

# 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  $\mathcal{D}_i = \{X_{it} : 1 \leq t \leq T\}$  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,  $X_{it} \sim \mathcal{P}_{\lambda_i(t/T)}$ . 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 [cite], we let the regression function  $\lambda_i$  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  $\lambda_i$  is a function of rescaled time  $t/T$ , the variables  $X_{it}$  in model (1.1) depend on the time series length  $T$ , 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  $\lambda_i$ . Hence, the question whether the outbreak patterns are comparable across countries amounts to the question whether the intensity functions  $\lambda_i$  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  $\lambda_i$ . 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  $\lambda_i$  and  $\lambda_j$  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 a consequence, we can 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  $\lambda_i$  and  $\lambda_j$  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  $X_{it} \sim \mathcal{P}_{\lambda_i(t/T)}$  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  $\eta_{it}$  has zero mean and unit variance. In this model, both the mean and the noise variance are described by the same function  $\lambda_i$ . In empirical applications, however, the noise variance often tends to be much larger than that implied by the Poisson distribution; cp. [cite]. To deal with this issue, so-called quasi-Poisson models 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  $\sigma$  is a scaling factor that allows the noise variance to be a multiple of the mean function  $\lambda_i$ . In what follows, we assume that the observed data  $X_{it}$  are produced by model (2.2), where the noise residuals  $\eta_{it}$  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 random variables  $\eta_{it}$  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  $\theta > 4$ .
- (C2) The intensity functions  $\lambda_i$  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  $0 \leq \lambda_{\min} \leq \min_{w \in [0, 1]} \lambda_i(w) \leq \max_{w \in [0, 1]} \lambda_i(w) \leq \lambda_{\max} < \infty$  for all  $i$ .

[Discuss why independence of  $\eta_{it}$  across  $i$  and  $t$  is reasonable/not too simplistic. Discuss relationship between nonstationarities and autocorrelations in this context. Refer to literature on simple volatility models.]

### 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  $\lambda_i$  and  $\lambda_j$  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 Section [number],  $p$  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

$\lambda_i$  and  $\lambda_j$  on the interval  $\mathcal{I}_k$ . Under (C1), 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  $\sigma^2$  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

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

with  $\bar{\lambda}_{ij}(u) = \{\lambda_i(u) + \lambda_j(u)\}/2$ , which is identical to the test statistic  $\hat{\psi}_{ijk,T}$  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  $\alpha$ , 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  $\alpha$  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  $\alpha$  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 [cite] and has been used in many other studies since then; cp. [cite]. In our context, the approach of [cite] leads to the critical values

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

where  $a_k = \{\log(e/h_k)\}^{1/2}/\log \log(e/h_k)$  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

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

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  $\alpha$ . 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. [cite]), 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

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

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  $\Phi_T$ , 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  $a_k = \{\log(e/h_k)\}^{1/2}/\log \log(e/h_k)$  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  $\alpha$ . 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  $H_0^{(ijk)}$ .*

In other words:

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

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  $\{Z_{it}^{(\ell)} : 1 \leq t \leq T, 1 \leq i \leq T\}$  for  $1 \leq \ell \leq N$ . Compute the value  $\Phi_T^{(\ell)}$  of the Gaussian statistic  $\Phi_T$  for each sample  $\ell$  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  $\lambda_i$  and  $\lambda_j$  differ on each interval in  $\mathcal{F}_{\text{reject}}(i, j)$ .*

Hence, we can claim with statistical confidence at least  $1 - \alpha$  that the functions  $\lambda_i$  and  $\lambda_j$  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  $\alpha$  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 should show by simulations that the method indeed controls the FWER as predicted by the theory! Here's a first suggestion for a simulation setting.

Simulation setting:

- We simulate data  $\{X_{it} : 1 \leq t \leq T, 1 \leq i \leq n\}$  for different  $n$  and  $T$  from the model

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

where  $\eta_{it}$  are i.i.d. standard normal.

- We assume that  $H_0^{(ijk)}$  holds true for all  $(i, j, k)$  under consideration, implying that  $\lambda_i = \lambda$  for all  $i$ . We in particular let

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

This function has a similar form as the estimated curves in the application. Moreover,  $\lambda(0) \approx 100$ , which fits our normalization that each time series starts at the first day with more than 100 new cases.

- We set  $\sigma = 10$ . (We may also take something like 15 or 20.) (Estimates of  $\sigma$  in the application take similar values.)
- We let  $\mathcal{S} = \{(i, j) : 1 \leq i < j \leq n\}$  and  $\mathcal{F}$  a family of 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) [we may even include longer intervals with 5, 6, etc. 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 of intervals  $\mathcal{F}$  is given in Figure ??.
- We may let  $T \in \{50, 100\}$  and  $n \in \{5, 10, 50, 100\}$ .

## 4.2 Analysis of COVID-19 data

(1) As I wrote in my previous email, I found the test results a bit too liberal. That's why I've played around with the data myself. In particular, I have run the simultaneous test with five countries (DE, ES, FR, IT, UK) on the basis of the R code that I have written at the beginning of the project. Figures 1 and 2 show part of the results (only the comparisons of DE with the other four countries). As you can see, the test results are less liberal than those you sent me. Hence, there must

be some issue with the code (either mine or yours ...). I double-checked mine but couldn't find any bug yet ... Please double-check yours!

(2) There are some issues with the data: there are negative values in the time series for Spain, there is a day with 0 new case reported for DE, the last observation for FR is way too high to be credible, etc. I guess some data cleaning is needed!

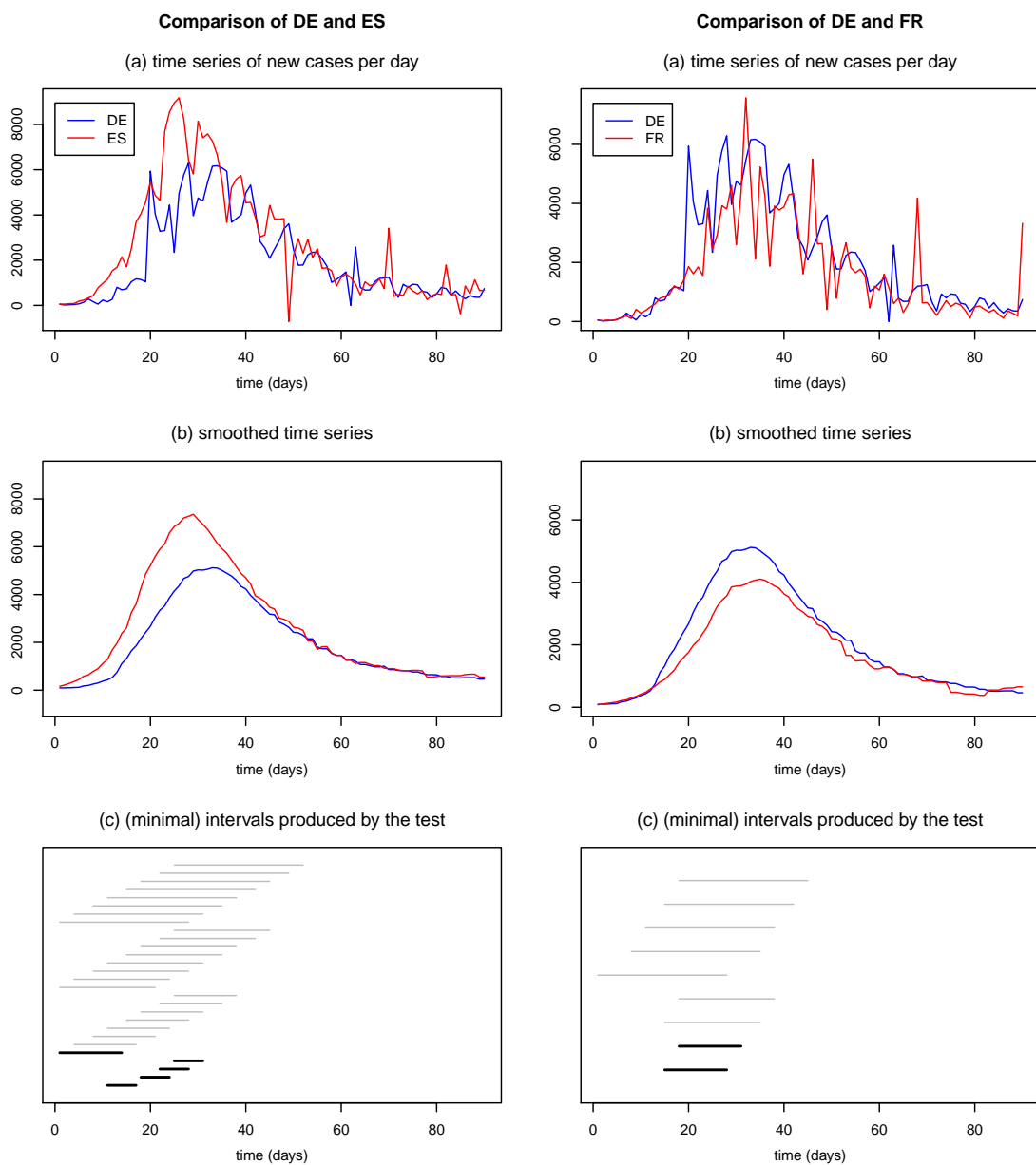


Figure 1

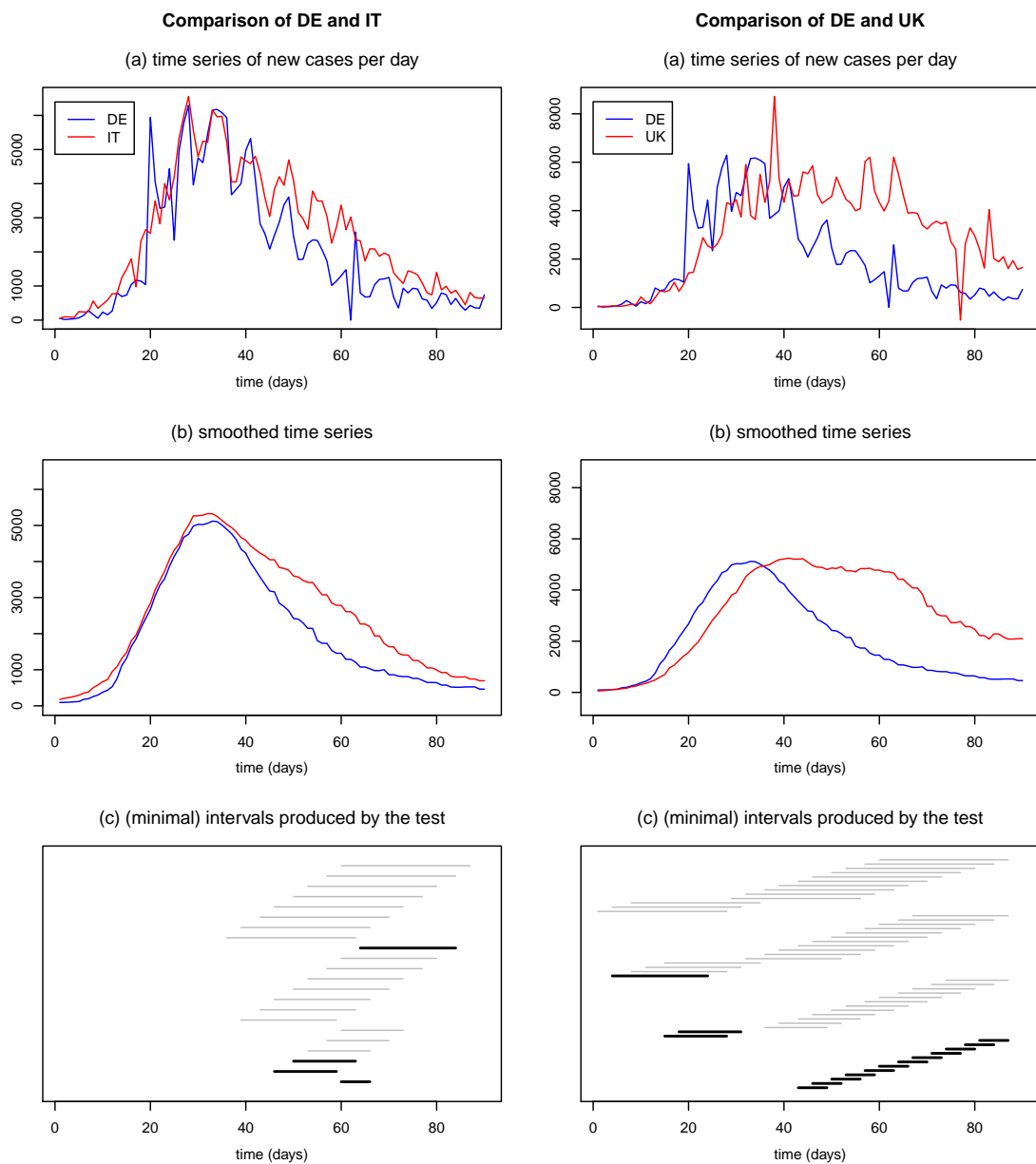


Figure 2

## A Appendix

In what follows, we state and prove the main theoretical results on our multiscale test. To do so, we define  $h_{\min} := \min_{1 \leq k \leq K} h_k$  and  $h_{\max} := \max_{1 \leq k \leq K} h_k$  to be the smallest and largest length of an interval in the family  $\mathcal{F}$ , respectively.

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

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

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

**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

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

where  $\{r_T\}$  is any null sequence that converges more slowly to zero than  $\rho_T = \sqrt{(\log p)/(Th_{\min})} + h_{\max}\sqrt{\log p}$ , 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

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

To do so, we rewrite the statistics  $\Psi_T$  and  $\Phi_T$  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

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

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 [cite]. 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)  $T^{-1} \sum_{t=1}^T \mathbb{E}[|V_t^{(ijk)}|^{2+r}] \leq C_T^r$  for all  $(i, j, k) \in \mathcal{M}$  and  $r = 1, 2$ , where  $C_T \geq 1$  are constants that may tend to infinity as  $T \rightarrow \infty$ .
- (c)  $\mathbb{E}[\{\max_{(i,j,k) \in \mathcal{M}} |V_t^{(ijk)}|/C_T\}^\theta] \leq 2$  for all  $t$  and some  $\theta > 4$ .

Then

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

where  $C$  depends only on  $\delta$  and  $\theta$ .

It is straightforward to verify that assumptions (a)–(c) are satisfied under the conditions of Theorem A.1 for sufficiently large  $T$ , where  $C_T$  can be chosen as  $C_T = C_0 p^{1/\theta} h_{\min}^{-1/2}$  with  $C_0$  sufficiently large. Moreover, it can be shown that the right-hand side of (A.3) is  $o(1)$  for this choice of  $C_T$  under our conditions. 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

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

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

Moreover,

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

where the last line follows from (A.1) and (A.2). Finally, by Nazarov's inequality [\[cite\]](#),

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

where  $C$  is a constant that depends only on the parameter  $\delta$  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

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

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

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

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

## S Supplementary Material

We now provide the technical details that were 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

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

where we set  $g(x) = x$ ,  $g(x) = |x|$  or  $g(x) = x^2$ . Using this uniform convergence result along with conditions (C1) and (C2), it is not difficult to show that

$$\begin{aligned} \max_{1 \leq i \leq n} \left| \frac{1}{T} \sum_{t=2}^T \lambda_i\left(\frac{t}{T}\right) (\eta_{it} - \eta_{it-1})^2 - \frac{2}{T} \sum_{t=1}^T \lambda_i\left(\frac{t}{T}\right) \right| &= O_p\left(\sqrt{\frac{\log p}{T}}\right) \\ \max_{1 \leq \ell \leq 5} \max_{1 \leq i \leq n} |R_{i,T}^{(\ell)}| &= O_p(T^{-1}). \end{aligned}$$



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

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

By similar but simpler arguments, we additionally get that

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

From (S.2) and (S.3), it follows that  $\max_{1 \leq i \leq n} |\hat{\sigma}_i^2 - \sigma^2| = O_p(\sqrt{\log p/T})$ , which in turn implies that  $|\hat{\sigma}^2 - \sigma^2| = O_p(\sqrt{\log p/T})$  as well.  $\square$

**Proof of (A.1).** We prove the claim

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

which implies (A.1). 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

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

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

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

(S.7) has already been verified in Lemma S.1. It thus remains to prove the other

two statements.

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

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

which immediately implies (S.6).

We next turn to the proof of (S.5). As the intensity functions  $\lambda_i$  are Lipschitz continuous by (C2), it is straightforward to show that

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

From this, the uniform convergence result (S.8) and condition (C2), 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) is a straightforward consequence of (S.10) and (S.11).  $\square$

**Proof of (A.5).**  $\square$