

Simultaneous statistical inference for epidemic trends

Marina Khismatullina
University of Bonn

Michael Vogt
University of Bonn

1 Introduction

There are many questions surrounding the current COVID-19 pandemic that are not well understood yet. One important question is the following: How does the outbreak pattern of COVID-19 compare across countries? Are the time trends of daily new infections comparable across countries, or is the virus spreading differently in different regions of the world? The main aim of this paper is to develop new statistical methods that help to shed light on this issue.

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 . 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 $\lambda_i(t/T) = \mathbb{E}[X_{it}] = \text{Var}(X_{it})$, we can model the observations X_{it} by the nonparametric regression equation

$$X_{it} = \lambda_i\left(\frac{t}{T}\right) + u_{it} \quad (1.1)$$

for $1 \leq t \leq T$, where $u_{it} = X_{it} - \mathbb{E}[X_{it}]$ with $\mathbb{E}[u_{it}] = 0$ and $\text{Var}(u_{it}) = \lambda_i(t/T)$. As usual in nonparametric regression (cp. Robinson, 1989), we let the regression function λ_i in model (1.1) depend on rescaled time t/T rather than on real time t . Hence, $\lambda_i : [0, 1] \rightarrow \mathbb{R}$ can be regarded as a function on the unit interval, which allows us to estimate it by techniques from nonparametric regression. Since λ_i is a function of rescaled time t/T , the variables X_{it} in model (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 model (1.1), the outbreak pattern of COVID-19 in country i is determined by the intensity function λ_i . Hence, the question whether the outbreak patterns 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 intensity functions λ_i . More specifically, let $\mathcal{F} = \{\mathcal{I}_k \subseteq [0, 1] : k = 1, \dots, K\}$ be a family of (rescaled) time intervals and let $H_0^{(ijk)}$ be the hypothesis that the intensity

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

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

$$X_{it} = \lambda_i\left(\frac{t}{T}\right) + u_{it} \quad \text{with} \quad u_{it} = \sqrt{\lambda_i\left(\frac{t}{T}\right)}\eta_{it}, \quad (2.1)$$

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

$$X_{it} = \lambda_i\left(\frac{t}{T}\right) + \varepsilon_{it} \quad \text{with} \quad \varepsilon_{it} = \sigma\sqrt{\lambda_i\left(\frac{t}{T}\right)}\eta_{it}, \quad (2.2)$$

where σ is a scaling factor that allows the noise 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 (2.2), where the noise residuals η_{it} have zero mean and unit variance but we do not impose any further distributional assumptions on them.

In order to derive our theoretical results, we impose the following regularity conditions on model (2.2):

(C1) The intensity 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 λ_{\min} and λ_{\max} with $0 \leq \lambda_{\min} \leq \min_{w \in [0, 1]} \lambda_i(w) \leq \max_{w \in [0, 1]} \lambda_i(w) \leq \lambda_{\max} < \infty$ for all i .

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

(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 error variance in model (2.2) equals zero if λ_i is equal to zero. Since we normalize our test statistics by an estimate of the error variance, we need the latter 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 . 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. Independence across time t is more debatable, but it is by no means unreasonable in our modelling framework: The time series process $\mathcal{X}_i = \{X_{it} : 1 \leq t \leq T\}$ produced by model (2.2) is nonstationary for each i . Specifically, both the mean $\mathbb{E}[X_{it}] = \lambda_i(t/T)$ and the variance $\text{Var}(X_{it}) = \sigma^2 \lambda_i(t/T)$ 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 (cp. for example Mikosch and Stărică, 2004; Fryzlewicz et al., 2008). 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 (cp. Mikosch and Stărică, 2000, 2004; Fryzlewicz et al., 2006). 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.

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 intensity 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 $\mathcal{M} := \mathcal{S} \times \{1, \dots, K\}$ 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; in particular, it amounts to the number of tests we simultaneously carry out. As shown by our theoretical results in the Appendix, 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.

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 introduce the expression

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

where $\mathbf{1}(t/T \in \mathcal{I}_k)$ can be regarded as a rectangular kernel weight. A simple application of the law of large numbers yields that $\hat{s}_{ijk,T}/\sqrt{T} = T^{-1} \sum_{t=1}^T \mathbf{1}(t/T \in \mathcal{I}_k) \{\lambda_i(t/T) - \lambda_j(t/T)\} + o_p(1)$ for any fixed pair of countries (i, j) . Hence, the statistic $\hat{s}_{ijk,T}/\sqrt{T}$ estimates the average distance between the intensity functions λ_i and λ_j on the interval \mathcal{I}_k . Under (C2), it holds that

$$\nu_{ijk,T}^2 := \text{Var}(\hat{s}_{ijk,T}) = \frac{\sigma^2}{T} \sum_{t=1}^T \mathbf{1}\left(\frac{t}{T} \in \mathcal{I}_k\right) \left\{ \lambda_i\left(\frac{t}{T}\right) + \lambda_j\left(\frac{t}{T}\right) \right\}.$$

In order to normalize the variance of the statistic $\hat{s}_{ijk,T}$, we scale it by an estimator of $\nu_{ijk,T}$. In particular, we estimate $\nu_{ijk,T}^2$ by

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

where $\hat{\sigma}^2$ is defined as follows: For each country i , let

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

and set $\hat{\sigma}^2 = |\mathcal{C}|^{-1} \sum_{i \in \mathcal{C}} \hat{\sigma}_i^2$ with $\mathcal{C} = \{\ell : \ell = i \text{ or } \ell = j \text{ for some } (i, j) \in \mathcal{S}\}$ denoting the set of countries that are taken into account by our test. As shown in the Supplementary Material, $\hat{\sigma}^2$ is a consistent estimator of σ^2 under our regularity conditions. Normalizing the statistic $\hat{s}_{ijk,T}$ by the estimator $\hat{\nu}_{ijk,T}$ yields the expression

$$\hat{\psi}_{ijk,T} := \frac{\hat{s}_{ijk,T}}{\hat{\nu}_{ijk,T}} = \frac{\sum_{t=1}^T \mathbf{1}\left(\frac{t}{T} \in \mathcal{I}_k\right) (X_{it} - X_{jt})}{\hat{\sigma} \left\{ \sum_{t=1}^T \mathbf{1}\left(\frac{t}{T} \in \mathcal{I}_k\right) (X_{it} + X_{jt}) \right\}^{1/2}}, \quad (3.1)$$

which serves as our test statistic of the hypothesis $H_0^{(ijk)}$. For later reference, we additionally introduce the statistic

$$\hat{\psi}_{ijk,T}^0 = \frac{\sum_{t=1}^T \mathbf{1}(\frac{t}{T} \in \mathcal{I}_k) \sigma \bar{\lambda}_{ij}^{1/2}(\frac{t}{T})(\eta_{it} - \eta_{jt})}{\hat{\sigma}\{\sum_{t=1}^T \mathbf{1}(\frac{t}{T} \in \mathcal{I}_k)(X_{it} + X_{jt})\}^{1/2}} \quad (3.2)$$

with $\bar{\lambda}_{ij}(u) = \{\lambda_i(u) + \lambda_j(u)\}/2$, which is identical to $\hat{\psi}_{ijk,T}$ under $H_0^{(ijk)}$.

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 at level α is

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

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. 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

$$1 - \mathbb{P}\left(\max_{(i,j,k) \in \mathcal{M}_0} |\hat{\psi}_{ijk,T}| \leq c_T(\alpha)\right) \leq \alpha. \quad (3.3)$$

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

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

where $\hat{\psi}_{ijk,T}^0$ has been introduced in (3.2). (Note that both the statistic $\tilde{\Psi}_T$ and the quantile $c_T(\alpha)$ depend on p 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.)

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; cp. for example Dümbgen (2002), 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

$$\begin{aligned} \text{FWER}(\alpha) &= \mathbb{P}\left(\exists(i, j, k) \in \mathcal{M}_0 : |\hat{\psi}_{ijk,T}| > c_T(\alpha, h_k)\right) \\ &= 1 - \mathbb{P}\left(\forall(i, j, k) \in \mathcal{M}_0 : |\hat{\psi}_{ijk,T}| \leq c_T(\alpha, h_k)\right) \\ &= 1 - \mathbb{P}\left(\forall(i, j, k) \in \mathcal{M}_0 : a_k(|\hat{\psi}_{ijk,T}| - b_k) \leq q_T(\alpha)\right) \\ &= 1 - \mathbb{P}\left(\max_{(i,j,k) \in \mathcal{M}_0} a_k(|\hat{\psi}_{ijk,T}| - b_k) \leq q_T(\alpha)\right), \end{aligned} \quad (3.4)$$

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

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

in order to ensure control of the FWER at level α . Comparing (3.4) with (3.3), 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 $\hat{\psi}_{ijk,T}$ 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 with better theoretical properties (cp. Dümbgen and Spokoiny, 2001), we restrict attention to the critical values $c_T(\alpha, h_k)$ in the sequel. There is, however, one complication we need to deal with: As the quantiles $q_T(\alpha)$ are not known in practice, we can not 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

$$\begin{aligned}\hat{\psi}_{ijk,T}^0 &= \frac{\sum_{t=1}^T \mathbf{1}(\frac{t}{T} \in \mathcal{I}_k) \sigma \bar{\lambda}_{ij}^{1/2}(\frac{t}{T})(\eta_{it} - \eta_{jt})}{\hat{\sigma}\{\sum_{t=1}^T \mathbf{1}(\frac{t}{T} \in \mathcal{I}_k)(X_{it} + X_{jt})\}^{1/2}} \\ &\approx \frac{1}{\sqrt{2Th_k}} \sum_{t=1}^T \mathbf{1}\left(\frac{t}{T} \in \mathcal{I}_k\right) \{\eta_{it} - \eta_{jt}\}.\end{aligned}$$

A Gaussian version of the statistic displayed in the final line above is given by

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

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

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

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 $q_{T,\text{Gauss}}(\alpha)$ 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 $\alpha \in (0, 1)$:

$$\text{For each } (i, j, k) \in \mathcal{M}, \text{ reject } H_0^{(ijk)} \text{ if } |\hat{\psi}_{ijk,T}| > c_{T,\text{Gauss}}(\alpha, h_k), \quad (3.5)$$

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

3.3 Formal properties of the test

In Theorem A.1 of the Appendix, we prove that under appropriate regularity conditions, the test defined in (3.5) (asymptotically) controls the familywise error rate $\text{FWER}(\alpha)$ for each pre-specified significance level α . As shown in Corollary A.1, this has the following implication:

$$\begin{aligned}\mathbb{P}\left(\forall (i, j, k) \in \mathcal{M} : \text{ If } |\hat{\psi}_{ijk,T}| > c_{T,\text{Gauss}}(\alpha, h_k), \text{ then } (i, j, k) \notin \mathcal{M}_0\right) \\ \geq 1 - \alpha + o(1),\end{aligned} \quad (3.6)$$

where \mathcal{M}_0 is the set of triples $(i, j, k) \in \mathcal{M}$ for which $H_0^{(ijk)}$ holds true. Verbally, (3.6) can be expressed as follows:

With (asymptotic) probability at least $1 - \alpha$, the null $H_0^{(ijk)}$ is violated for all $(i, j, k) \in \mathcal{M}$ for which the test rejects $H_0^{(ijk)}$. (3.7)

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 $H_0^{(ijk)}$. (3.8)

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

3.4 Implementation of the test in practice

For a given significance level $\alpha \in (0, 1)$, the test procedure defined in (3.5) 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 $N = 1000$) samples of independent standard normal random variables $\{Z_{it}^{(\ell)} : 1 \leq t \leq T, 1 \leq t \leq T\}$ for $1 \leq \ell \leq N$. 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 $\{\Phi_T^{(\ell)} : 1 \leq \ell \leq N\}$. 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 $1 \leq k \leq K$ 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 variables $r_{ijk, T} = \mathbf{1}(|\hat{\psi}_{ijk, T}| > c_{T, \text{Gauss}}(\alpha, h_k))$ 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 $(i, j) \in \mathcal{S}$ which shows the intervals \mathcal{I}_k in the set $\mathcal{F}_{\text{reject}}(i, j) = \{\mathcal{I}_k \in \mathcal{F} : r_{ijk, T} = 1\}$, that is, the intervals \mathcal{I}_k for which the test rejects the null $H_0^{(ijk)}$. We in particular use this graphical device to present the test results of our empirical application in Section 4. According to (3.6), we can make the following simultaneous confidence statement about the intervals in $\mathcal{F}_{\text{reject}}(i, j)$:

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 $\mathcal{F}_{\text{reject}}(i, j)$. (3.9)

Hence, we can claim with statistical confidence at least $1 - \alpha$ that the functions λ_i and λ_j differ on each time interval \mathcal{I}_k which is shown in the plots of our graphical device.

4 Empirical application to COVID-19 data

We now use our test to analyze the outbreak patterns of the COVID-19 virus. We proceed in two steps. In Section 4.1, we conduct some Monte-Carlo experiments to assess the finite sample performance of our test. Specifically, we run some experiments which show that the test controls the FWER at level α as predicted by the theory. Having established this, we then apply the test to a sample of current COVID-19 data in Section 4.2.

4.1 Simulation experiments

We consider the following simulation setup:

- We simulate count data $\mathcal{X} = \{X_{it} : 1 \leq i \leq n, 1 \leq t \leq T\}$ by drawing X_{it} independently from the double Poisson distribution $dP_{\lambda_i(t/T), \sigma^2}$ with parameters $\lambda_i(t/T)$ and σ^2 . The double Poisson distribution dP_{λ, σ^2} with parameters λ and σ^2 was rigorously introduced by Efron (1986). As shown there, dP_{λ, σ^2} is a probability distribution on $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ with mean and variance approximately equal to λ and $\sigma^2\lambda$, respectively. Hence, by letting $X_{it} \sim dP_{\lambda_i(t/T), \sigma^2}$, we approximately obtain a nonparametric regression model of the form

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

where the noise variables η_{it} have zero mean and unit variance.

- Throughout the simulation study, 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 the function

$$\lambda(u) = 5000 \exp\left(-\frac{(10u - 3)^2}{2}\right) + 50,$$

which has a similar shape as some of the estimated curves in the application of Section 4.2. A plot of the function λ is provided in Figure 1a.

- We set the overdispersion parameter $\sigma = 10$. (Alternatively, we may pick $\sigma = 15$ or $\sigma = 20$.) The estimated values of σ in the application take similar values.

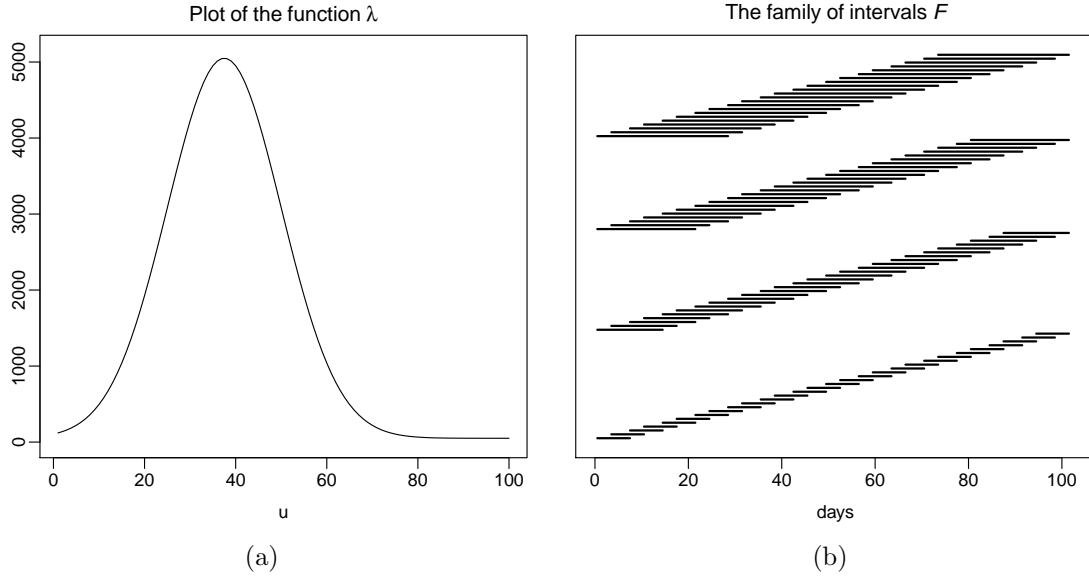


Figure 1: (a) Plot of the function λ ; (b) plot of the family \mathcal{F} .

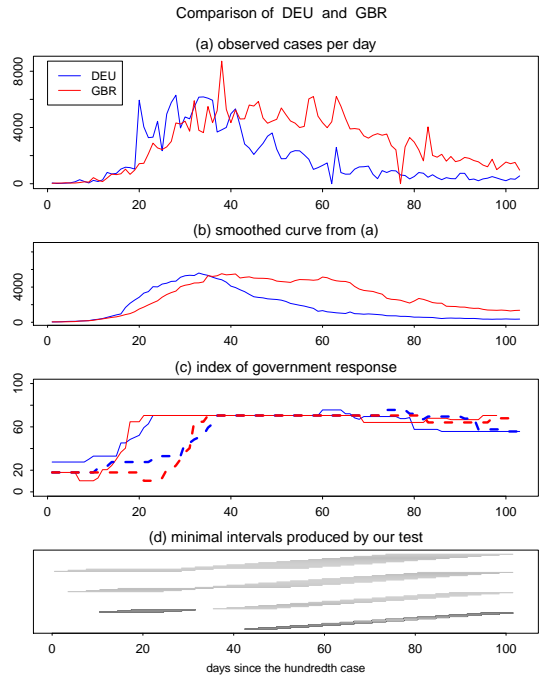
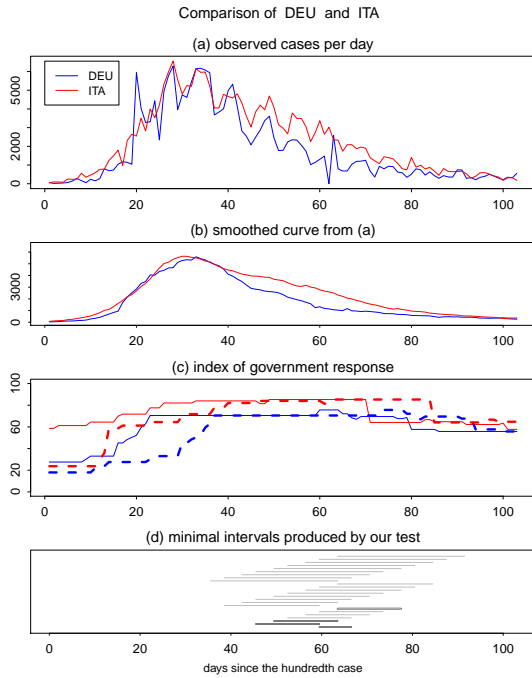
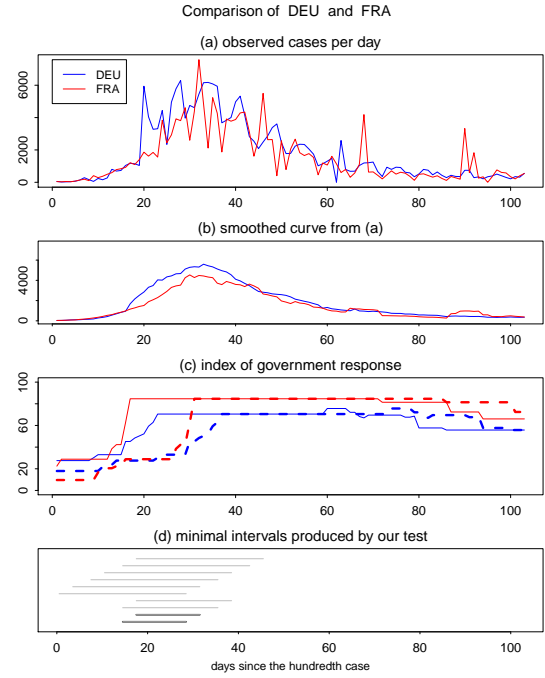
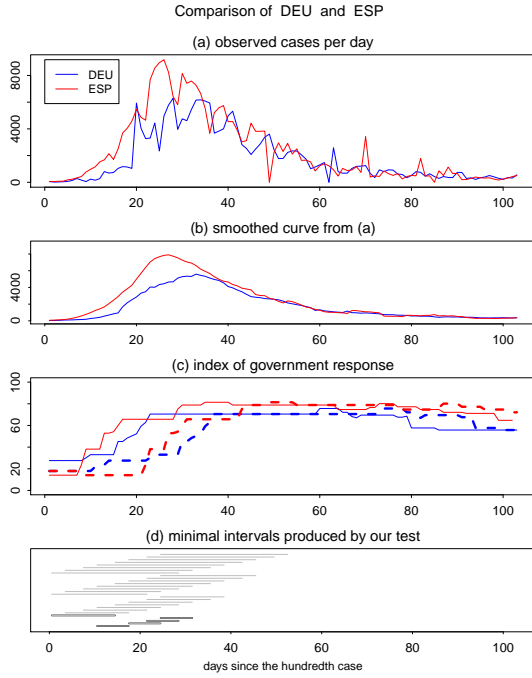
Table 1: Size of the test for different number of time series $n \in \{5, 10, 50\}$.

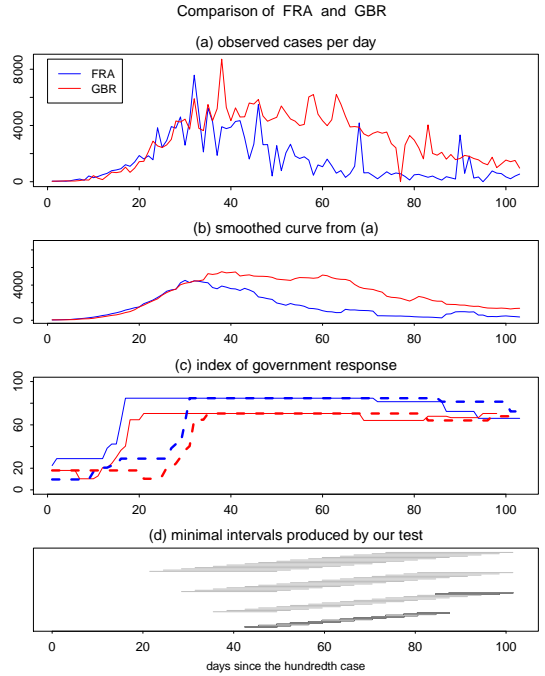
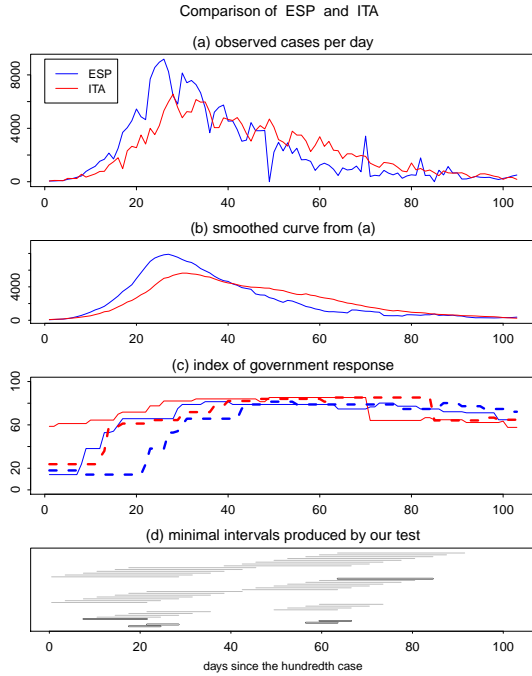
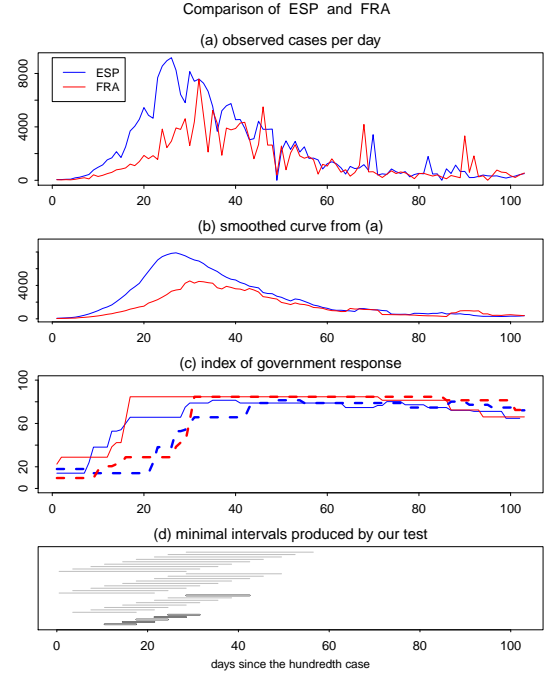
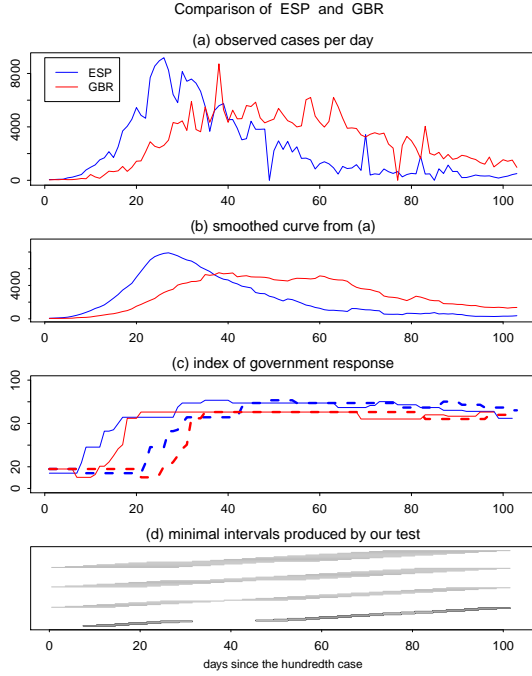
	$n = 5$			$n = 10$			$n = 50$		
	nominal size α			nominal size α			nominal size α		
	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
$T = 100$	0.013	0.040	0.086	0.016	0.054	0.106	0.009	0.045	0.094
$T = 250$	0.013	0.044	0.102	0.008	0.041	0.089	0.013	0.057	0.107
$T = 500$	0.011	0.052	0.090	0.007	0.057	0.114	0.011	0.049	0.106

- We consider different values for T and n , in particular, $T \in \{100, 250, 500\}$ and $n \in \{5, 10, 50\}$.
- 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 with $h_k \in \{7/T, 14/T, 21/T, 28/T\}$. Hence, all intervals in \mathcal{F} have length either 7 days (1 week), 14 days (2 weeks), 21 days (3 weeks), or 28 days (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 $j = 1, 2, \dots$. A graphical presentation of the family \mathcal{F} is given in Figure 1b.

Each simulation experiment is based on $R = 5000$ simulated samples \mathcal{X} of count data.

4.2 Analysis of COVID-19 data





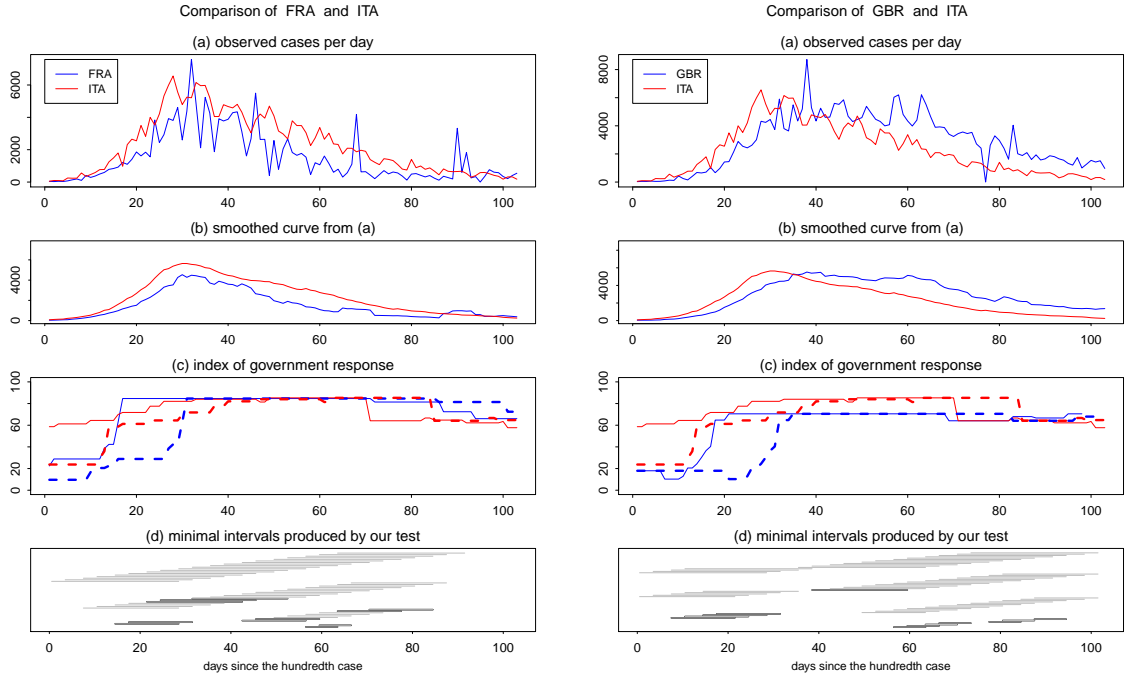


Figure 2: Summary of the results for the daily numbers of the new cases of COVID-19 in Germany, Spain, France, Italy and The United Kingdom with one subfigure corresponding to the results of one pairwise comparison of these countries. Panel (a) of each subfigure shows the two observed time series. Panel (b) in each subfigure shows the smoothed versions of the time series from (a) using the bandwidth of 7 days. Panel (c) of each subfigure show the corresponding time series of the Government Response Index (solid line) against the lagged version of itself (dashed line), with lag being 14 days. Panel (d) of each subfigure depicts the set of intervals $\mathcal{F}_{\text{reject}}(i, j)$ in gray and the set of minimal intervals constructed on the basis of $\mathcal{F}_{\text{reject}}(i, j)$ in black.

A Appendix

In what follows, we state and prove the main theoretical results on the multiscale test developed in Section 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} .

Theorem A.1. *Let (C2) and (C1) be satisfied. Moreover, assume that (i) $h_{\max} = o(1/\log T)$, (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 $\alpha \in (0, 1)$,*

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

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.

According to Theorem A.1, the multiscale test asymptotically controls the FWER at level α under conditions (C2)–(C1) and the restrictions (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 (C1) 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 A.1. It provides the theoretical justification needed to make simultaneous confidence statements of the form (3.7)–(3.9).

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

$$\mathbb{P}\left(\forall(i, j, k) \in \mathcal{M} : \text{If } |\hat{\psi}_{ijk,T}| > c_{T,\text{Gauss}}(\alpha, h_k), \text{ then } (i, j, k) \notin \mathcal{M}_0\right) \geq 1 - \alpha + o(1)$$

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

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

Step 1. Let $\hat{\Psi}_T = \max_{(i,j,k) \in \mathcal{M}} a_k(|\hat{\psi}_{ijk,T}^0| - b_k)$ with $\hat{\psi}_{ijk,T}^0$ introduced in (3.2) and define $\Psi_T = \max_{(i,j,k) \in \mathcal{M}} a_k(|\psi_{ijk,T}^0| - b_k)$ with

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

To start with, we prove that

$$|\hat{\Psi}_T - \Psi_T| = o_p(r_T), \quad (\text{A.1})$$

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, $\rho_T / r_T \rightarrow 0$ as $T \rightarrow \infty$. Since the proof of (A.1) is rather technical and lengthy, the details are provided in the Supplementary Material.

Step 2. We next prove that

$$\sup_{q \in \mathbb{R}} \left| \mathbb{P}(\Psi_T \leq q) - \mathbb{P}(\Phi_T \leq q) \right| = o(1). \quad (\text{A.2})$$

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

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

for $(i, j, k) \in \mathcal{M}$ and let $\mathbf{V}_t = (V_t^{(ijk)} : (i, j, k) \in \mathcal{M})$ 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

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

Analogously, we define

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

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

$\phi_{ijk,T} = T^{-1/2} \sum_{t=1}^T W_t^{(ijk)}$ and

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

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

$$\begin{aligned}\mathbb{P}(\Psi_T \leq q) &= \mathbb{P}\left(\max_{(i,j,k) \in \mathcal{M}} a_k \left\{ \left| \frac{1}{\sqrt{T}} \sum_{t=1}^T V_t^{(ijk)} \right| - b_k \right\} \leq q\right) \\ &= \mathbb{P}\left(\left| \frac{1}{\sqrt{T}} \sum_{t=1}^T V_t^{(ijk)} \right| \leq q_{ijk} \text{ for all } (i,j,k) \in \mathcal{M}\right) \\ &= \mathbb{P}\left(\left| \frac{1}{\sqrt{T}} \sum_{t=1}^T \mathbf{V}_t \right| \leq \mathbf{q}\right),\end{aligned}$$

where \mathbf{q} is the \mathbb{R}^p -vector with the entries $q_{ijk} = q/a_k + b_k$. Analogously, we have

$$\mathbb{P}(\Phi_T \leq q) = \mathbb{P}\left(\left| \frac{1}{\sqrt{T}} \sum_{t=1}^T \mathbf{W}_t \right| \leq \mathbf{q}\right).$$

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 A.1. *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 $T \rightarrow \infty$.
- (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

$$\begin{aligned}\sup_{\mathbf{q} \in \mathbb{R}^p} \left| \mathbb{P}\left(\left| \frac{1}{\sqrt{T}} \sum_{t=1}^T \mathbf{V}_t \right| \leq \mathbf{q}\right) - \mathbb{P}\left(\left| \frac{1}{\sqrt{T}} \sum_{t=1}^T \mathbf{W}_t \right| \leq \mathbf{q}\right) \right| \\ \leq C \left\{ \left(\frac{B_T^2 \log^7(pT)}{T} \right)^{1/6} + \left(\frac{B_T^2 \log^3(pT)}{T^{1-2/\theta}} \right)^{1/3} \right\}, \quad (\text{A.3})\end{aligned}$$

where C depends only on δ and θ .

It is straightforward to verify that assumptions (a)–(c) are satisfied under the conditions of Theorem 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 (A.3) is $o(1)$ for this choice of B_T . Hence, Proposition A.1 yields that

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

which in turn implies (A.2).

Step 3. With the help of (A.1) and (A.2), we now show that

$$\sup_{q \in \mathbb{R}} \left| \mathbb{P}(\hat{\Psi}_T \leq q) - \mathbb{P}(\Phi_T \leq q) \right| = o(1). \quad (\text{A.4})$$

To start with, the above supremum can be bounded by

$$\begin{aligned} & \sup_{q \in \mathbb{R}} \left| \mathbb{P}(\hat{\Psi}_T \leq q) - \mathbb{P}(\Phi_T \leq q) \right| \\ &= \sup_{q \in \mathbb{R}} \left| \mathbb{P}\left(\Psi_T \leq q + \{\Psi_T - \hat{\Psi}_T\}\right) - \mathbb{P}(\Phi_T \leq q) \right| \\ &\leq \max \left\{ \sup_{q \in \mathbb{R}} \left| \mathbb{P}\left(\Psi_T \leq q + |\Psi_T - \hat{\Psi}_T|\right) - \mathbb{P}(\Phi_T \leq q) \right|, \right. \\ &\quad \left. \sup_{q \in \mathbb{R}} \left| \mathbb{P}\left(\Psi_T \leq q - |\Psi_T - \hat{\Psi}_T|\right) - \mathbb{P}(\Phi_T \leq q) \right| \right\} \\ &\leq \max \left\{ \sup_{q \in \mathbb{R}} \left| \mathbb{P}\left(\Psi_T \leq q + r_T\right) - \mathbb{P}(\Phi_T \leq q) \right| + \mathbb{P}\left(|\Psi_T - \hat{\Psi}_T| > r_T\right), \right. \\ &\quad \left. \sup_{q \in \mathbb{R}} \left| \mathbb{P}\left(\Psi_T \leq q - r_T\right) - \mathbb{P}(\Phi_T \leq q) \right| + \mathbb{P}\left(|\Psi_T - \hat{\Psi}_T| > r_T\right) \right\}. \end{aligned}$$

Moreover,

$$\begin{aligned} & \sup_{q \in \mathbb{R}} \left| \mathbb{P}\left(\Psi_T \leq q \pm r_T\right) - \mathbb{P}(\Phi_T \leq q) \right| + \mathbb{P}\left(|\Psi_T - \hat{\Psi}_T| > r_T\right) \\ &\leq \sup_{q \in \mathbb{R}} \left| \mathbb{P}\left(\Psi_T \leq q \pm r_T\right) - \mathbb{P}\left(\Phi_T \leq q \pm r_T\right) \right| \\ &\quad + \sup_{q \in \mathbb{R}} \left| \mathbb{P}\left(\Phi_T \leq q \pm r_T\right) - \mathbb{P}(\Phi_T \leq q) \right| + \mathbb{P}\left(|\Psi_T - \hat{\Psi}_T| > r_T\right) \\ &= \sup_{q \in \mathbb{R}} \left| \mathbb{P}\left(\Phi_T \leq q \pm r_T\right) - \mathbb{P}(\Phi_T \leq q) \right| + o(1), \end{aligned}$$

where the last line follows from (A.1) and (A.2). Finally, by Nazarov's inequality (cp. Nazarov, 2003 and Lemma A.1 in Chernozhukov et al., 2017),

$$\begin{aligned} & \sup_{q \in \mathbb{R}} \left| \mathbb{P}\left(\Phi_T \leq q \pm r_T\right) - \mathbb{P}(\Phi_T \leq q) \right| \\ &= \sup_{\mathbf{q} \in \mathbb{R}^p} \left| \mathbb{P}\left(\left|\frac{1}{\sqrt{T}} \sum_{t=1}^T \mathbf{W}_t\right| \leq \mathbf{q} \pm r_T\right) - \mathbb{P}\left(\left|\frac{1}{\sqrt{T}} \sum_{t=1}^T \mathbf{W}_t\right| \leq \mathbf{q}\right) \right| \leq Cr_T \sqrt{\log(2p)}, \end{aligned}$$

where C is a constant that depends only on the parameter δ defined in condition (a) of Proposition A.1.

Step 4. By definition of the quantile $q_{T,\text{Gauss}}(\alpha)$, it holds that $\mathbb{P}(\Phi_T \leq q_{T,\text{Gauss}}(\alpha)) \geq 1 - \alpha$. As shown in the Supplementary Material, we even have that

$$\mathbb{P}(\Phi_T \leq q_{T,\text{Gauss}}(\alpha)) = 1 - \alpha \quad (\text{A.5})$$

for any $\alpha \in (0, 1)$. From this and (A.4), it immediately follows that

$$\mathbb{P}(\hat{\Psi}_T \leq q_{T,\text{Gauss}}(\alpha)) = 1 - \alpha + o(1), \quad (\text{A.6})$$

which in turn yields the statement of Theorem A.1. \square

Proof of Corollary A.1. By Theorem A.1,

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

which gives the statement of Corollary A.1. \square

References

- CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2017). Central limit theorems and bootstrap in high dimensions. *Annals of Probability*, **45** 2309–2352.
- 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.
- 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.
- 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.
- ROBINSON, P. M. (1989). Nonparametric estimation of time-varying parameters. In *Statistical Analysis and Forecasting of Economic Structural Change* (P. Hackl, ed.). Springer, 253–264.
- ROHDE, A. (2008). Adaptive goodness-of-fit tests based on signed ranks. *Annals of Statistics*, **36** 1346–1374.
- 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.

S Supplementary Material

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

Lemma S.1. *Under the conditions of Theorem A.1, it holds that*

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

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

$$\frac{1}{T} \sum_{t=2}^T (X_{it} - X_{it-1})^2 = \frac{\sigma^2}{T} \sum_{t=2}^T \lambda_i\left(\frac{t}{T}\right) (\eta_{it} - \eta_{it-1})^2 + \{R_{i,T}^{(1)} + \dots + R_{i,T}^{(5)}\}, \quad (\text{S.1})$$

where

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

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

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

where we let $g(x) = x$, $g(x) = |x|$ or $g(x) = x^2$, 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 (C2) and (C1), we obtain that

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

and

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

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

$$\max_{i \in \mathcal{C}} \left| \frac{1}{T} \sum_{t=2}^T (X_{it} - X_{it-1})^2 - \frac{2}{T} \sum_{t=1}^T \lambda_i\left(\frac{t}{T}\right) \right| = O_p\left(\sqrt{\frac{\log p}{T}}\right). \quad (\text{S.2})$$

By similar but simpler arguments, we additionally get that

$$\max_{i \in \mathcal{C}} \left| \frac{1}{T} \sum_{t=1}^T X_{it} - \frac{1}{T} \sum_{t=1}^T \lambda_i\left(\frac{t}{T}\right) \right| = O_p\left(\sqrt{\frac{\log p}{T}}\right). \quad (\text{S.3})$$

From (S.2) and (S.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. \square

Proof of (A.1). Since

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

it suffices to prove that

$$\max_{(i,j,k) \in \mathcal{M}} |\hat{\psi}_{ijk,T}^0 - \psi_{ijk,T}^0| = o_p\left(\frac{r_T}{\sqrt{\log T}}\right). \quad (\text{S.4})$$

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

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

where

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

With this notation, we can establish the bound

$$\begin{aligned} \max_{(i,j,k) \in \mathcal{M}} |\hat{\psi}_{ijk,T}^0 - \psi_{ijk,T}^0| &\leq \max_{(i,j,k) \in \mathcal{M}} |\hat{\psi}_{ijk,T}^* - \psi_{ijk,T}^0| \\ &\quad + \left| \frac{\sigma}{\hat{\sigma}} - 1 \right| \max_{(i,j,k) \in \mathcal{M}} |\hat{\psi}_{ijk,T}^* - \psi_{ijk,T}^0| \\ &\quad + \left| \frac{\sigma}{\hat{\sigma}} - 1 \right| \max_{(i,j,k) \in \mathcal{M}} |\psi_{ijk,T}^0|, \end{aligned}$$

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

$$\max_{(i,j,k) \in \mathcal{M}} |\hat{\psi}_{ijk,T}^* - \psi_{ijk,T}^0| = O_p\left(\frac{\log p}{\sqrt{Th_{\min}}} + h_{\max}\sqrt{\log p}\right) \quad (\text{S.5})$$

$$\max_{(i,j,k) \in \mathcal{M}} |\psi_{ijk,T}^0| = O_p(\sqrt{\log p}) \quad (\text{S.6})$$

$$|\hat{\sigma}^2 - \sigma^2| = O_p\left(\sqrt{\frac{\log p}{T}}\right). \quad (\text{S.7})$$

Since (S.7) has already been verified in Lemma S.1, it remains to prove the statements (S.5) and (S.6).

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

$$\max_{i \in \mathcal{C}} \max_{1 \leq k \leq K} \left| \frac{1}{\sqrt{Th_k}} \sum_{t=1}^T \mathbf{1}\left(\frac{t}{T} \in \mathcal{I}_k\right) w_i\left(\frac{t}{T}\right) \eta_{it} \right| = O_p(\sqrt{\log p}), \quad (\text{S.8})$$

where $w_i(t/T)$ are general 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} . This immediately implies (S.6).

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

$$\max_{i \in \mathcal{C}} \max_{1 \leq k \leq K} \left| \frac{1}{Th_k} \sum_{t=1}^T \mathbf{1}\left(\frac{t}{T} \in \mathcal{I}_k\right) \lambda_i\left(\frac{t}{T}\right) - \frac{1}{h_k} \int_{w \in \mathcal{I}_k} \lambda_i(w) dw \right| \leq \frac{C}{Th_{\min}}. \quad (\text{S.9})$$

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

$$\begin{aligned} \max_{(i,j,k) \in \mathcal{M}} \left| \frac{1}{Th_k} \sum_{t=1}^T \mathbf{1}\left(\frac{t}{T} \in \mathcal{I}_k\right) (X_{it} + X_{jt}) \right. \\ \left. - \frac{1}{h_k} \int_{w \in \mathcal{I}_k} \{\lambda_i(w) + \lambda_j(w)\} dw \right| = O_p\left(\sqrt{\frac{\log p}{Th_{\min}}}\right) \end{aligned} \quad (\text{S.10})$$

and

$$\begin{aligned} \max_{(i,j,k) \in \mathcal{M}} \left| \frac{1}{\sqrt{Th_k}} \sum_{t=1}^T \mathbf{1}\left(\frac{t}{T} \in \mathcal{I}_k\right) \bar{\lambda}_{ij}^{1/2}\left(\frac{t}{T}\right) (\eta_{it} - \eta_{jt}) \right. \\ \left. - \left\{ \frac{\int_{w \in \mathcal{I}_k} \bar{\lambda}_{ij}(w) dw}{h_k} \right\}^{1/2} \frac{1}{\sqrt{Th_k}} \sum_{t=1}^T \mathbf{1}\left(\frac{t}{T} \in \mathcal{I}_k\right) (\eta_{it} - \eta_{jt}) \right| \\ = O_p\left(h_{\max}\sqrt{\log p}\right). \end{aligned} \quad (\text{S.11})$$

The claim (S.5) follows from (S.10) and (S.11) along with straightforward calculations. \square

Proof of (A.5). The proof is by contradiction. Suppose that (A.5) 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,

$$\mathbb{P}(\Phi_T \leq q_{T,\text{Gauss}}(\alpha)) - \mathbb{P}(\Phi_T \leq q_{T,\text{Gauss}}(\alpha) - \eta) \leq C\eta\sqrt{\log(2p)}$$

for any $\eta > 0$ with C depending only on the parameter δ specified in condition (a) of Proposition A.1. Hence,

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

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