wu (2016) Assume that* $||X_\cdot||_{q,\alpha}^q < \infty$, *where* $q > 2$ *and* $\alpha > 0$, *and* $\sum_{t=1}^T a_t^2 = T$. *Moreover, assume that* $\alpha > 1/2 - 1/q$. *Denote* $S_T = a_1 X_1 + \ldots + a_T X_T$. *Then for all* $x > 0$,

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

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

Theorem 3.B.3. *Wu (2007) Let* $(\xi_i)_{i\in\mathbb{Z}}$ *be a stationary and ergodic Markov chain and* $g(\cdot)$ *be a measurable function. Let* $g(\xi_1) \in \mathcal{L}^q, q > 2, \mathbb{E}[g(\xi_0)] = 0$ *and* l *be a positive, nondecreasing slowly varying function. Assume that*

$$\sum_{i=n}^{\infty} \Big|\Big|\mathbb{E}[g(\xi_i)|\xi_0] - \mathbb{E}[g(\xi_i)|\xi_{-1}]\Big|\Big|_q = O\Big([\log n]^{-\beta}\Big),$$

where $0 \leq \beta < 1/q$ *and*

$$\sum_{k=1}^{\infty} \frac{k^{-\beta q}}{[l(2^k)]^q} < \infty.$$

Then $S_n = g(\xi_1) + \ldots + g(\xi_n) = o_{a.s.}\big[\sqrt{n} l(n)\big]$.

Proposition 3.B.6. *Wu (2007) Let* $(\epsilon_n)_{n\in\mathbb{Z}}$ *be i.i.d. random variables,* $\xi_n = (\ldots, \epsilon_{n-1}, \epsilon_n)$ *and* $g(\cdot)$ *be a measurable function such that* $g(\xi_n)$ *is a proper random variable for each* $n \geq 0$. *For* $k \geq 0$ *let* $\tilde{\xi}_k = (\ldots, \epsilon_{-1}, \epsilon_0', \epsilon_1, \ldots, \epsilon_{k-1}, \epsilon_k)$, *where* ϵ_0' *is an i.i.d. copy of* ϵ_0. *Let* $g(\xi_0) \in \mathcal{L}^q, q > 1$ *and* $\mathbb{E}[g(\xi_0)] = 0$. *For* $n \geq 1$ *we have*

$$\Big|\Big|\mathbb{E}[g(\xi_n)|\xi_0] - \mathbb{E}[g(\xi_n)|\xi_{-1}]\Big|\Big|_q \leq 2\Big|\Big|g(\xi_n) - g(\tilde{\xi}_n)\Big|\Big|_q.$$

Proposition 3.B.7. *Under the conditions of Theorem 3.4.1, for all* $i \in \{1, \ldots, n\}$ *it holds that*

$$\bar{X}_i = \frac{1}{T}\sum_{t=1}^{T} \mathbf{H}_i(\mathcal{U}_{it}) = o_P(1). \tag{3.B.1}$$

Proof of Proposition 3.B.7. Take any $i \in \{1, \ldots, n\}$. To prove (3.B.1), we will use two results from Wu (2007) stated above. First, fix $j \in \{1, \ldots, d\}$. Denote $\xi_t = \mathcal{U}_{it}$, $\tilde{\xi}_t = \mathcal{U}_{it}'$ and $g(\cdot) = H_{i,j}(\cdot)$. Then by Assumption (C6), $g(\xi_0) = H_{i,j}(\mathcal{U}_{i0}) \in \mathcal{L}^{q'}$ for $q' > 4$ and $\mathbb{E}[g(\xi_0)] = \mathbb{E}[H_{i,j}(\mathcal{U}_{i0})] = 0$ and we can apply Proposition 3.B.6 (Proposition 3(ii) in Wu (2007)) that says that for all $s \geq 1$ we have:

$$\big|\big|\mathbb{E}[g(\xi_s)|\xi_0] - \mathbb{E}[g(\xi_s)|\xi_{-1}]\big|\big|_{q'} \leq 2\big|\big|g(\xi_s) - g(\tilde{\xi}_s)\big|\big|_{q'},$$

or, equivalently,

$$\big\|\mathbb{E}[H_{i,j}(\mathcal{U}_{is})|\mathcal{U}_{i0}] - \mathbb{E}[H_{i,j}(\mathcal{U}_{is})|\mathcal{U}_{i(-1)}]\big\|_{q'} \leq 2\big\|H_{i,j}(\mathcal{U}_{is}) - H_{i,j}(\mathcal{U}'_{is})\big\|_{q'}.$$

Since this holds simultaneously for all $j \in \{1, \ldots, d\}$, we can use the obvious bound $\big\|H_{i,j}(\mathcal{U}_{is}) - H_{i,j}(\mathcal{U}'_{is})\big\|_{q'} \leq \big\|\mathbf{H}_i(\mathcal{U}_{is}) - \mathbf{H}_i(\mathcal{U}'_{is})\big\|_{q'} = \delta_{q'}(\mathbf{H}_i, s)$ and Assumption (C8) to write

$$0 \leq \sum_{s=t}^{\infty} \big\|\mathbb{E}[g(\xi_s)|\xi_0] - \mathbb{E}[g(\xi_s)|\xi_{-1}]\big\|_{q'} \leq \sum_{s=t}^{\infty} \delta_{q'}(\mathbf{H}_i, s) = O(t^{-\alpha}),$$

where $\alpha > 1/2 - 1/q'$.

Now we want to apply Theorem 3.B.3 (Corollary 2(i) in Wu (2007)). As a parameter β in the theorem we can take any value satisfying assumption $0 \leq \beta < 1/q'$ because for every $\beta \geq 0$ we have

$$\sum_{s=t}^{\infty} \big\|\mathbb{E}[g(\xi_s)|\xi_0] - \mathbb{E}[g(\xi_s)|\xi_{-1}]\big\|_{q'} \leq \sum_{s=t}^{\infty} \delta_{q'}(\mathbf{H}_i, s) = O(t^{-\alpha}) = O\big([\log t]^{-\beta}\big).$$

Furthermore, as a positive, nondecreasing slowly varying function l we can take $l(x) = \log^{2/q'-\beta}(x)$. Then,

$$\begin{aligned}
\sum_{k=1}^{\infty} \frac{k^{-\beta q'}}{[l(2^k)]^{q'}} &= \sum_{k=1}^{\infty} \frac{k^{-\beta q'}}{\big[\log^{2/q'-\beta}(2^k)\big]^{q'}} \\
&= \sum_{k=1}^{\infty} \frac{k^{-\beta q'}}{k^{2-\beta q'}(\log 2)^{2-\beta q'}} \\
&= \frac{1}{(\log 2)^{2-\beta q'}} \sum_{k=1}^{\infty} \frac{1}{k^2} \\
&< \infty.
\end{aligned}$$

Hence, $S_T = g(\xi_1) + \ldots + g(\xi_T) = o_{a.s.}[\sqrt{T}\log^{2/q'-\beta}(T)]$, or, equivalently, $\bar{X}_{i,j} = S_T/T = o_{a.s.}[\log^{2/q'-\beta}(T)/\sqrt{T}] = o_P(1)$ for each $j \in \{1, \ldots, d\}$. Trivially, this means that $\bar{X}_i = o_P(1)$. □

3.C Proofs of theoretical properties of the test

Proof of Theorem 3.4.1

The main steps of the proof of the Theorem 3.4.1 are described below. We will build the proof on the auxiliary results stated in 3.B.

1. First, we introduce the intermediate statistic $\widehat{\widehat{\Phi}}_{n,T}$ that can be regarded as the version of $\widehat{\Phi}_{n,T}$ where we excluded the regressors X_{it} from the construction of the kernel

averages. Next, we show that we can replace $\widehat{\widehat{\Phi}}_{n,T}$ by an identically distributed version $\widetilde{\Phi}_{n,T}$ which is close to the Gaussian statistics $\Phi_{n,T}$ defined in (3.3.10). Formally, in Proposition 3.C.8 we prove that there exist statistics $\widetilde{\Phi}_{n,T}$ for $T = 1, 2, \dots$ which are distributed as $\widehat{\widehat{\Phi}}_{n,T}$ for any $T \geq 1$ and which have the property that

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

where $\Phi_{n,T}$ is the Gaussian statistic.

2. Second, in Proposition 3.C.10 we demonstrate that $\Phi_{n,T}$ does not concentrate too strongly in small regions of the form $[x - \delta_T, x + \delta_T]$ with δ_T converging to zero as $T \to \infty$. Or, in other words, it holds that

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

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

3. Then, we make use of Lemma 3.C.10 to show that

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

This statement directly follows from the previous two steps and the fact that $\widetilde{\Phi}_{n,T}$ is distributed as $\widehat{\widehat{\Phi}}_{n,T}$ for any $n \geq 2$, $T \geq 1$.

4. In the fourth step, in Propositions 3.C.11 and 3.C.12 we formally show that the introduced intermediate statistic $\widehat{\widehat{\Phi}}_{n,T}$ is close to $\widehat{\Phi}_{n,T}$, i.e. there exists a sequence of positive numbers $\gamma_{n,T}$ that converges to 0 as $T \to \infty$ such that for all $x \in \mathbb{R}$

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

and

$$\mathbb{P}\Big(\big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| > \gamma_{n,T}\Big) = o(1). \tag{3.C.1}$$

Note that (3.C.1) does not involve x. Hence, this result is uniform over all $x \in \mathbb{R}$.

5. And finally, by the means of Proposition 3.C.13 we prove that

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

which immediately implies the statement of Theorem 3.4.1.

Step 1

The auxiliary statistics $\widehat{\Phi}_{n,T}$ defined in (3.4.1) is equal to our multiscale statistics $\widehat{\Psi}_{n,T}$ under the null hypothesis, but has the property that it depends on the known covariates $\mathbf{X}_{it}$, whereas the Gaussian version $\Phi_{n,T}$ defined in (3.3.10) is independent of them. This is the reason why we need to introduce additional intermediate test statistics that do not include the covariates and connect $\widehat{\Phi}_{n,T}$ and $\Phi_{n,T}$.

We do it in the following way. For each i and j, consider the kernel averages

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

We can view these kernel averages as constructed (under the null) based on the unobserved variables $\widehat{\widehat{Y}}_{it}$ and $\widehat{\widehat{Y}}_{jt}$ defined by

$$\begin{aligned}\widehat{\widehat{Y}}_{it} :&= Y_{it} - \boldsymbol{\beta}_i^\top \mathbf{X}_{it} - \frac{1}{T}\sum_{t=1}^{T}\big(Y_{it} - \boldsymbol{\beta}_i^\top \mathbf{X}_{it}\big) = \\ &= m_i\Big(\frac{t}{T}\Big) - \frac{1}{T}\sum_{t=1}^{T} m_i\Big(\frac{t}{T}\Big) + \varepsilon_{it} - \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}.\end{aligned}$$

The intermediate statistic is then defined as

$$\widehat{\widehat{\Phi}}_{n,T} = \max_{1\le i<j\le n}\max_{(u,h)\in\mathcal{G}_T}\left\{\left|\frac{\widehat{\widehat{\phi}}_{ij,T}(u,h)}{\big(\widehat{\widehat{\sigma}}_i^2 + \widehat{\widehat{\sigma}}_j^2\big)^{1/2}}\right| - \lambda(h)\right\} \tag{3.C.2}$$

with $\widehat{\widehat{\sigma}}_i^2$ being an estmator 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 unobserved sample $\{\widehat{\widehat{Y}}_{it} : 1 \le t \le T\}$. We thus regard $\widehat{\widehat{\sigma}}_i^2 = \widehat{\widehat{\sigma}}_i^2(\widehat{\widehat{Y}}_{i1}, \dots, \widehat{\widehat{Y}}_{iT})$ as a function of the variables $\widehat{\widehat{Y}}_{it}$ for $1 \le t \le T$. As with the estimator $\widehat{\sigma}_i^2$, we assume that $\widehat{\widehat{\sigma}}_i^2 = \sigma_i^2 + o_p(\rho_T)$ with $\rho_T = o(\sqrt{h_{\min}}/\log T)$.

The statistics $\widehat{\widehat{\Phi}}_{n,T}$ can thus be viewed as a version of the statistic $\widehat{\Phi}_{n,T}$ without the covariates. We formally prove that these two statistics are close in Step 4.

Here, we are interested in another matter. Specifically, the main theoretical result of this step is the fact that there exists a version of the multiscale statistic $\widehat{\widehat{\Phi}}_{n,T}$ with the same distributional properties and that is close to the Gaussian statistics $\Phi_{n,T}$ which distribution is known. More specifically, we prove the following result.

Proposition 3.C.8. *Under the conditions of Theorem 3.4.1, there exist statistics $\widetilde{\Phi}_{n,T}$ for $T = 1, 2, \dots$ with the following two properties: (i) $\widetilde{\Phi}_{n,T}$ has the same distribution as $\widehat{\widehat{\Phi}}_{n,T}$ as defined in* (3.C.2) *for any T, and (ii)*

$$\big|\widetilde{\Phi}_{n,T} - \Phi_{n,T}\big| = o_p\Big(\frac{T^{1/q}}{\sqrt{T h_{\min}}} + \rho_T\sqrt{\log T}\Big), \tag{3.C.3}$$

where $\Phi_{n,T}$ is a Gaussian statistic as defined in (3.3.10).

Proof of Proposition 3.C.8. For the proof, we draw on strong approximation theory for each stationary process $\mathcal{E}_i = \{\varepsilon_{it} : 1 \leq t \leq T\}$ 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 exists a standard Brownian motion $\mathbb{B}_i$ and a sequence $\{\widetilde{\varepsilon}_{it} : t \in \mathbb{N}\}$ such that $[\widetilde{\varepsilon}_{i1}, \ldots, \widetilde{\varepsilon}_{iT}] \overset{\mathcal{D}}{=} [\varepsilon_{i1}, \ldots, \varepsilon_{iT}]$ for each T and

$$\max_{1 \leq t \leq T} \Big| \sum_{s=1}^{t} \widetilde{\varepsilon}_{is} - \sigma_i \mathbb{B}_i(t) \Big| = o\big(T^{1/q}\big) \quad \text{a.s.}, \tag{3.C.4}$$

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

We apply this result for each stationary process $\mathcal{E}_i = \{\varepsilon_{it} : 1 \leq t \leq T\}$ so that each process $\widetilde{\mathcal{E}}_i = \{\widetilde{\varepsilon}_{it} : t \in \mathbb{N}\}$ is independent of $\widetilde{\mathcal{E}}_j = \{\widetilde{\varepsilon}_{jt} : t \in \mathbb{N}\}$ for $i \neq j$.

Furthermore, we define

$$\widetilde{\Phi}_{n,T} = \max_{1 \leq i < j \leq n} \max_{(u,h) \in \mathcal{G}_T} \left\{ \left| \frac{\widetilde{\phi}_{ij,T}(u,h)}{\big(\widetilde{\sigma}_i^2 + \widetilde{\sigma}_j^2\big)^{1/2}} \right| - \lambda(h) \right\}$$
$$\text{with} \quad \widetilde{\phi}_{ij,T}(u,h) = \sum\nolimits_{t=1}^{T} w_{t,T}(u,h) \big\{ (\widetilde{\varepsilon}_{it} - \bar{\widetilde{\varepsilon}}_i) - (\widetilde{\varepsilon}_{jt} - \bar{\widetilde{\varepsilon}}_j) \big\}.$$

where $\widetilde{\sigma}_i^2$ are the same estimators as $\widehat{\sigma}_i^2$ with $\widehat{Y}_{it} = (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top \mathbf{X}_{it} + m_i(t/T) + \big(\alpha_i - \widehat{\alpha}_i\big) + \varepsilon_{it}$ replaced by $\widetilde{Y}_{it} = (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top \mathbf{X}_{it} + m_i(t/T) + \big(\alpha_i - \widehat{\alpha}_i\big) + \widetilde{\varepsilon}_{it}$ for $1 \leq t \leq T$. Since $[\widetilde{\varepsilon}_{i1}, \ldots, \widetilde{\varepsilon}_{iT}] \overset{\mathcal{D}}{=} [\varepsilon_{i1}, \ldots, \varepsilon_{iT}]$, we have $\sum_{\ell=-\infty}^{\infty} \operatorname{Cov}(\widetilde{\varepsilon}_{i0}, \widetilde{\varepsilon}_{i\ell}) = \sum_{\ell=-\infty}^{\infty} \operatorname{Cov}(\varepsilon_{i0}, \varepsilon_{i\ell}) = \sigma_i^2$. Hence, by construction, $\widetilde{\sigma}_i^2 = \sigma_i^2 + o_P(\rho_T)$.

In addition, we let

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

with $\phi_{ij,T}(u,h) = \sum_{t=1}^{T} w_{t,T}(u,h) \big\{ \sigma_i(Z_{it} - \bar{Z}_i) - \sigma_j(Z_{jt} - \bar{Z}_j) \big\}$ as defined in (3.3.9) with $Z_{it} = \mathbb{B}_i(t) - \mathbb{B}_i(t-1)$. With this notation, we can write

$$\big|\widetilde{\Phi}_{n,T} - \Phi_{n,T}\big| \leq \big|\widetilde{\Phi}_{n,T} - \Phi_{n,T}^{\diamond}\big| + \big|\Phi_{n,T}^{\diamond} - \Phi_{n,T}\big|. \tag{3.C.5}$$

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

$$\big|\widetilde{\Phi}_{n,T} - \Phi_{n,T}^{\diamond}\big| \leq \max_{1 \leq i < j \leq n} \Big(\big(\widetilde{\sigma}_i^2 + \widetilde{\sigma}_j^2\big)^{-1/2} \max_{(u,h) \in \mathcal{G}_T} \big|\widetilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h)\big| \Big). \tag{3.C.6}$$

We have already noted that $\widetilde{\sigma}_i^2 = \sigma_i^2 + o_P(\rho_T)$. Moreover, for all $i \in \{1, \ldots, n\}$ we know that $\sigma_i^2 \neq 0$. Hence,

$$\max_{1 \leq i < j \leq n} \big(\widetilde{\sigma}_i^2 + \widetilde{\sigma}_j^2\big)^{-1/2} = O_P(1). \tag{3.C.7}$$

Next, using summation by parts, ($\sum_{i=1}^{n} a_i b_i = \sum_{i=1}^{n-1} A_i(b_i - b_{i+1}) + A_n b_n$ with $A_j = \sum_{j=1}^{i} a_j$) we obtain that

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

where

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

Note that by construction $A_{ij,T} = 0$ for all pairs (i,j). Denoting

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

we have

$$\begin{aligned}
\big|\widetilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h)\big| &= \Big| \sum_{t=1}^{T-1} A_{ij,t}\big(w_{t,T}(u,h) - w_{t+1,T}(u,h)\big)\Big| \\
&\le W_T(u,h) \max_{1\le t\le T} |A_{ij,t}|.
\end{aligned} \tag{3.C.8}$$

Now consider $\max_{1\le t\le T} |A_{ij,t}|$. Straightforward application of the triangle inequality provides the following bound:

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

Applying the strong approximation result (3.C.4), we can infer that

$$\max_{1\leq t\leq T}|A_{ij,t}| = o_P\big(T^{1/q}\big). \tag{3.C.9}$$

Standard arguments show that $\max_{(u,h)\in\mathcal{G}_T} W_T(u,h) = O(1/\sqrt{Th_{\min}})$. Plugging (3.C.9) in (3.C.8), and taking the result together with (3.C.7) and plugging them in (3.C.6), we can thus infer that

$$\begin{aligned}\big|\widetilde{\Phi}_{n,T} - \Phi^{\diamond}_{n,T}\big| &\leq \big(\widetilde{\sigma}_i^2 + \widetilde{\sigma}_j^2\big)^{-1/2} \max_{(u,h)\in\mathcal{G}_T} W_T(u,h) \max_{1\leq i<j\leq n}\max_{1\leq t\leq T}|A_{ij,t}| \\ &= O_P(1)\cdot O\Big(\frac{1}{\sqrt{Th_{\min}}}\Big)\cdot o_P\big(T^{1/q}\big) \\ &= o_P\Big(\frac{T^{1/q}}{\sqrt{Th_{\min}}}\Big).\end{aligned} \tag{3.C.10}$$

Now consider $|\Phi^{\diamond}_{n,T} - \Phi_{n,T}|$. Trivially,

$$\begin{aligned}\big|\Phi^{\diamond}_{n,T} - \Phi_{n,T}\big| &\leq \max_{1\leq i<j\leq n}\max_{(u,h)\in\mathcal{G}_T}\Big|\frac{\phi_{ij,T}(u,h)}{\{\widetilde{\sigma}_i^2+\widetilde{\sigma}_j^2\}^{1/2}} - \frac{\phi_{ij,T}(u,h)}{\{\sigma_i^2+\sigma_j^2\}^{1/2}}\Big| \\ &\leq \max_{1\leq i<j\leq n}\left(\Big|\big(\widetilde{\sigma}_i^2+\widetilde{\sigma}_j^2\big)^{-1/2} - \big(\widehat{\sigma}_i^2+\widehat{\sigma}_j^2\big)^{-1/2}\Big| \max_{(u,h)\in\mathcal{G}_T}|\phi_{ij,T}(u,h)|\right)\end{aligned} \tag{3.C.11}$$

Since $\widetilde{\sigma}_i^2 = \sigma_i^2 + o_P(\rho_T)$ by the note above and $\widehat{\sigma}_i^2 = \sigma_i^2 + o_P(\rho_T)$ by our assumptions, we have that

$$\max_{1\leq i<j\leq n}\Big|\big(\widetilde{\sigma}_i^2+\widetilde{\sigma}_j^2\big)^{-1/2} - \big(\widehat{\sigma}_i^2+\widehat{\sigma}_j^2\big)^{-1/2}\Big| = o_P(\rho_T). \tag{3.C.12}$$

Then, $\phi_{ij,T}(u,h) = \sum_{t=1}^T w_{t,T}(u,h)\,(\sigma_i Z_{it} - \sigma_j Z_{jt}) - \sum_{t=1}^T w_{t,T}(u,h)\,(\sigma_i\bar{Z}_i - \sigma_j\bar{Z}_j)$, where the first part is distributed as $N(0,\sigma_i^2+\sigma_j^2)$ and the second part is distributed as $N\Big(0,(\sigma_i^2+\sigma_j^2)(\sum_{t=1}^T w_{t,T}(u,h))^2/T\Big)$ for all $(u,h)\in\mathcal{G}_T$ and all $1\leq i<j\leq n$. Note that $(\sum_{t=1}^T w_{t,T}(u,h))^2 \leq C\cdot T$ by (3.C.33), $|\mathcal{G}_T| = O(T^\theta)$ for some large but fixed constant θ by Assumption (C12), n is fixed. Hence, by the well-known results in probability theory,

$$\max_{1\leq i<j\leq n}\max_{(u,h)\in\mathcal{G}_T}\big|\phi_{ij,T}(u,h)\big| = O_P(\sqrt{\log T}), \tag{3.C.13}$$

which together with (3.C.11) and (3.C.12) leads to

$$\big|\Phi^{\diamond}_{n,T} - \Phi_{n,T}\big| = o_P(\rho_T)\cdot O_P(\sqrt{\log T}) = o_P(\rho_T\sqrt{\log T}). \tag{3.C.14}$$

Plugging (3.C.10) and (3.C.14) in (3.C.5) completes the proof. □

Step 2

In this step, we establish some properties of the Gaussian statistic $\Phi_{n,T}$ defined in (3.3.10). We in particular show that $\Phi_{n,T}$ does not concentrate too strongly in small regions of the form $[x-\delta_T, x+\delta_T]$ with δ_T converging to zero.

The main technical tool for proving these results (specifically, Proposition 3.C.10) 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 3.C.9. *Khismatullina and Vogt (2020) Let $(X_1,\dots,X_p)^\top$ 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 $\overline{\sigma}$.

Proposition 3.C.10. *Under the conditions of Theorem 3.4.1, it holds that*

$$\sup_{x\in\mathbb{R}}\mathbb{P}\big(|\Phi_{n,T}-x|\le\delta_T\big)=o(1), \tag{3.C.15}$$

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

Proof of Proposition 3.C.10. We write $x=(u,h)$ along with $\mathcal{G}_T=\{x: x\in\mathcal{G}_T\}=\{x_1,\dots,x_p\}$, where $p:=|\mathcal{G}_T|\le O(T^\theta)$ for some large but fixed $\theta>0$ by our assumptions. Moreover, for $k=1,\dots,p$, we set

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

with $x_k=(x_{k1},x_{k2})$. This notation allows us to write

$$\Phi_{n,T}=\max_{1\le i<j\le n}\max_{1\le k\le 2p}U_{ij,k}=\max_{1\le l\le(n-1)np}U_l'$$

where $(U_1',\dots,U_{(n-1)np}')^\top\in\mathbb{R}^{n(n-1)p}$ is a Gaussian random vector with the following properties: (i) $\mu_l:=\mathbb{E}[U_l']=\{\mathbb{E}[U_{ij,2k}]$ or $\mathbb{E}[U_{ij,2k-1}]\}=-\lambda(x_{k2})$ and thus

$$\overline{\mu}=\max_{1\le l\le(n-1)np}|\mu_l|\le C\sqrt{\log T},$$

and (ii) $\sigma_l^2:=\mathrm{Var}(U_l')=1$ for all $1\le l\le(n-1)np$. We would like to apply Proposition 3.C.9 (Proposition S.3 in Khismatullina and Vogt (2020)) to $(U_1',\dots,U_{(n-1)np}')^\top$,

and for this, we need to check the assumptions therein. First,

$$a_{(n-1)np} := \mathbb{E}\big[\max_{1\le j\le (n-1)np}(U_l' - \mu_j)/\sigma_j\big] = \mathbb{E}\big[\max_{1\le j\le (n-1)np}(U_l' - \mu_j)\big] =: b_{(n-1)np}.$$

Moreover, as the variables $(U_l' - \mu_l)/\sigma_l$ are standard normal, we have that $a_{(n-1)np} = b_{(n-1)np} \le C\sqrt{\log((n-1)np)} \le C\sqrt{\log T}$. With this notation at hand, we can apply Proposition 3.C.9 to obtain that

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

with $\delta_T = T^{1/q}/\sqrt{Th_{\min}} + \rho_T\sqrt{\log T}$, which is the statement of Proposition 3.C.10. □

Step 3

Lemma 3.C.10. *Khismatullina and Vogt (2020) Let V_T and W_T be real-valued random variables for $T = 1, 2, \dots$ such that $V_T - W_T = o_p(\delta_T)$ with some $\delta_T = o(1)$. If*

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

then

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

Applying Lemma 3.C.10 to $\widetilde{\Phi}_{n,T}$ and $\Phi_{n,T}$ (taking $V_T = \Phi_{n,T}$ and $W_T = \widetilde{\Phi}_{n,T}$) together with the results (3.C.3) and (3.C.15) and noting the fact that $\widetilde{\Phi}_{n,T}$ is distributed as $\widehat{\widetilde{\Phi}}_{n,T}$ for any $n \ge 2$, $T \ge 1$ immediately leads to

$$\sup_{x\in\mathbb{R}} \big|\mathbb{P}(\widehat{\widetilde{\Phi}}_{n,T} \le x) - \mathbb{P}(\Phi_{n,T} \le x)\big| = o(1). \tag{3.C.16}$$

Step 4

As was already mentioned in Step 1, the statistics $\widehat{\widetilde{\Phi}}_{n,T}$ can be viewed as an approximation of the statistics $\widehat{\Phi}_{n,T}$. Heuristically, the kernel averages $\widehat{\widetilde{\phi}}_{ij,T}(u,h)$ are close to the kernel averages $\widehat{\phi}_{ij,T}(u,h)$ because of the properties of our estimators $\widehat{\boldsymbol{\beta}}_i$, $\widehat{\widetilde{\sigma}}_i^2$ and assumptions on $\mathbf{X}_{it}$. In the following two propositions we prove it formally.

Proposition 3.C.11. *For any $x \in \mathbb{R}$ and any $\gamma > 0$, we have*

$$\begin{aligned}\mathbb{P}\Big(\widehat{\widetilde{\Phi}}_{n,T} \le x - \gamma\Big) - \mathbb{P}\Big(\big|\widehat{\widetilde{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| > \gamma\Big) &\le \mathbb{P}(\widehat{\Phi}_{n,T} \le x) \\ &\le \mathbb{P}\Big(\widehat{\widetilde{\Phi}}_{n,T} \le x + \gamma\Big) + \mathbb{P}\Big(\big|\widehat{\widetilde{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| > \gamma\Big).\end{aligned} \tag{3.C.17}$$

Proof of Proposition 3.C.11. From the law of total probability and the monotonic

property of the probability function, we have

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

Analogously,

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

Combining these two inequalities together, we arrive at the desired result. □

The aim of the next proposition is to determine the sequence of values of $\gamma_{n,T}$ that may depend on n and T such that the difference between the distributions of $\widehat{\Phi}_{n,T}$ and $\widehat{\widehat{\Phi}}_{n,T}$ is not too big. In other words,

Proposition 3.C.12. *There exists a sequence of positive random numbers $\{\gamma_{n,T}\}_T$, that converges to 0 as $T \to \infty$, such that*

$$\mathbb{P}\Big(\big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| > \gamma_{n,T}\Big) = o(1). \tag{3.C.18}$$

Proof of Proposition 3.C.12. Straightforward calculations yield that

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

Obviously,

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

and

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

Furthermore, the difference of the kernel averages $\widehat{\widehat{\phi}}_{ij,T}(u,h) - \widehat{\phi}_{ij,T}(u,h)$ does not include the error term (it cancels out) and can be written as

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

Hence,

$$\begin{aligned}\big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| \leq{}& \max_{1\leq i<j\leq n}\Big|\big(\widehat{\widehat{\sigma}}_i^2 + \widehat{\widehat{\sigma}}_j^2\big)^{-1/2} - \big(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\big)^{-1/2}\Big| \max_{1\leq i<j\leq n}\max_{(u,h)\in\mathcal{G}_T}\Big|\widehat{\widehat{\phi}}_{ij,T}(u,h)\Big| \\ &+ 2\max_{1\leq i<j\leq n}\big(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\big)^{-1/2}\max_{1\leq i\leq n}\max_{(u,h)\in\mathcal{G}_T}\Big|(\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top\sum_{t=1}^{T} w_{t,T}(u,h)\mathbf{X}_{it}\Big| \\ &+ 2\max_{1\leq i<j\leq n}\big(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\big)^{-1/2}\max_{1\leq i\leq n}\big|(\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top\bar{\mathbf{X}}_i\big|\max_{(u,h)\in\mathcal{G}_T}\Big|\sum_{t=1}^{T} w_{t,T}(u,h)\Big|.\end{aligned} \tag{3.C.19}$$

We consider each of the three summands separately.

We start by looking at the first summand in (3.C.19). Since $\widehat{\widehat{\sigma}}_i^2 = \sigma_i^2 + o_P(\rho_T)$ and $\widehat{\sigma}_i^2 = \sigma_i^2 + o_P(\rho_T)$ by our assumptions, we have that

$$\max_{1\leq i<j\leq n}\Big|\big(\widehat{\widehat{\sigma}}_i^2 + \widehat{\widehat{\sigma}}_j^2\big)^{-1/2} - \big(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\big)^{-1/2}\Big| = o_P(\rho_T). \tag{3.C.20}$$

Then, we investigate $\max_{(u,h)\in\mathcal{G}_T}\big|\widehat{\widehat{\phi}}_{ij,T}(u,h)\big|$. Specifically, we are interested in its distribution. We know by Proposition 3.C.8 that there exists $\widetilde{\phi}_{ij,T}(u,h)$ that has the same distribution as $\widehat{\widehat{\phi}}_{ij,T}(u,h)$ for all $1 \leq i < j \leq n$ and all $(u,h) \in \mathcal{G}_T$.

$$\mathbb{P}\left(\max_{(u,h)\in\mathcal{G}_T}\Big|\widehat{\widehat{\phi}}_{ij,T}(u,h)\Big| \leq C\right) = \mathbb{P}\left(\max_{(u,h)\in\mathcal{G}_T}\Big|\widetilde{\phi}_{ij,T}(u,h)\Big| \leq C\right).$$

So instead of looking at the distribution of $\max_{(u,h)\in\mathcal{G}_T}\big|\widehat{\widehat{\phi}}_{ij,T}(u,h)\big|$, we now turn our attention at the distribution of $\max_{(u,h)\in\mathcal{G}_T}\big|\widetilde{\phi}_{ij,T}(u,h)\big|$ instead.

In bounding this probability, we can use the strategy from the second part of the proof of Proposition 3.C.11. For any $c_T \in \mathbb{R}$ (taking $x = \gamma = c_T/2$) we have

$$\begin{aligned}&\mathbb{P}\left(\max_{(u,h)\in\mathcal{G}_T}|\phi_{ij,T}(u,h)| \leq c_T/2\right) \\ &\leq \mathbb{P}\left(\max_{(u,h)\in\mathcal{G}_T}\Big|\widetilde{\phi}_{ij,T}(u,h)\Big| \leq c_T\right) + \mathbb{P}\left(\Big|\max_{(u,h)\in\mathcal{G}_T}\Big|\widetilde{\phi}_{ij,T}(u,h)\Big| - \max_{(u,h)\in\mathcal{G}_T}|\phi_{ij,T}(u,h)|\Big| > \frac{c_T}{2}\right)\end{aligned}$$

$$\leq \mathbb{P}\left(\max_{(u,h)\in\mathcal{G}_T}\left|\widetilde{\phi}_{ij,T}(u,h)\right| \leq c_T\right) + \mathbb{P}\left(\max_{(u,h)\in\mathcal{G}_T}\left|\widetilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h)\right| > \frac{c_T}{2}\right).$$

Hence,

$$\begin{aligned}\mathbb{P}\left(\max_{(u,h)\in\mathcal{G}_T}\left|\widehat{\widetilde{\phi}}_{ij,T}(u,h)\right| \leq c_T\right) &= \mathbb{P}\left(\max_{(u,h)\in\mathcal{G}_T}\left|\widetilde{\phi}_{ij,T}(u,h)\right| \leq c_T\right)\\ &\geq \mathbb{P}\left(\max_{(u,h)\in\mathcal{G}_T}|\phi_{ij,T}(u,h)| \leq c_T/2\right) - \mathbb{P}\left(\max_{(u,h)\in\mathcal{G}_T}\left|\widetilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h)\right| > \frac{c_T}{2}\right).\end{aligned} \tag{3.C.21}$$

By (3.C.14) we have

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

Furthermore, $\phi_{ij,T}(u,h)$ is distributed as $N(0,\sigma_i^2+\sigma_j^2)$ for all $(u,h)\in\mathcal{G}_T$ and all $1\leq i<j\leq n$, and $|\mathcal{G}_T| = O(T^\theta)$ for some large but fixed constant θ by Assumption (C12). By the standard results from the probability theory, we know that

$$\max_{(u,h)\in\mathcal{G}_T}|\phi_{ij,T}(u,h)| = O_P(\sqrt{\log T}).$$

Since $T^{1/q}/\sqrt{Th_{\min}} \ll \sqrt{\log T}$, we can take $c_T = o(\sqrt{\log T})$ in (3.C.21) to get the following:

$$\begin{aligned}&\mathbb{P}\left(\max_{(u,h)\in\mathcal{G}_T}\left|\widehat{\widetilde{\phi}}_{ij,T}(u,h)\right| \leq c_T\right)\\ &\geq \mathbb{P}\left(\max_{(u,h)\in\mathcal{G}_T}|\phi_{ij,T}(u,h)| \leq \frac{c_T}{2}\right) - \mathbb{P}\left(\max_{(u,h)\in\mathcal{G}_T}\left|\widetilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h)\right| > \frac{c_T}{2}\right)\\ &= 1 - o(1) - o(1)\\ &= 1 - o(1),\end{aligned}$$

which means that

$$\max_{(u,h)\in\mathcal{G}_T}\left|\widehat{\widetilde{\phi}}_{ij,T}(u,h)\right| = o_P(\sqrt{\log T}). \tag{3.C.22}$$

Combining (3.C.20) and (3.C.22) and taking into consideration that n is fixed, we get the following:

$$\begin{aligned}\max_{1\leq i<j\leq n}\left|\left(\widehat{\widetilde{\sigma}}_i^2+\widehat{\widetilde{\sigma}}_j^2\right)^{-1/2} - \left(\widehat{\sigma}_i^2+\widehat{\sigma}_j^2\right)^{-1/2}\right| \max_{1\leq i<j\leq n}\max_{(u,h)\in\mathcal{G}_T}\left|\widehat{\widetilde{\phi}}_{ij,T}(u,h)\right|&\\ = o_P(\rho_T)\cdot o_P(\sqrt{\log T})&\\ = o_P(1)&\end{aligned} \tag{3.C.23}$$

since by our assumption $\rho_T = O(\sqrt{h_{\min}}/\log T)$.

Now we evaluate the second summand in (3.C.19).

First, by our assumptions $\widehat{\sigma}_i^2 = \sigma_i^2 + o_P(\rho_T)$. Moreover, for all $i \in \{1, \dots, n\}$ we know that $\sigma_i^2 \neq 0$. Hence,

$$\max_{1 \leq i < j \leq n} \left(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\right)^{-1/2} = O_P(1). \tag{3.C.24}$$

Then, by Theorem 3.5.1, we know that

$$\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i = O_P\Big(\frac{1}{\sqrt{T}}\Big). \tag{3.C.25}$$

Now consider $\sum_{t=1}^{T} w_{t,T}(u,h)\mathbf{X}_{it}$. Without loss of generality, we can regard the covariates $\mathbf{X}_{it}$ to be scalars X_{it}, not vectors. The proof in case of vectors proceeds analogously.

By construction the weights $w_{t,T}(u,h)$ are not equal to 0 if and only if $T(u-h) \leq t \leq T(u+h)$. We can use this fact to rewrite

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

Note that

$$\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} w_{t,T}^2(u,h) = \sum_{t=1}^{T} w_{t,T}^2(u,h) = \sum_{t=1}^{T} \frac{\Lambda_{t,T}^2(u,h)}{\sum_{s=1}^{T}\Lambda_{s,T}^2(u,h)} = 1. \tag{3.C.26}$$

Denoting by $D_{T,u,h}$ the number of integers between $\lfloor T(u-h)\rfloor$ and $\lceil T(u+h)\rceil$ incl. (with obvious bounds $2Th \leq D_{T,u,h} \leq 2Th+2$) and using (3.C.26), we can normalize the weights as follows:

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

According to Theorem 3.B.2 (Theorem 2(ii) in Wu and Wu (2016)), if we define the weights from the theorem as $a_t = \sqrt{D_{T,u,h}} \cdot w_{t,T}(u,h)$, we can bound the probability as follows:

$$\begin{aligned}
&\mathbb{P}\left(\Big|\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} \sqrt{D_{T,u,h}} \cdot w_{t,T}(u,h)X_{it}\Big| \geq x\right) \\
&\leq C_1 \frac{\left(\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} |\sqrt{D_{T,u,h}} \cdot w_{t,T}(u,h)|^{q'}\right)||X_{i\cdot}||_{q',\alpha}^{q'}}{x^{q'}} + C_2 \exp\left(-\frac{C_3 x^2}{D_{T,u,h}||X_{i\cdot}||_{2,\alpha}^2}\right),
\end{aligned} \tag{3.C.27}$$

where $||X_{i\cdot}||_{q,\alpha}^{q}$ is the dependence adjusted norm as defined by Definition 3.B.1. Taking

any $\delta > 0$ and applying Boole's inequality and (3.C.27) subsequently, we get

$$
\begin{aligned}
&\mathbb{P}\left(\max_{(u,h)\in\mathcal{G}_T}\Big|\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} w_{t,T}(u,h)X_{it}\Big| \geq \delta\sqrt{T}\right)\\
&\leq \sum_{(u,h)\in\mathcal{G}_T}\mathbb{P}\left(\Big|\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} w_{t,T}(u,h)X_{it}\Big| \geq \delta\sqrt{T}\right)\\
&= \sum_{(u,h)\in\mathcal{G}_T}\mathbb{P}\left(\Big|\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} \sqrt{D_{T,u,h}}\cdot w_{t,T}(u,h)X_{it}\Big| \geq \delta\sqrt{D_{T,u,h}T}\right)\\
&\leq \sum_{(u,h)\in\mathcal{G}_T}\left[C_1\frac{(\sqrt{D_{T,u,h}})^{q'}\big(\sum|w_{t,T}(u,h)|^{q'}\big)||X_{i\cdot}||_{q',\alpha}^{q'}}{\big(\delta\sqrt{D_{T,u,h}T}\big)^{q'}} + C_2\exp\left(-\frac{C_3\big(\delta\sqrt{D_{T,u,h}T}\big)^2}{D_{T,u,h}||X_{i\cdot}||_{2,\alpha}^2}\right)\right]\\
&= \sum_{(u,h)\in\mathcal{G}_T}\left[C_1\frac{\big(\sum|w_{t,T}(u,h)|^{q'}\big)||X_{i\cdot}||_{q',\alpha}^{q'}}{\big(\delta\sqrt{T}\big)^{q'}} + C_2\exp\left(-\frac{C_3\delta^2T}{||X_{i\cdot}||_{2,\alpha}^2}\right)\right]\\
&\leq C_1\frac{T^\theta||X_{i\cdot}||_{q',\alpha}^{q'}}{T^{q'/2}\cdot\delta^{q'}}\max_{(u,h)\in\mathcal{G}_T}\left(\sum\nolimits_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil}|w_{t,T}(u,h)|^{q'}\right) + C_2T^\theta\exp\left(-\frac{C_3\delta^2T}{||X_{i\cdot}||_{2,\alpha}^2}\right)\\
&= C\frac{T^{\theta-q'/2}}{\delta^{q'}} + CT^\theta\exp\left(-CT\delta^2\right).
\end{aligned}
$$

where the symbol C denotes a universal real constant that does not depend neither on T, nor on δ, and takes a different value on each occurrence. Here in the last equality we used the following facts:

1. $||X_{i\cdot}||_{q',\alpha}^{q'} = \sup_{t\geq 0}(t+1)^\alpha\sum_{s=t}^\infty\delta_{q'}(H_i,s) < \infty$ holds true since $\sum_{s=t}^\infty\delta_{q'}(H_i,s) = O(t^{-\alpha})$ by Assumption (C8);

2. $\max_{(u,h)\in\mathcal{G}_T}\left(\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil}|w_{t,T}(u,h)|^{q'}\right) \leq 1$ because for every $x\in[0,1]$ we have $0\leq x^{q'/2}\leq x\leq 1$. Thus, since $\sum_{t=1}^T w_{t,T}^2(u,h) = 1$ by (3.C.26) we have $0\leq w_{t,T}^2(u,h)\leq 1$ for all $t\in\{1,\ldots,T\}$ and all $(u,h)\in\mathcal{G}_T$, we get

$$0\leq|w_{t,T}(u,h)|^{q'} = (w_{t,T}^2(u,h))^{q'/2}\leq w_{t,T}^2(u,h)\leq 1.$$

   This leads to a bound

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

3. $||X_{i\cdot}||_{2,\alpha}^2 < \infty$ (follows from 1).

By Assumption (C6), $\theta - q'/2 < 0$ and the term on the RHS of the above inequality is

converging to 0 as $T \to \infty$ for any fixed $\delta > 0$. Hence,

$$\max_{(u,h)\in\mathcal{G}_T} \left| \sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} w_{t,T}(u,h) X_{it} \right| = o_P(\sqrt{T}),$$

and similarly,

$$\max_{(u,h)\in\mathcal{G}_T} \left| \sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} w_{t,T}(u,h) \mathbf{X}_{it} \right| = o_P(\sqrt{T}). \tag{3.C.28}$$

Combining (3.C.24), (3.C.25) and (3.C.28), we get the following:

$$\begin{aligned} \max_{1\le i<j\le n} \left(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\right)^{-1/2} \max_{1\le i\le n} \max_{(u,h)\in\mathcal{G}_T} \left| (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top \sum_{t=1}^{T} w_{t,T}(u,h) \mathbf{X}_{it} \right| \\ = O_P(1) \cdot O_P(1/\sqrt{T}) \cdot o_P(\sqrt{T}) \\ = o_P(1). \end{aligned} \tag{3.C.29}$$

Now consider the third summand in (3.C.19). Similarly as before,

$$\max_{1\le i<j\le n} \left(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\right)^{-1/2} = O_P(1) \tag{3.C.30}$$

and

$$\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i = O_P\Big(\frac{1}{\sqrt{T}}\Big). \tag{3.C.31}$$

Furthermore, by Proposition 3.B.7,

$$\bar{X}_i = o_P(1). \tag{3.C.32}$$

Finally, consider the local linear kernel weights $w_{t,T}(u,h)$ defined in (3.3.5). Again, by construction the weights $w_{t,T}(u,h)$ are not equal to 0 if and only if $T(u-h) \le t \le T(u+h)$. We can use this fact to bound $\left|\sum_{t=1}^{T} w_{t,T}(u,h)\right|$ for all $(u,h) \in \mathcal{G}_T$ using the Cauchy-Schwarz inequality:

$$\begin{aligned} \left| \sum_{t=1}^{T} w_{t,T}(u,h) \right| &= \left| \sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} w_{t,T}(u,h) \cdot 1 \right| \\ &\le \sqrt{\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} w_{t,T}^2(u,h)} \sqrt{\sum_{t=\lfloor T(u-h)\rfloor}^{\lceil T(u+h)\rceil} 1^2} \\ &= \sqrt{1} \cdot \sqrt{D_{T,u,h}} \\ &\le \sqrt{2Th+2} \\ &\le \sqrt{2Th_{\max}+2} \end{aligned}$$

$$\leq \sqrt{T+2}.$$

Hence,

$$\max_{(u,h)\in\mathcal{G}_T} \Big| \sum_{t=1}^{T} w_{t,T}(u,h) \Big| = O(\sqrt{T}). \tag{3.C.33}$$

Combining (3.C.30), (3.C.31), (3.C.32) and (3.C.33), we get the following:

$$\begin{aligned} \max_{1\leq i<j\leq n} \left(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\right)^{-1/2} & \max_{1\leq i\leq n} \left|(\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top \bar{\mathbf{X}}_i\right| \max_{(u,h)\in\mathcal{G}_T} \Big| \sum_{t=1}^{T} w_{t,T}(u,h) \Big| \\ &= O_P(1) \cdot O_P(1/\sqrt{T}) \cdot o_P(1) \cdot O(\sqrt{T}) \\ &= o_P(1). \end{aligned} \tag{3.C.34}$$

Plugging (3.C.23), (3.C.29) and (3.C.34) in (3.C.19), we get that $\big|\widehat{\widehat{\Phi}}_{n,T} - \widehat{\Phi}_{n,T}\big| = o_P(1)$ and the statement of the theorem follows. □

Step 5

Proposition 3.C.13. *Under the conditions of Theorem 3.4.1, it holds that*

$$\sup_{x\in\mathbb{R}} \big|\mathbb{P}(\widehat{\Phi}_{n,T} \leq x) - \mathbb{P}(\Phi_{n,T} \leq x)\big| = o(1). \tag{3.C.35}$$

Proof of Proposition 3.C.13. First, we consider those $x \in \mathbb{R}$ such that $\mathbb{P}(\widehat{\Phi}_{n,T} \leq x) \geq \mathbb{P}(\Phi_{n,T} \leq x)$. Then by Proposition 3.C.11 for a sequence $\gamma_{n,T} > 0$ that satisfies the conditions of the Proposition 3.C.12 we have

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

Now consider such $x \in \mathbb{R}$ that $\mathbb{P}(\widehat{\Phi}_{n,T} \leq x) < \mathbb{P}(\Phi_{n,T} \leq x)$. Analogously,

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

Note that since $\gamma_{n,T} \to 0$, we can use the anti-concentration results (3.C.15) for the

Gaussian statistic $\Phi_{n,T}$: $\sup_{x\in\mathbb{R}}\mathbb{P}\Big(\big|\Phi_{n,T}-x\big|\le\gamma_{n,T}\Big)=o(1)$. Moreover,

$$\mathbb{P}\Big(\big|\widehat{\widehat{\Phi}}_{n,T}-\widehat{\Phi}_{n,T}\big|>\gamma_{n,T}\Big)=o(1)$$

by Proposition 3.C.12 and this probability does not depend on x.

Thus,

$$\begin{aligned}
&\sup_{x\in\mathbb{R}}\big|\mathbb{P}(\widehat{\Phi}_{n,T}\le x)-\mathbb{P}(\Phi_{n,T}\le x)\big|\le\\
&\qquad\le\max\Bigg\{\sup_{x\in\mathbb{R}}\bigg|\mathbb{P}\Big(\Phi_{n,T}\le x-\gamma_{n,T}\Big)-\mathbb{P}\Big(\widehat{\widehat{\Phi}}_{n,T}\le x-\gamma_{n,T}\Big)\bigg|,\\
&\qquad\qquad\sup_{x\in\mathbb{R}}\bigg|\mathbb{P}\Big(\Phi_{n,T}\le x+\gamma_{n,T}\Big)-\mathbb{P}\Big(\widehat{\widehat{\Phi}}_{n,T}\le x+\gamma_{n,T}\Big)\bigg|\Bigg\}+\\
&\qquad\qquad+\sup_{x\in\mathbb{R}}\mathbb{P}\Big(\big|\Phi_{n,T}-x\big|\le\gamma_{n,T}\Big)+\sup_{x\in\mathbb{R}}\mathbb{P}\Big(\big|\widehat{\widehat{\Phi}}_{n,T}-\widehat{\Phi}_{n,T}\big|>\gamma_{n,T}\Big)=\\
&\qquad=\sup_{y\in\mathbb{R}}\bigg|\mathbb{P}\Big(\Phi_{n,T}\le y\Big)-\mathbb{P}\Big(\widehat{\widehat{\Phi}}_{n,T}\le y\Big)\bigg|+o(1)+o(1)=o(1).
\end{aligned}$$

□

Proof of Proposition 3.4.2

Proof. To start with, note that for some constant C we have

$$\lambda(h)=\sqrt{2\log\{1/(2h)\}}\le\sqrt{2\log\{1/(2h_{\min})\}}\le C\sqrt{\log T}. \tag{3.C.36}$$

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

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

where $\bar{m}_{i,T}=T^{-1}\sum_{t=1}^{T}m_{i,T}(t/T)$.

Without loss of generality, consider the first scenario: 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

$$m_{i,T}(w)-m_{j,T}(w)\ge c_T\sqrt{\log T/(Th_0)} \tag{3.C.37}$$

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,

$$w_{t,T}(u_0, h_0) = \frac{K\Big(\frac{\frac{t}{T}-u_0}{h_0}\Big) S_{T,2}(u_0, h_0)}{\Big\{\sum_{t=1}^T K^2\Big(\frac{\frac{t}{T}-u_0}{h_0}\Big) S_{T,2}^2(u_0, h_0)\Big\}^{1/2}} \tag{3.C.38}$$

$$= \frac{K\Big(\frac{\frac{t}{T}-u_0}{h_0}\Big)}{\Big\{\sum_{t=1}^T K^2\Big(\frac{\frac{t}{T}-u_0}{h_0}\Big)\Big\}^{1/2}} \geq 0. \tag{3.C.39}$$

Together with (3.C.37), this implies that

$$\widehat{\psi}_{ij,T}^B(u_0, h_0) \geq c_T \sqrt{\frac{\log T}{T h_0}} \sum_{t=1}^T w_{t,T}(u_0, h_0). \tag{3.C.40}$$

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

$$\Big|\frac{1}{Th}\sum_{t=1}^T K\Big(\frac{\frac{t}{T}-u}{h}\Big)\Big(\frac{\frac{t}{T}-u}{h}\Big)^\ell - \int_0^1 \frac{1}{h} K\Big(\frac{w-u}{h}\Big)\Big(\frac{w-u}{h}\Big)^\ell dw\Big| \leq \frac{C}{Th}, \tag{3.C.41}$$

where the constant C does not depend on u, h and T. With the help of (3.C.41), we obtain that for any $(u, h) \in \mathcal{G}_T$ with $[u-h, u+h] \subseteq [0, 1]$,

$$\Big|\sum_{t=1}^T w_{t,T}(u, h) - \frac{\sqrt{Th}}{\kappa}\Big| \leq \frac{C}{\sqrt{Th}}, \tag{3.C.42}$$

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 (3.C.42), it follows that $\sum_{t=1}^T w_{t,T}(u, h) \geq \sqrt{Th}/(2\kappa)$ for sufficiently large T and any $(u, h) \in \mathcal{G}_T$ with $[u-h, u+h] \subseteq [0, 1]$. This together with (3.C.40) allows us to infer that

$$\widehat{\psi}_{ij,T}^B(u_0, h_0) \geq \frac{c_T \sqrt{\log T}}{2\kappa} \tag{3.C.43}$$

for sufficiently large T.

Furthermore, since $\widehat{\psi}_{ij,T}^A(u, h) = \widehat{\phi}_{ij,T}(u, h) + (\bar{m}_{j,T} - \bar{m}_{i,T}) \sum_{t=1}^T w_{t,T}(u, h)$, by the arguments completely analogous to those for the proof of Proposition 3.C.12, we have

$$\max_{(u,h)\in\mathcal{G}_T} \big|\widehat{\psi}_{ij,T}^A(u, h)\big| = O_p\big(\sqrt{\log T}\big). \tag{3.C.44}$$

With the help of (3.C.43), (3.C.44) and (3.C.36) and the assumption that $\widehat{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$, we finally arrive at

$$\widehat{\Psi}_T \geq \max_{1\leq i<j\leq n} \max_{(u,h)\in\mathcal{G}_T} \frac{|\widehat{\psi}_{ij,T}^B(u, h)|}{\{\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\}^{1/2}} - \max_{1\leq i<j\leq n} \max_{(u,h)\in\mathcal{G}_T} \left\{\frac{|\widehat{\psi}_{ij,T}^A(u, h)|}{\{\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\}^{1/2}} + \lambda(h)\right\}$$

$$= \max_{1\le i<j\le n} \max_{(u,h)\in\mathcal{G}_T} \frac{|\widehat{\psi}^B_{ij,T}(u,h)|}{\{\widehat{\sigma}_i^2+\widehat{\sigma}_j^2\}^{1/2}} + O_p(\sqrt{\log T})$$
$$\ge \frac{c_T\sqrt{\log T}}{2\kappa\{\widehat{\sigma}_i^2+\widehat{\sigma}_j^2\}^{1/2}} + O_p(\sqrt{\log T}). \tag{3.C.45}$$

Since $q_T(\alpha) = O(\sqrt{\log T})$ for any fixed $\alpha \in (0,1)$ and $c_T \to \infty$, (3.C.45) immediately implies that $\mathbb{P}(\widehat{\Psi}_T \le q_T(\alpha)) = o(1)$. □

Proof of Proposition 3.4.3

Proof. By Proposition 3.C.13, we have

$$\sup_{x\in\mathbb{R}} \big|\mathbb{P}(\widehat{\Phi}_{n,T} \le x) - \mathbb{P}(\Phi_{n,T} \le x)\big| = o(1). \tag{3.C.46}$$

By definition of the quantile $q_{n,T}(\alpha)$, it holds that $\mathbb{P}(\Phi_T \le q_{n,T}(\alpha)) \ge 1-\alpha$, which together with (3.C.46) immediately yields

$$\mathbb{P}\big(\widehat{\Phi}_T \le q_{n,T}(\alpha)\big) \ge 1-\alpha+o(1). \tag{3.C.47}$$

Now for the sake of simplifying notation, denote by $\mathcal{M}_0$ the set of quadruples $(i,j,u,h) \in \{1\dots,n\}^2 \times \mathcal{G}_T$ that has the property that $H_0^{[i,j]}(u,h)$ is true. Analogously, denote by $\mathcal{M}$ the full set of quadruples: $\mathcal{M} := \{1\dots,n\}^2 \times \mathcal{G}_T$. Then we can write FWER as

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

where the third equality holds true because under $H_0^{[i,j]}(u,h)$, $\hat{\psi}^0_{ijk,T} = \hat{\phi}^0_{ijk,T}$ by the observation in the beginning of Section 3.4. □

Proof of Corollary 3.4.1

Proof. By Proposition 3.4.3,

$$\begin{aligned}
1-\alpha+o(1) &\le 1-\mathrm{FWER}(\alpha)\\
&= \mathbb{P}\Big(\nexists(i,j,u,h)\in\mathcal{M}_0 : |\hat{\psi}^0_{ij,T}(u,h)| > q_{n,T}(\alpha)\Big)\\
&= \mathbb{P}\Big(\forall(i,j,u,h)\in\mathcal{M}_0 : |\hat{\psi}^0_{ij,T}(u,h)| \le q_{n,T}(\alpha)\Big)
\end{aligned}$$

$$= \mathbb{P}\Big(\forall\, i,j \in \{1,\dots,n\}, (u,h) \in \mathcal{G}_T \text{ such that } H_0^{[i,j]}(u,h) \text{ is true} : |\hat{\psi}^0_{ij,T}(u,h)| \leq q_{n,T}(\alpha)\Big),$$

which gives the statement of Corollary 3.4.1. □

3.D Asymptotic consistency of the estimators

3.D.1 Asymptotic consistency of $\widehat{\boldsymbol{\beta}}_i$

Before proceeding to the proof of Theorem 3.5.1, we first prove several auxiliary results. In order to do that, we define the first-differenced regressors as follows.

$$\Delta \mathbf{X}_{it} = \mathbf{H}_i(\mathcal{U}_{it}) - \mathbf{H}_i(\mathcal{U}_{it-1}) := \Delta \mathbf{H}_i(\mathcal{U}_{it}).$$

Similarly,

$$\Delta \varepsilon_{it} = \varepsilon_{it} - \varepsilon_{it-1} = G_i(\mathcal{J}_{it}) - G_i(\mathcal{J}_{it-1}) = \Delta G_i(\mathcal{J}_{it}).$$

We now can prove the following propositions.

Proposition 3.D.14. *Under Assumptions (C4) and (C6),* $||\Delta \mathbf{H}_i(\mathcal{U}_{it})||_4 < \infty$.

Proof of Proposition 3.D.14. By Assumption (C6) and the triangle inequality,

$$||\Delta \mathbf{H}_i(\mathcal{U}_{it})||_4 \leq ||\mathbf{H}_i(\mathcal{U}_{it})||_4 + ||\mathbf{H}_i(\mathcal{U}_{it-1})||_4 < \infty.$$

□

Proposition 3.D.15. *Under Assumption (C9),* $\Delta \mathbf{X}_{it}$ *(elementwise) and* $\Delta \varepsilon_{it}$ *are uncorrelated for each* $t \in \{1,\dots,T\}$.

Proof of Proposition 3.D.15. By Assumption (C9),

$$\begin{aligned}
\mathbb{E}[\Delta \mathbf{X}_{it} \Delta \varepsilon_{it}] &= \mathbb{E}\big[(\mathbf{X}_{it} - \mathbf{X}_{it-1})(\varepsilon_{it} - \varepsilon_{it-1})\big] \\
&= \mathbb{E}[\mathbf{X}_{it}\varepsilon_{it}] - \mathbb{E}[\mathbf{X}_{it-1}\varepsilon_{it}] - \mathbb{E}[\mathbf{X}_{it}\varepsilon_{it-1}] + \mathbb{E}[\mathbf{X}_{it-1}\varepsilon_{it-1}] \\
&= \mathbb{E}[\mathbf{X}_{it}]\mathbb{E}[\varepsilon_{it}] - \mathbb{E}[\mathbf{X}_{it-1}]\mathbb{E}[\varepsilon_{it}] - \mathbb{E}[\mathbf{X}_{it}]\mathbb{E}[\varepsilon_{it-1}] + \mathbb{E}[\mathbf{X}_{it-1}]\mathbb{E}[\varepsilon_{it-1}] \\
&= \big(\mathbb{E}[\mathbf{X}_{it}] - \mathbb{E}[\mathbf{X}_{it-1}]\big)\big(\mathbb{E}[\varepsilon_{it}] - \mathbb{E}[\varepsilon_{it-1}]\big) \\
&= \mathbb{E}[\Delta \mathbf{X}_{it}]\mathbb{E}[\Delta \varepsilon_{it}]
\end{aligned}$$

□

Proposition 3.D.16. *Define*

$$\Delta \mathbf{U}_i(\mathcal{I}_{it}) := \Delta \mathbf{H}_i(\mathcal{U}_{it}) \Delta G_i(\mathcal{J}_{it}).$$

Under Assumptions (C2), (C3), (C6), (C7) and (C10), we have that $\sum_{s=1}^{\infty} \delta_2(\Delta \mathbf{U}_i, s) < \infty$.

Proof of Proposition 3.D.16. By the triangle inequality and the definition of the physical dependence measure δ_2, we have that

$$
\begin{aligned}
&\delta_2(\Delta\mathbf{U}_i, t) = ||\Delta\mathbf{U}_i(\mathcal{I}_{it}) - \Delta\mathbf{U}_i(\mathcal{I}'_{it})|| \\
&= ||\Delta\mathbf{H}_i(\mathcal{U}_{it})\Delta G_i(\mathcal{J}_{it}) - \Delta\mathbf{H}_i(\mathcal{U}'_{it})\Delta G_i(\mathcal{J}'_{it})|| \\
&= ||\mathbf{H}_i(\mathcal{U}_{it})G_i(\mathcal{J}_{it}) - \mathbf{H}_i(\mathcal{U}_{it-1})G_i(\mathcal{J}_{it}) - \mathbf{H}_i(\mathcal{U}_{it})G_i(\mathcal{J}_{it-1}) + \mathbf{H}_i(\mathcal{U}_{it-1})G_i(\mathcal{J}_{it-1}) \\
&\quad - \mathbf{H}_i(\mathcal{U}'_{it})G_i(\mathcal{J}'_{it}) + \mathbf{H}_i(\mathcal{U}'_{it-1})G_i(\mathcal{J}'_{it}) + \mathbf{H}_i(\mathcal{U}'_{it})G_i(\mathcal{J}'_{it-1}) - \mathbf{H}_i(\mathcal{U}'_{it-1})G_i(\mathcal{J}'_{it-1})|| \\
&\leq ||\mathbf{H}_i(\mathcal{U}_{it})G_i(\mathcal{J}_{it}) - \mathbf{H}_i(\mathcal{U}'_{it})G_i(\mathcal{J}'_{it})|| + ||\mathbf{H}_i(\mathcal{U}_{it-1})G_i(\mathcal{J}_{it-1}) - \mathbf{H}_i(\mathcal{U}'_{it-1})G_i(\mathcal{J}'_{it-1})|| \\
&\quad + ||\mathbf{H}_i(\mathcal{U}_{it-1})G_i(\mathcal{J}_{it}) - \mathbf{H}_i(\mathcal{U}'_{it-1})G_i(\mathcal{J}'_{it})|| \\
&\quad + ||\mathbf{H}_i(\mathcal{U}_{it})G_i(\mathcal{J}_{it-1}) - \mathbf{H}_i(\mathcal{U}'_{it})G_i(\mathcal{J}'_{it-1})|| \\
&= \delta_2(\mathbf{U}_i, t) + \delta_2(\mathbf{U}_i, t-1) \\
&\quad + ||\mathbf{H}_i(\mathcal{U}_{it-1})G_i(\mathcal{J}_{it}) - \mathbf{H}_i(\mathcal{U}'_{it-1})G_i(\mathcal{J}_{it}) + \mathbf{H}_i(\mathcal{U}'_{it-1})G_i(\mathcal{J}_{it}) - \mathbf{H}_i(\mathcal{U}'_{it-1})G_i(\mathcal{J}'_{it})|| \\
&\quad + ||\mathbf{H}_i(\mathcal{U}_{it})G_i(\mathcal{J}_{it-1}) - \mathbf{H}_i(\mathcal{U}'_{it})G_i(\mathcal{J}_{it-1}) + \mathbf{H}_i(\mathcal{U}'_{it})G_i(\mathcal{J}_{it-1}) - \mathbf{H}_i(\mathcal{U}'_{it})G_i(\mathcal{J}'_{it-1})|| \\
&\leq \delta_2(\mathbf{U}_i, t) + \delta_2(\mathbf{U}_i, t-1) \\
&\quad + ||\big(\mathbf{H}_i(\mathcal{U}_{it-1}) - \mathbf{H}_i(\mathcal{U}'_{it-1})\big)G_i(\mathcal{J}_{it})|| + ||\mathbf{H}_i(\mathcal{U}'_{it-1})\big(G_i(\mathcal{J}_{it}) - G_i(\mathcal{J}'_{it})\big)|| \\
&\quad + ||\big(\mathbf{H}_i(\mathcal{U}_{it}) - \mathbf{H}_i(\mathcal{U}'_{it})\big)G_i(\mathcal{J}_{it-1})|| + ||\mathbf{H}_i(\mathcal{U}'_{it})\big(G_i(\mathcal{J}_{it-1}) - G_i(\mathcal{J}'_{it-1})\big)|| \\
&\leq \delta_2(\mathbf{U}_i, t) + \delta_2(\mathbf{U}_i, t-1) \\
&\quad + \big(\delta_2(\mathbf{H}_i, t-1) + \delta_2(\mathbf{H}_i, t)\big)||G_i|| + \big(\delta_2(G_i, t-1) + \delta_2(G_i, t)\big)||\mathbf{H}_i||.
\end{aligned}
$$

Here $\mathcal{U}'_{it} = (\ldots, u_{i(-1)}, u'_{i0}, u_{i1}, \ldots, u_{it-1}, u_{it})$, $\mathcal{U}'_{i(t-1)} = (\ldots, u_{i(-1)}, u'_{i0}, u_{i1}, \ldots, u_{it-1})$, $\mathcal{J}'_{it} = (\ldots, \eta_{i(-1)}, \eta'_{i0}, \eta_{i1}, \ldots, \eta_{it-1}, \eta_{it})$, $\mathcal{J}'_{i(t-1)} = (\ldots, \eta_{i(-1)}, \eta'_{i0}, \eta_{i1}, \ldots, \eta_{it-1})$ are coupled processes with u'_{i0} being an i.i.d. copy of u_{i0} and η'_{i0} being an i.i.d. copy of η_{i0}.

Therefore,

$$
\begin{aligned}
\sum_{s=1}^{\infty}\delta_2(\Delta\mathbf{U}_i, s) &\leq \sum_{s=0}^{\infty}\delta_2(\mathbf{U}_i, s) + \sum_{s=1}^{\infty}\delta_2(\mathbf{U}_i, s-1) \\
&\quad + \sum_{s=1}^{\infty}\big(\delta_2(\mathbf{H}_i, s-1) + \delta_2(\mathbf{H}_i, s)\big)||G_i|| + \sum_{s=1}^{\infty}\big(\delta_2(G_i, s-1) + \delta_2(G_i, s)\big)||\mathbf{H}_i||.
\end{aligned}
$$

By Assumptions (C2), (C3), (C6), (C7) and (C10), the RHS is finite. Statement of the proposition follows. □

Proposition 3.D.17. *Under Assumptions (C1) - (C10),*

$$
\Big|\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}\Big| = O_P(1).
$$

Proof of Proposition 3.D.17. For this proof, we will need the following notation:

$$
\mathcal{P}_{i,t}(\cdot) := \mathbb{E}[\cdot|\mathcal{I}_{it}] - \mathbb{E}[\cdot|\mathcal{I}_{i,t-1}],
$$

$$\begin{aligned}\kappa_i &:= \frac{1}{T-1}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta\varepsilon_{it},\\ \kappa_{i,s}^{\mathcal{P}} &:= \frac{1}{T-1}\sum_{t=2}^{T}\mathcal{P}_{i,t-s}\big(\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}\big).\end{aligned}$$

Since $\mathcal{P}_{i,t}(\cdot)$ is a projection operator, we have that

$$\begin{aligned}||\kappa_{i,s}^{\mathcal{P}}||^2 &\le \frac{1}{(T-1)^2}\sum_{t=2}^{T}\Big|\Big|\mathcal{P}_{i,t-s}\big(\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}\big)\Big|\Big|^2\\ &= \frac{1}{(T-1)^2}\sum_{t=2}^{T}\Big|\Big|\mathbb{E}\big(\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}|\mathcal{I}_{i(t-s)}\big)-\mathbb{E}\big(\Delta\mathbf{X}_{it,s}\Delta\varepsilon_{it,s}|\mathcal{I}_{i(t-s-1)}\big)\Big|\Big|^2\\ &= \frac{1}{(T-1)^2}\sum_{t=2}^{T}\Big|\Big|\mathbb{E}\big(\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}|\mathcal{I}_{i(t-s)}\big)-\mathbb{E}\big(\Delta\mathbf{X}'_{it,s}\Delta\varepsilon'_{it,s}|\mathcal{I}_{i(t-s)}\big)\Big|\Big|^2,\end{aligned}$$

where $\Delta\mathbf{X}'_{it,s}\Delta\varepsilon'_{it,s}$ denotes $\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}$ with $\{\zeta_{i,t-s}\}$ replaced by its i.i.d. copy $\{\zeta'_{i,t-s}\}$. In this case $\mathbb{E}\big(\Delta\mathbf{X}'_{it,s}\Delta\varepsilon'_{it,s}|\mathcal{I}_{i(t-s-1)}\big)=\mathbb{E}\big(\Delta\mathbf{X}'_{it,s}\Delta\varepsilon'_{it,s}|\mathcal{I}_{i(t-s)}\big)$. Furthermore, by linearity of the expectation and Jensen's inequality, we have

$$\begin{aligned}||\kappa_{i,s}^{\mathcal{P}}||^2 &\le \frac{1}{(T-1)^2}\sum_{t=2}^{T}\Big|\Big|\mathbb{E}(\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}|\mathcal{I}_{i(t-s)})-\mathbb{E}(\Delta\mathbf{X}'_{it,s}\Delta\varepsilon'_{it,s}|\mathcal{I}_{i(t-s)})\Big|\Big|^2\\ &\le \frac{1}{(T-1)^2}\sum_{t=2}^{T}\Big|\Big|\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}-\Delta\mathbf{X}'_{it,s}\Delta\varepsilon'_{it,s}\Big|\Big|^2\\ &= \frac{1}{(T-1)^2}\sum_{t=2}^{T}\Big|\Big|\Delta\mathbf{H}_i(\mathcal{U}_{it})\Delta G_i(\mathcal{J}_{it})-\Delta\mathbf{H}_i(\mathcal{U}'_{it,s})\Delta G_i(\mathcal{J}'_{it,s})\Big|\Big|^2\\ &= \frac{1}{(T-1)^2}\sum_{t=2}^{T}\Big|\Big|\Delta\mathbf{U}_i(\mathcal{I}_{it})-\Delta\mathbf{U}_i(\mathcal{I}'_{it,s})\Big|\Big|^2\\ &\le \frac{1}{(T-1)^2}\sum_{t=2}^{T}\delta_2^2(\Delta\mathbf{U}_i,s)\\ &= \frac{1}{T-1}\delta_2^2(\Delta\mathbf{U}_i,s)\end{aligned}$$

with $\mathcal{U}'_{it,s}=(\ldots,u_{i(t-s-1)},u'_{i(t-s)},u_{i(t-s+1)},\ldots,u_{it})$, $u'_{i(t-s)}$ being an i.i.d. copy of $u_{i(t-s)}$, $\mathcal{J}'_{it,s}=(\ldots,\eta_{i(t-s-1)},\eta'_{i(t-s)},\eta_{i(t-s+1)},\ldots,\eta_{it})$, $\eta'_{i(t-s)}$ being an i.i.d. copy of $\eta_{i(t-s)}$, and $\zeta'_{it}=(u'_{it},\eta'_{it})^\top$ and $\mathcal{I}'_{i,t,s}=(\ldots,\zeta_{i(t-s-1)},\zeta'_{i(t-s)},\zeta_{i(t-s+1)},\ldots,\zeta_{it})$. Moreover,

$$\begin{aligned}\kappa_i-\mathbb{E}\kappa_i &= \frac{1}{T-1}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}-\mathbb{E}\kappa_i\\ &= \frac{1}{T-1}\sum_{t=2}^{T}\mathbb{E}(\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}|\mathcal{I}_{it})-\mathbb{E}\kappa_i\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{T-1}\sum_{t=2}^{T}\big(\mathbb{E}(\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}|\mathcal{I}_{it}) - \mathbb{E}(\mathbf{X}_{it}\Delta\varepsilon_{it})\big) \\
&= \frac{1}{T-1}\sum_{t=2}^{T}\sum_{s=0}^{\infty}\big(\mathbb{E}(\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}|\mathcal{I}_{i(t-s)}) - \mathbb{E}(\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}|\mathcal{I}_{i(t-s-1)})\big) \\
&= \frac{1}{T-1}\sum_{t=2}^{T}\sum_{s=0}^{\infty}\mathcal{P}_{i,t-s}(\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}) = \sum_{s=0}^{\infty}\kappa_{i,s}^{\mathcal{P}}.
\end{aligned}
$$

Thus, by Proposition 3.D.16,

$$
||\kappa_i - \mathbb{E}\kappa_i|| \leq \sum_{s=0}^{\infty}||\kappa_{i,s}^{\mathcal{P}}|| \leq \frac{1}{\sqrt{T-1}}\sum_{s=0}^{\infty}\delta_2(\Delta\mathbf{U}_i, s) = O\left(\frac{1}{\sqrt{T}}\right)
$$

Since $\mathbb{E}\kappa_i = 0$ by Proposition 3.D.15, we conclude that

$$
\Big|\Big|\frac{1}{T}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}\Big|\Big| = O\left(\frac{1}{\sqrt{T}}\right).
$$

Therefore, the proposition follows. □

Proof of Theorem 3.5.1. Before we begin, we need to introduce some additional notation that we will use throughout the proof. First, define $\Delta m_{it} = m_i\left(\frac{t}{T}\right) - m_i\left(\frac{t-1}{T}\right)$. Then, by Assumption (C4), we can rewrite the first-differenced regressors $\Delta\mathbf{X}_{it}$ as

$$
\Delta\mathbf{X}_{it} = \mathbf{H}_i(\mathcal{U}_{it}) - \mathbf{H}_i(\mathcal{U}_{it-1}) := \Delta\mathbf{H}_i(\mathcal{U}_{it})
$$

with $\Delta\mathbf{H}_i(\mathcal{U}_{it}) := (\Delta H_{i1}, \Delta H_{i2}, \ldots, \Delta H_{id})^\top$.

Similarly, by Assumption (C1), we have

$$
\Delta\varepsilon_{it} = \varepsilon_{it} - \varepsilon_{it-1} = G_i(\mathcal{J}_{it}) - G_i(\mathcal{J}_{it-1}) = \Delta G_i(\mathcal{J}_{it}).
$$

Then, the differencing estimator $\widehat{\boldsymbol{\beta}}_i$ can be written as

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

Hence,

$$\begin{aligned}\sqrt{T}(\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i) =& \Big(\frac{1}{T}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^{\top}\Big)^{-1}\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta m_{it} \\ &+ \Big(\frac{1}{T}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^{\top}\Big)^{-1}\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}.\end{aligned} \tag{3.D.1}$$

We look at the parts that constitute (3.D.1) independently and for clarification purposes, we break the proof into three steps.

For the sake of simplicity, we focus our attention on the individual vector components and we prove the necessary bounds and inequalities for each of the components separately, combining them together in the end.

Step 1.

First, we take a closer look at the part of the first summand in (3.D.1), specifically, $\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta m_{it}$.

Fix $j \in 1,\dots,d$. By Chebyshev's inequality, for any $a > 0$ we have

$$\mathbb{P}\left(\frac{1}{T}\sum_{t=2}^{T}|\Delta H_{ij}(\mathcal{U}_{it})| > a\right) \leq \frac{\mathbb{E}\Big[\big(\sum_{t=2}^{T}|\Delta H_{ij}(\mathcal{U}_{it})|\big)^2\Big]}{(T-1)^2a^2} \tag{3.D.2}$$

and

$$\mathbb{E}\Big[\big(\sum_{t=2}^{T}|\Delta H_{ij}(\mathcal{U}_{it})|\big)^2\Big] = \sum_{t=2}^{T}\mathbb{E}\left[\Delta H_{ij}^2(\mathcal{U}_{it})\right] + \sum_{\substack{t=2,s=2,\\ t\neq s}}^{T}\mathbb{E}\big[\,|\Delta H_{ij}(\mathcal{U}_{it})\Delta H_{ij}(\mathcal{U}_{is})|\,\big]. \tag{3.D.3}$$

Note that by the Cauchy-Schwarz inequality for all t and s we have

$$\mathbb{E}\big[\,|H_{ij}(\mathcal{U}_{it})H_{ij}(\mathcal{U}_{is})|\,\big] \leq \sqrt{\mathbb{E}\big[H_{ij}^2(\mathcal{U}_{it})\big]}\sqrt{\mathbb{E}\big[H_{ij}^2(\mathcal{U}_{is})\big]} = \mathbb{E}\left[H_{ij}^2(\mathcal{U}_{i0})\right] \tag{3.D.4}$$

and

$$\big|\mathbb{E}\big[H_{ij}(\mathcal{U}_{it})H_{ij}(\mathcal{U}_{is})\big]\big| \leq \mathbb{E}\big[\,|H_{ij}(\mathcal{U}_{it})H_{ij}(\mathcal{U}_{is})|\,\big] \leq \mathbb{E}\left[H_{ij}^2(\mathcal{U}_{i0})\right].$$

Hence,

$$\begin{aligned}\mathbb{E}\left[\Delta H_{ij}^2(\mathcal{U}_{it})\right] &= \mathbb{E}\left[H_{ij}^2(\mathcal{U}_{it})\right] - 2\mathbb{E}\left[H_{ij}(\mathcal{U}_{it})H_{ij}(\mathcal{U}_{it-1})\right] + \mathbb{E}\left[H_{ij}^2(\mathcal{U}_{it-1})\right] \\ &\leq \mathbb{E}\left[H_{ij}^2(\mathcal{U}_{i0})\right] + 2\mathbb{E}\left[H_{ij}^2(\mathcal{U}_{i0})\right] + \mathbb{E}\left[H_{ij}^2(\mathcal{U}_{i0})\right] \\ &= 4\mathbb{E}\left[H_{ij}^2(\mathcal{U}_{i0})\right]\end{aligned}$$

and the first summand in (3.D.3) can be bounded by $4(T-1)\mathbb{E}\left[H_{ij}^2(\mathcal{U}_{i0})\right]$, where the expectation is finite due to Assumption (C6).

Now to the second summand in (3.D.3):

$$\begin{aligned}\mathbb{E}\big[\,|\Delta H_{ij}(\mathcal{U}_{it})\Delta H_{ij}(\mathcal{U}_{is})|\,\big] &\le \mathbb{E}\big[\,|H_{ij}(\mathcal{U}_{it})H_{ij}(\mathcal{U}_{is})|\,\big] + \mathbb{E}\big[\,|H_{ij}(\mathcal{U}_{it-1})H_{ij}(\mathcal{U}_{is})|\,\big] \\ &\quad + \mathbb{E}\big[\,|H_{ij}(\mathcal{U}_{it})H_{ij}(\mathcal{U}_{is-1})|\,\big] + \mathbb{E}\big[\,|H_{ij}(\mathcal{U}_{it-1})H_{ij}(\mathcal{U}_{is-1})|\,\big] \\ &\le 4\mathbb{E}\left[H_{ij}^2(\mathcal{U}_{i0})\right],\end{aligned}$$

where in the last inequality we used (3.D.4). This means that the second summand in (3.D.3) can be bounded by $4(T-1)(T-2)\mathbb{E}\left[H_{ij}^2(\mathcal{U}_{i0})\right]$.

Plugging these bounds in (3.D.3), we get

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{t=2}^{T}\big|\Delta H_{ij}(\mathcal{U}_{it})\big|\right)^2\right] &\le 4(T-1)\mathbb{E}\left[H_{ij}^2(\mathcal{U}_{i0})\right] + 4(T-1)(T-2)\mathbb{E}\left[H_{ij}^2(\mathcal{U}_{i0})\right] \\ &= 4(T-1)^2\mathbb{E}\left[H_{ij}^2(\mathcal{U}_{i0})\right],\end{aligned}$$

which together with (3.D.2) leads to $\frac{1}{T}\sum_{t=2}^{T}\big|\Delta H_{ij}(\mathcal{U}_{it})\big| = O_P(1)$.

Next, by the assumption in Theorem 3.5.1, $m_i(\cdot)$ is Lipschitz continuous, that is, $|\Delta m_{it}| = \big|m_i\left(\frac{t}{T}\right) - m_i\left(\frac{t-1}{T}\right)\big| \le C\frac{1}{T}$ for all $t \in \{1,\dots,T\}$ and some constant $C > 0$. Hence,

$$\begin{aligned}\Big|\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\Delta H_{ij}(\mathcal{U}_{it})\Delta m_{it}\Big| &\le \frac{1}{\sqrt{T}}\sum_{t=2}^{T}\big|\Delta H_{ij}(\mathcal{U}_{it})\big|\cdot|\Delta m_{it}| \\ &\le \frac{C}{\sqrt{T}}\cdot\frac{1}{T}\sum_{t=2}^{T}|\Delta H_{ij}(\mathcal{U}_{it})| \\ &= O_P\Big(\frac{1}{\sqrt{T}}\Big).\end{aligned}$$

Since it holds for each $j \in \{1,\dots,d\}$ (and d is fixed), it is obvious that

$$\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta m_{it} = \frac{1}{\sqrt{T}}\sum_{t=2}^{T}\Delta\mathbf{H}_i(\mathcal{U}_{it})\Delta m_{it} = O_P\Big(\frac{1}{\sqrt{T}}\Big). \tag{3.D.5}$$

Step 2.

Now we look at the other part of the first summand in (3.D.1), specifically, $\big(\frac{1}{T}\sum_{t=2}^{T}\Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^\top\big)^{-1}$. Using similar arguments as in Step 1 and applying Proposition 3.D.14, we can show that

$$\Big|\frac{1}{T}\sum_{t=2}^{T}\Delta H_{ij}(\mathcal{U}_{it})\Delta H_{ik}(\mathcal{U}_{it})\Big| = O_P(1),$$

for each $j, k \in \{1, \dots, d\}$, which trivially leads to

$$\Big|\frac{1}{T}\sum_{t=2}^{T} \Delta\mathbf{H}_i(\mathcal{U}_{it})\Delta\mathbf{H}_i(\mathcal{U}_{it})^\top\Big| = \Big|\frac{1}{T}\sum_{t=2}^{T} \Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^\top\Big| = O_P(1),$$

where $|A|$ with A being a matrix is any matrix norm.

Furthermore, by Assumption (C5), we know that $\mathbb{E}[\Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^\top] = \mathbb{E}[\Delta\mathbf{X}_{i0}\Delta\mathbf{X}_{i0}^\top]$ is invertible, thus,

$$\left|\Big(\frac{1}{T}\sum_{t=2}^{T} \Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^\top\Big)^{-1}\right| = O_P(1). \tag{3.D.6}$$

Step 3

Here we turn our attention to the second summand in (3.D.1). We already know that $\Big|\big(\frac{1}{T}\sum_{t=2}^{T} \Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^\top\big)^{-1}\Big| = O_P(1)$. Moreover, by Proposition 3.D.17,

$$\left|\frac{1}{\sqrt{T}}\sum_{t=2}^{T} \Delta\mathbf{X}_{it}\Delta\varepsilon_{it}\right| = O_P(1).$$

Taking these two facts together, we have that

$$\Big(\frac{1}{T}\sum_{t=2}^{T} \Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^\top\Big)^{-1}\frac{1}{\sqrt{T}}\sum_{t=2}^{T} \Delta\mathbf{X}_{it}\Delta\varepsilon_{it} = O_P(1). \tag{3.D.7}$$

Finally, from (3.D.5) and (3.D.6) we get that the first summand in (3.D.1) is $O_P(1/\sqrt{T})$, and by (3.D.7) the second summand is $O_P(1)$. The statement of the theorem follows. □

3.D.2 Asymptotic consistency of $\widehat{\sigma}_i^2$

Lemma 3.D.11. *Let $s_T \asymp T^{1/3}$. Then, under Assumptions (C1) - (C10), for each $i \in \{1, \dots, n\}$ we have*

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

where $\widehat{\sigma}_i^2$ is the subseries variance estimate of σ_i^2 introduced by (3.5.4).

Proof of Lemma 3.D.11. For notational convenience, we let $Y_{it}^* = Y_{it} - \boldsymbol{\beta}_i^\top X_{it}$. Note that

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

where Y_{it}° is the dependent variable in a well-studied standard nonparametric regression discussed in Section 3.3.1.

Now, using simple arithmetic calculations, we can rewrite $\widehat{\sigma}_i^2$ as

$$\begin{aligned}\widehat{\sigma}_i^2 = & \frac{1}{2(M-1)s_T}\sum_{m=1}^{M}\left[\sum_{t=1}^{s_T}\left(Y_{i(t+ms_T)}^{\circ}-Y_{i(t+(m-1)s_T)}^{\circ}\right)\right]^2 \\ & + \frac{1}{2(M-1)s_T}\sum_{m=1}^{M}\left[\sum_{t=1}^{s_T}(\widehat{\boldsymbol{\beta}}_i-\boldsymbol{\beta}_i)^{\top}\left(X_{i(t+ms_T)}-X_{i(t+(m-1)s_T)}\right)\right]^2 \\ & - \frac{1}{(M-1)s_T}\sum_{m=1}^{M}\left[\sum_{t=1}^{s_T}\left(Y_{i(t+ms_T)}^{\circ}-Y_{i(t+(m-1)s_T)}^{\circ}\right)\sum_{t=1}^{s_T}(\widehat{\boldsymbol{\beta}}_i-\boldsymbol{\beta}_i)^{\top}\left(X_{i(t+ms_T)}-X_{i(t+(m-1)s_T)}\right)\right],\end{aligned} \tag{3.D.8}$$

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

$$\frac{1}{2(M-1)s_T}\sum_{m=1}^{M}\left[\sum_{t=1}^{s_T}\left(Y_{i(t+ms_T)}^{\circ}-Y_{i(t+(m-1)s_T)}^{\circ}\right)\right]^2 = \sigma_i^2 + O_P(T^{-1/3}). \tag{3.D.9}$$

Furthermore, by our assumption that $s_T \asymp T^{1/3}$, Assumption (C5) and Theorem 3.5.1, we have

$$\frac{1}{2(M-1)s_T}\sum_{m=1}^{M}\left[\sum_{t=1}^{s_T}(\widehat{\boldsymbol{\beta}}_i-\boldsymbol{\beta}_i)^{\top}\left(X_{i(t+ms_T)}-X_{i(t+(m-1)s_T)}\right)\right]^2 = O_P(T^{-2/3}). \tag{3.D.10}$$

Finally, applying (3.D.9) and (3.D.10) together with the Cauchy-Schwarz inequality, we obtain

$$\frac{1}{(M-1)s_T}\sum_{m=1}^{M}\left[\sum_{t=1}^{s_T}\left(Y_{i(t+ms_T)}^{\circ}-Y_{i(t+(m-1)s_T)}^{\circ}\right)\sum_{t=1}^{s_T}(\widehat{\boldsymbol{\beta}}_i-\boldsymbol{\beta}_i)^{\top}\left(X_{i(t+ms_T)}-X_{i(t+(m-1)s_T)}\right)\right] = O_P(T^{-1/3}). \tag{3.D.11}$$

Applying (3.D.9) - (3.D.11) to (3.D.8), the lemma trivially follows. □

Bibliography

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

Bakirov, N. K., Rizzo, M. L. and Székely, G. J. (2006). A multivariate nonparametric test of independence. *Journal of multivariate analysis*, **97** 1742–1756.

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.

Blum, J. R., Kiefer, J. and Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *Annals of Mathematical Statistics*, **32** 485–498.

Brockwell, P. J. and Davis, R. A. (1991). *Time series: theory and methods*. New York, Springer.

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

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

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. (2014). Gaussian approximation of suprema of empirical processes. *Annals of Statistics*, **42** 1564–1597.

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

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

Cho, H. and Fryzlewicz, P. (2012). Multiscale and multilevel technique for consistent segmentation of nonstationary time series. *Statistica Sinica*, **22** 207–229.

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

Cohen, J. and Kupferschmidt, K. (2020). Countries test tactics in 'war' against COVID-19. *Science*, **367** 1287–1288.

Cox, D. R. (1983). Some remarks on overdispersion. *Biometrika*, **70** 269–274.

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

De Salazar, P. M., Niehus, R., Taylor, A., Buckee, C. and Lipsitch, M. (2020). Using predicted imports of 2019-nCoV cases to determine locations that may not be identifying all imported cases. *medRxiv*.

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

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

Donoho, D., Johnstone, I., Kerkyacharian, G. and Picard, D. (1995). Wavelet shrinkage: Asymptopia? *Journal of the Royal Statistical Society: Series B*, **57** 301–369.

Donoho, D. L. and Johnstone, I. M. (1995). Adapting to unknown smoothness via Wavelet shrinkage. *Journal of the American Statistical Association*, **90** 1200–1224.

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.

Dümbgen, L. and Walther, G. (2008). Multiscale inference about a density. *Annals of Statistics*, **36** 1758–1785.

Dunker, F., Eckle, K., Proksch, K. and Schmidt-Hieber, J. (2019). Tests for qualitative features in the random coefficients model. *Electronic Journal of Statistics*, **13** 2257–2306.

Eckle, K., Bissantz, N. and Dette, H. (2017). Multiscale inference for multivariate deconvolution. *Electronic Journal of Statistics*, **11** 4179–4219.

Efron, B. (1986). Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association*, **81** 709–721.

Fryzlewicz, P., Sapatinas, T. and Subba Rao, S. (2006). A Haar-Fisz technique for locally stationary volatility estimation. *Biometrika*, **93** 687–704.

Fryzlewicz, P., Sapatinas, T. and Subba Rao, S. (2008). Normalized least-squares estimation in time-varying ARCH models. *Annals of Statistics*, **36** 742–786.

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

Hafner, C. M. and Linton, O. (2010). Efficient estimation of a multivariate multiplicative volatility model. *Journal of Econometrics*, **159** 55–73.

Hale, T., Petherick, A., Phillips, T. and Webster, S. (2020a). Variation in government responses to COVID-19. *Blavatnik school of government working paper*, **31**.

Hale, T., Webster, S., Petherick, A., Phillips, T. and Kira, B. (2020b). Oxford COVID-19 government response tracker. Blavatnik school of government. `http://www.bsg.ox.ac.uk/covidtracker`.

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

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.

Hannig, J. and Marron, J. S. (2006). Advanced distribution theory for SiZer. *Journal of the American Statistical Association*, **101** 484–499.

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

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

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

Inselberg, A. (1985). The plane with parallel coordinates. *The Visual Computer*, **1** 69–91.

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

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

Khismatullina, M. and Vogt, M. (2021). Nonparametric comparison of epidemic time trends: the case of covid-19. *arXiv preprint arXiv:2007.15931*.

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

King, E. C., Hart, J. D. and Wehrly, T. E. (1991). Testing the equality of regression curves using linear smoothers. *Statistics & Probability Letters*, **12** 239–247.

Kulasekera, K. B. (1995). Comparison of regression curves using quasi-residuals. *Journal of the American Statistical Association*, **90** 1085–1093.

Lavergne, P. (2001). An equality test across nonparametric regressions. *Journal of Econometrics*, **103** 307–344.

Ledoux, M. (2001). *Concentration of Measure Phenomenon*. American Mathematical Society.

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

Marron, J. S., Adak, S., Johnstone, I. M., Neumann, M. and Patil, P. (1998). Exact risk analysis of Wavelet regression. *Journal of Computational and Graphical Statistics*, **7** 278–309.

McCullagh, P. and Nelder, J. (1989). *Generalized linear models*. Chapman and Hall.

Mikosch, T. and Stărică, C. (2000). Is it really long memory we see in financial returns? In *Extremes and Integrated Risk Management* (P. Embrechts, ed.). 149–168.

Mikosch, T. and Stărică, C. (2004). Non-stationarities in financial time series, the long-range dependence, and IGARCH effects. *The Review of Economics and Statistics*, **86** 378–390.

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.

Munk, A. and Dette, H. (1998). Nonparametric comparison of several regression functions: exact and asymptotic theory. *Annals of Statistics*, **26** 2339–2368.

Nazarov, F. (2003). On the maximal perimeter of a convex set in $\mathbb{R}^n$ with respect to a Gaussian measure. In *Geometric Aspects of Functional Analysis*, vol. 1807 of *Lecture Notes in Mathematics*. Springer, 169–187.

Neumeyer, N. and Dette, H. (2003). Nonparametric comparison of regression curves: an empirical process approach. *Annals of Statistics*, **31** 880–920.

Pardo-Fernández, J. C., Van Keilegom, I. and González-Manteiga, W. (2007). Testing for the equality of k regression curves. *Statistica Sinica*, **17** 1115–1137.

Park, C., Hannig, J. and Kang, K.-H. (2009a). Improved SiZer for time series. *Statistica Sinica*, **19** 1511–1530.

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. (2009b). 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.

Pellis, L., Scarabel, F., Stage, H. B., Overton, C. E., Chappell, L. H., Lythgoe, K. A., Fearon, E., Bennett, E., Curran-Sebastian, J., Das, R. et al. (2020). Challenges in control of Covid-19: short doubling time and long delay to effect of interventions. *arXiv:2004.00117*.

Proksch, K., Werner, F. and Munk, A. (2018). Multiscale scanning in inverse problems. *Forthcoming in Annals of Statistics*.

Qiu, D., Shao, Q. and Yang, L. (2013). Efficient inference for autoregressive coefficients in the presence of trends. *Journal of Multivariate Analysis*, **114** 40–53.

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

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

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

Rohde, A. (2008). Adaptive goodness-of-fit tests based on signed ranks. *Annals of Statistics*, **36** 1346–1374.

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.

Rufibach, K. and Walther, G. (2010). The block criterion for multiscale inference about a density, with applications to other multiscale problems. *Journal of Computational and Graphical Statistics*, **19** 175–190.

Schmidt-Hieber, J., Munk, A. and Dümbgen, L. (2013). Multiscale methods for shape constraints in deconvolution: confidence statements for qualitative features. *Annals of Statistics*, **41** 1299–1328.

Shao, Q. and Yang, L. J. (2011). Autoregressive coefficient estimation in nonparametric analysis. *Journal of Time Series Analysis*, **32** 587–597.

Sinha, B. K. and Wieand, H. S. (1977). Multivariate nonparametric tests for independence. *Journal of Multivariate Analysis*, **7** 572–583.

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

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.

Tobías, A., Valls, J., Satorra, P. and Tebé, C. (2020). COVID19-Tracker: A shiny app to produce comprehensive data visualization for SARS-CoV-2 epidemic in Spain. *medRxiv*.

Truong, Y. K. (1991). Nonparametric curve estimation with time series errors. *Journal of Statistical Planning and Inference*, **28** 167–183.

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

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

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

Vogt, M. and Linton, O. (2020). Multiscale clustering of nonparametric regression curves. *Journal of Econometrics*, **216** 305–325.

Von Sachs, R. and MacGibbon, B. (2000). Non-parametric curve estimation by Wavelet thresholding with locally stationary errors. *Scandinavian Journal of Statistics*, **27** 475–499.

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

Wu, W. B. (2007). Strong invariance principles for dependent random variables. *Annals of Probability*, **35** 2294–2320.

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

Wu, W. B., Woodroofe, M. and Mentz, G. (2001). Isotonic regression: another look at the changepoint problem. *Biometrika*, **88** 793–804.

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

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

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

Young, S. G. and Bowman, A. W. (1995). Nonparametric analysis of covariance. *Biometrics*, **51** 920–931.

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