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 $\mathcal{X}_i = \{X_{it} : 1 \leq 1 \leq T\}$ for n different countries i. A simple way to model the count data X_{it} is to use a Poisson distribution. For example, De Salazar et al. (2020) assume that the observed case count in country i follows a Poisson distribution with parameter λ_i being proportional to the daily air travel volume of the i-th location. In our paper, we relax the assumption of De Salazar et al. (2020) that λ_i stays constant over time. 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

$$X_{it} = \lambda_i\Big(\frac{t}{T}\Big) + u_{it} \tag{1.1}$$

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, $\lambda_i : [0,1] \to \mathbb{R}$ 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 $\mathcal{F} = \{\mathcal{I}_k \subseteq [0,1] : k = 1, \ldots, K\}$ 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.

Our method contributes to the literature on statistical tests for comparing nonparametric regression and trend curves. Examples include the tests of Härdle and Marron (1990), Hall and Hart (1990), King et al. (1991), Delgado (1993), Kulasekera (1995), Young and Bowman (1995), Munk and Dette (1998), 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 (2019). 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. (2009). 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. Sections 2 and 3 lay out the statistical methodology. In Section 2, we introduce the model setting in detail which underlies our analysis, while in Section 3, we develop the multiscale test step by step. To keep the presentation as clear as possible, the technical details are deferred to the Appendix and the Supplementary Material. Section 4 contains the empirical part of the paper. There, we apply our methods to a sample of current COVID-19 data. In addition, we run some simulation experiments to demonstrate that the multiscale test has the formal properties predicted by the theory. Even though our multiscale test is motivated by the current COVID-19 crisis, its applicability is by no means restricted to this specific issue. Indeed, it is a general method to compare nonparametric trends in epidemiological count data. In Section [number], we show that it is even useful beyond this epidemiological context, for example, in the context of financial volatility modelling. We briefly discuss this as well as some further extensions of our methods in Section [number].

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

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

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. 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 (McCullagh and Nelder, 1989; Efron, 1986) are frequently used. In our context, a quasi-Poisson model of X_{it} has the form

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

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.

Similar, but less general approach to modelling the number of new cases of was used in Pellis et al. (2020). They also consider quasi-Poisson model for the

observed data with a constant parameter of overdispersion, but they parametrically restrict the mean function to be exponentially growing with a constant growth rate. Tobías et al. (2020) analyze the data on the accumulated number of cases using quasi-Poisson regression with an interaction model to estimate the change in trend. However, both this approaches lacks the ability to capture day-to-day time variation, whereas our testing strategy allows us to do it without imposing any additional structure on the mean function λ_i.

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

(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 $0 \leq \lambda_{\min} \leq \min_{w\in[0,1]} \lambda_i(w) \leq \max_{w\in[0,1]} \lambda_i(w) \leq \lambda_{\max} < \infty$ 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 $\mathrm{Var}(X_{it}) = \sigma^2\lambda_i(t/T)$ 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 (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 $\mathcal{M} := \mathcal{S} \times \{1, \ldots, K\}$ 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; 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

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

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

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

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

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

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 $\hat{\sigma}^2 = |\mathcal{C}|^{-1} \sum_{i=\mathcal{C}} \hat{\sigma}_i^2$ 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

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

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

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

with $\overline{\lambda}_{ij}(u) = \{\lambda_i(u) + \lambda_j(u)\}/2$, 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

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

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. 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

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

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

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

where $\hat{\psi}^0_{ijk,T}$ 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 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

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

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

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

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 fol-

lowing 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

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

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

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

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 $\alpha \in (0,1)$:

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

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 $b_k = \sqrt{2\log(1/h_k)}$.

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 FWER(α) for each pre-specified significance level α. As shown in Corollary A.1, this has the following implication:

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

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

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

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

- 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

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

  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.

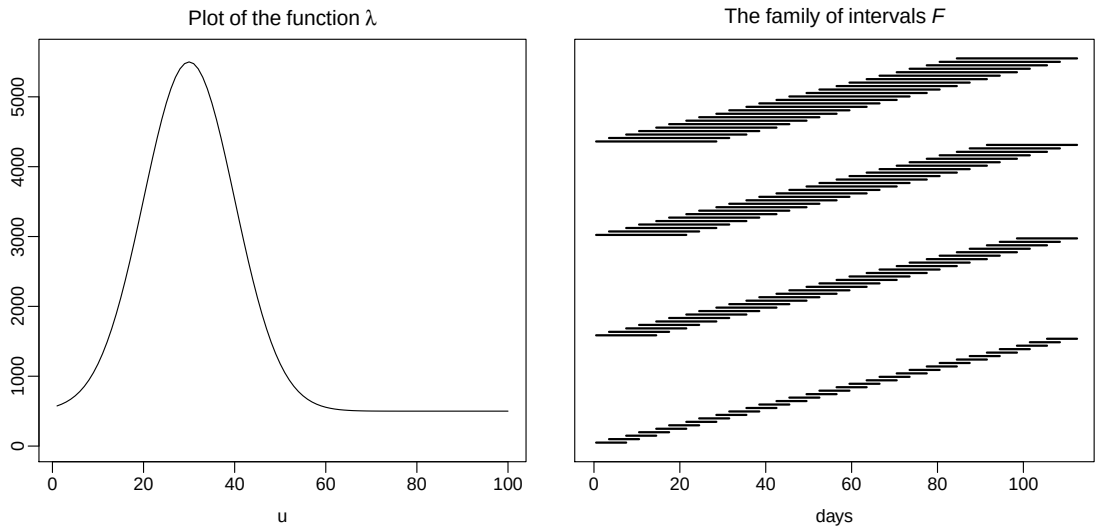
- We set the overdispersion parameter $\sigma = 15$ (the results of the robustness checks for $\sigma = 10$ and $\sigma = 20$ are provided in the Supplementary Material). The estimated values of σ in the application take similar values.

- 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 $j = 1, 2, \ldots$ A graphical presentation of the family $\mathcal{F}$ is given in Figure 1b.

We estimate the probability of Type I error (i.e. the test size) by calculating the percentage of at least one rejection of the null hypothesis $H_0^{(i,j,k)}$ for some (i, j, k) from $R = 5000$ simulated samples $\mathcal{X}$ of count data.

The results of the size simulations are presented in Table 1. As can be seen from the table, in all scenarios the empirical FWER coincides with the nominal size α. Moreover, size numbers are biased downwards, but this bias slowly diminished as the sample size or the number of time series increases. This slow performance increase can be explained by the following fact. For all the sample sizes considered we use the same family of intervals $\mathcal{F}$, hence, the length of the smallest interval (7 days, i.e. one week) relative to the rescaled time decreases, which leads to the effective sample size staying at the same level rather than increasing with the sample size. If we would let the length of the smallest interval be dependent on the sample size, this would probably diminish the bias in the empirical size compared to the nominal one. To summarize, (a) in finite samples, the test allows us to control empirical FWER(α) for pre-specified significance level α and (b) the test is somewhat conservative, but this issue may be overcome by using a different set of intervals that depends on the sample size.



(a) (b)

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	0.006	0.035	0.070	0.004	0.026	0.059	0.002	0.021	0.045
T = 250	0.010	0.047	0.099	0.008	0.042	0.084	0.006	0.034	0.067
T = 500	0.011	0.052	0.090	0.011	0.046	0.089	0.011	0.040	0.074

4.2 Analysis of COVID-19 data

Currently, the analysis of the ongoing global pandemic of coronavirus disease 2019 (COVID-19) is of the the utmost importance in many fields of science. One of the most relevant questions for governments, policy makers, economists and even general public, is the differences in the development of the pandemic in various countries.

One of the ways to address this question is to compare the underlying trends of the new cases of COVID-19 per day for different countries. Information on the regions where two trends are significantly different from each other can then be used in order to assess which government policies are more efficient and which are less efficient. In what follows, we use our multiscale test to compare the spread of COVID-19 in five European countries.

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 4.1.

4.2.1 Data

Epidemiological data for COVID-19 confirmed cases in Germany, Spain, France, Italy and United Kingdom was collected from the database created by European Center for Disease Prevention and Control (http://ecdc.europa.eu). The data was obtained on 26 June 2020. In order to be able to account for the differences in the onset of the COVID-19 pandemic in various countries, we take the day of the 100th case in each country as the starting point of the time series for that country. As the length of the dataset, we take the minimal number of days for which we have observations for all five countries. The resulting dataset consists of $n = 5$ time series, each with $T = 112$ observation (as of June 25). We then replaced all negative values by 0. Overall, this resulted in 4 replacements. Plots of the time series are given in panels (a) of Figure 2.

The dataset was then matched with the Government Response Index (GRI) from The Oxford COVID-19 Government Response Tracker (OxCGRT) (Hale et al. (2020b)). GRI is calculated based on several common policies that governments have taken as a response to the pandemic such as school closures and travel restrictions. GRI ranges from 0 to 100, with 0 corresponding to no response from the government 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 methodology of calculating indices is provided in Hale et al. (2020a). Plots of the time series of GRI are given in panels (c) of Figure 2.

4.2.2 Results

We assume that for each country $i \in \{1, \dots, 5\}$ the data X_{it} follows the nonparametric trend model

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

where the noise variables η_{it} have zero mean and unit variance. We estimate the overdispersion parameter σ by the procedure described in Section 3.1. This results in the estimator $\widehat{\sigma} = 14.49$.

As in Section 4.1, we let $\mathcal{S} = \{(i, j) : 1 \le i < j \le 5\}$, 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 $j = 1, 2, \dots$ Finally, with $K := |\mathcal{F}|$ being the cardinality of $\mathcal{F}$, we write $\mathcal{M} = \mathcal{S} \times \{1, \dots, K\}$.

With the help of our multiscale method, we now test the null hypotheses $H_0^{(i,j,k)}$

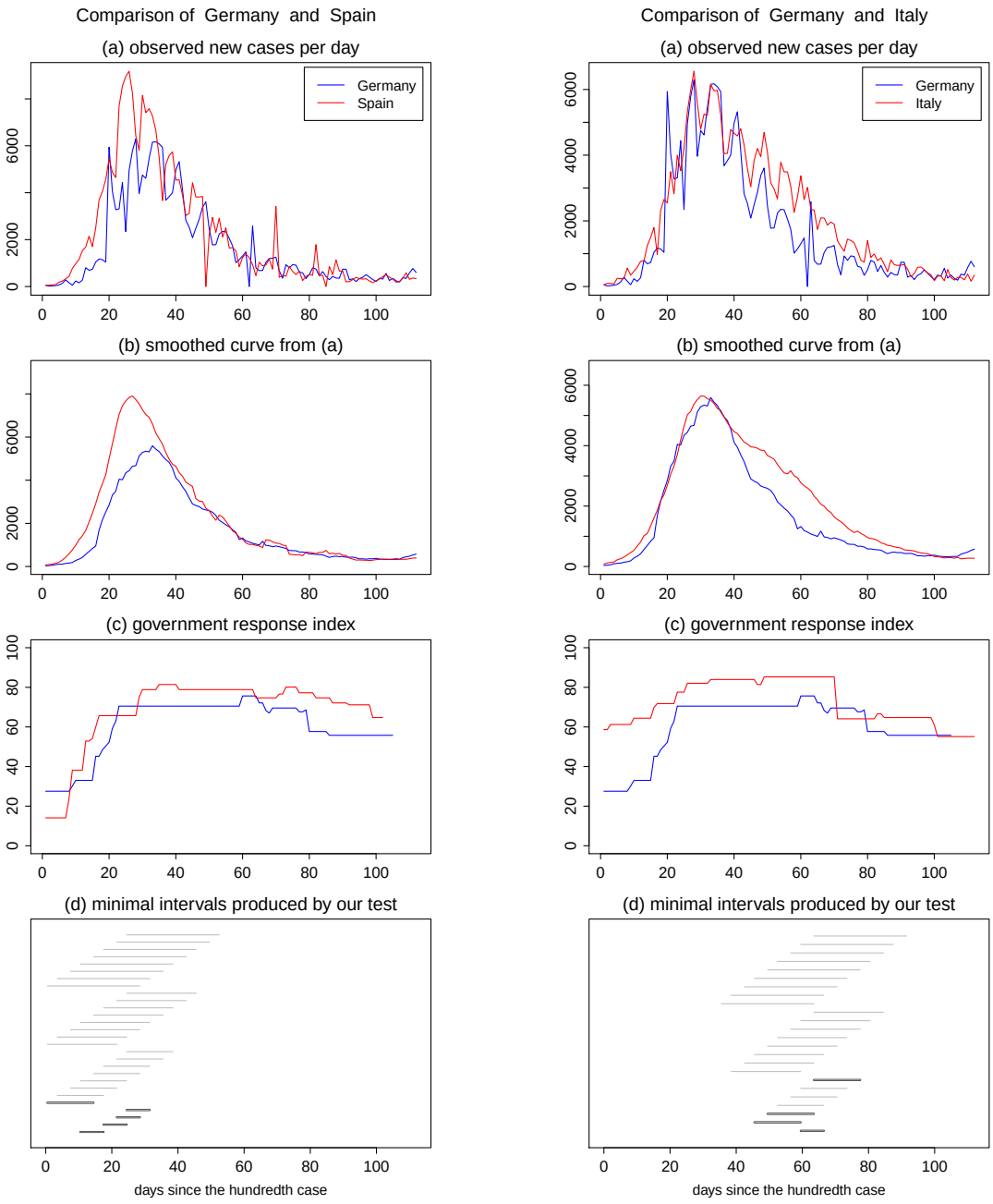
for each $(i, j, k) \in \mathcal{M}$ that $\lambda_i(\cdot) = \lambda_j(\cdot)$ on the interval $\mathcal{I}_k$. The results are presented in Figure 2. As was mentioned before, we produce a plot for each pair of countries. However, for the sake of brevity, we present here only the results of pairwise comparison between Germany and each one of the other four countries. The rest of the results can be found in the Supplementary Material.

On Figure 2, panels (a) show the observed time series for two countries that are being compared, while panels (b) shows the smoothed versions of the time series from (a) using the bandwidth of 7 days, panels (c) display the corresponding Government Response Index, and panels (d) present the results of out test procedure. Specifically, each panel (d) depicts in grey the set $\mathcal{F}_{reject}(i, j)$ of all the intervals $\mathcal{I}_k$ for which the test rejects the null $H_0^{(i,j,k)}$ where (i, j) is the pair of countries being compared. The corresponding set of minimal intervals $\mathcal{F}_{reject,minimal}(i, j)$ is depicted in black. According to (3.6), we can make the following simultaneous confidence statement about the intervals $\mathcal{F}_{reject}(i, j)$ plotted in panel (d). We can claim, with confidence of about 95%, that the functions λ_i and λ_j are different on each of the intervals.

As we can see from the plots, time series for these five countries are relatively similar to each other. For example, we do not detect any significant difference between trends in Germany and Spain after day 31 since the start of the pandemic, or between Germany and Italy before day 46 and after day 77, or between Germany and France after day 31. The only pairwise comparison that shows significance difference between functions λ_i and λ_j during the most part of the considered time period is the comparison between Germany and United Kingdom.

After visual inspection of the plots, it seems that the level of government response to the pandemic is not always the main driven force of the trend behind the spread of COVID-19. For example, consider the case of Germany and Italy. The GRI of Italy is almost always higher than GRI of Germany, but while the time trends of daily new infections in these countries do not differ in the beginning, starting from day 46 Italy tends to have more new cases of COVID-19 per day than Germany. This tendency nearly vanishes after day 77, but again, we can not attribute that to change of policies introduced by government. Most probably, there are other factors that affect the spread of COVID-19 in Germany and Italy.

But the severity of government measures can be an important factor in containing the pandemic. For example, consider the case of Germany and Spain. According to our test, there are significant differences in the onset of pandemic between Spain and Germany up to day 31, with Spain having on average more new cases of daily infections than Germany. But starting from day 31, the trends in these countries become similar and start to decrease with approximately the same rate. That may be due to the fact that in general Spain introduced more severe measures of lockdown than Germany (as can be seen on Figure 2, panel (c)), which helped battle the spread



of infection.

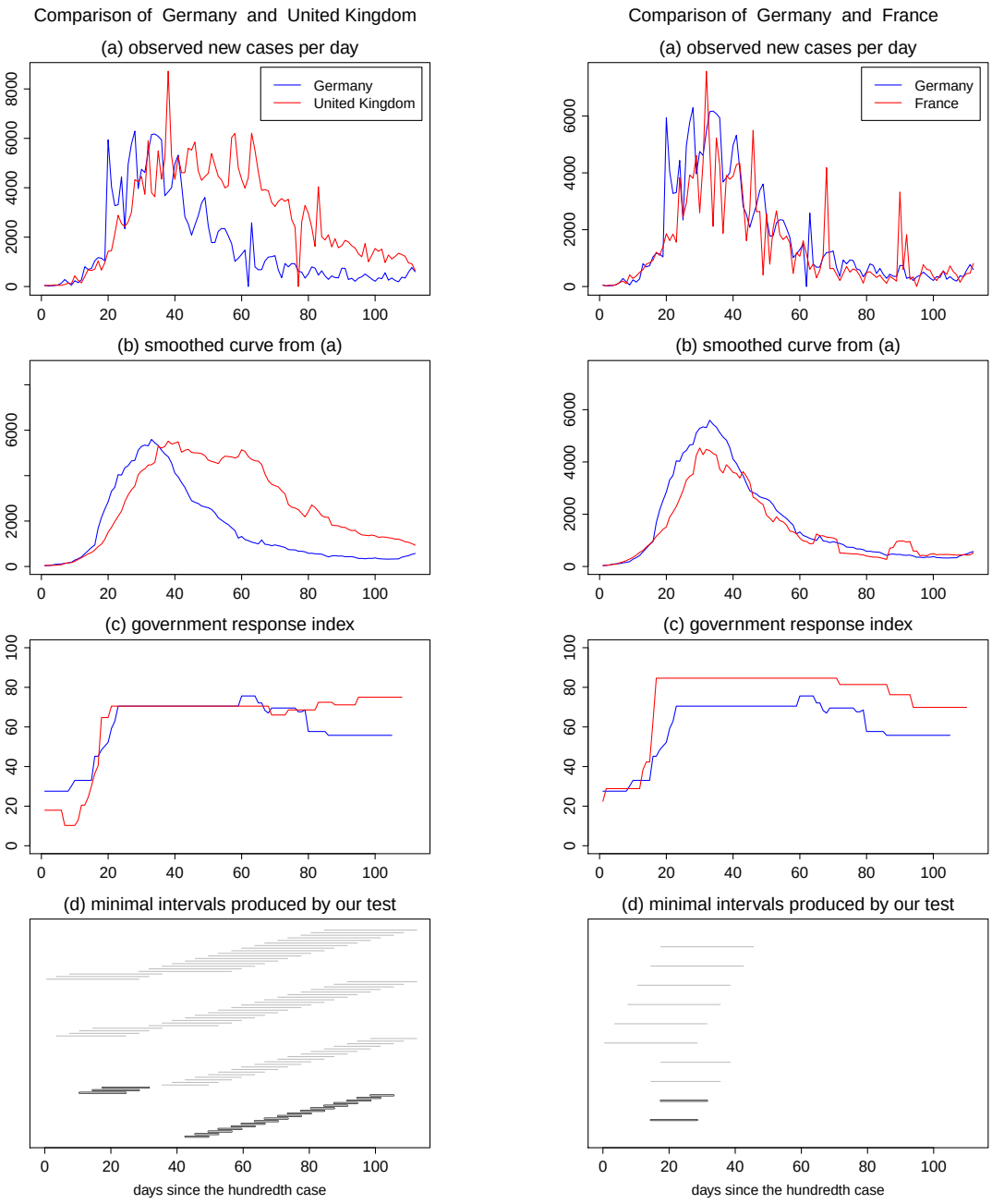


Figure 2: Summary of the results for the daily numbers of the new cases of COVID-19 in Germany, Spain, France, Italy and The United Kingdom with one subfigure corresponding to the results of pairwise comparison of Italy and one of the other countries. Panel (a) of each subfigure shows the two observed 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 shows the corresponding Government Response Index. Panel (d) of each subfigure depicts the set of intervals $\mathcal{F}_{\text{reject}}(i,j)$ in grey and the set of minimal intervals constructed on the basis of $\mathcal{F}_{\text{reject}}(i,j)$ in black.

4.2.3 Discussion

Direct comparison of the spread of COVID-19 infection is challenging in many ways. One of the most cited arguments against any possibility of reasonable cross-country comparison is different testing regimes. For example, UK mostly tested workers on the frontline and people who need to be hospitalized. On the other hand, Germany is often considered as a country with extensive early testing of contacts to known cases. This means one country's data may be fundamentally different than others, making direct comparison not . conclusive.

However, there are two reasons that our results are still valid. First, even visual inspection of the raw data confirms our finding that the underlying trends of the spread of infection are similar in different countries even despite possible differences in testing policies. Moreover, our test does not find any significant differences between countries on large parts of the time period, confirming our assumption that the trends of COVID-19 pandemic are comparable across countries.

Second, if we assume that the testing regime is country-specific and does not change over the time period considered, this does not affect our studies in any way. Even if the real number of infected people differ across the country, the trends either follow the same pattern and we do not detect any significant difference between countries; or the trends diverge on some part of the time period and we reject the null hypothesis for the corresponding interval. In both scenarios we do not aim to provide any causal interpretation, but merely reveal the discrepancies in the underlying deterministic trend.

It is worth to emphasise that our test, being mathematically rigorous, only allows researchers to determine the regions of interest where two trends differ from each other, but without telling what caused this difference. The proposed testing procedure provides the necessary basis for further explanatory analysis of the data, distinguishing significant differences in the structure of the data from the artifacts of the sampling noise. However, it is clear that in order to be able to identify the reason for these differences, more detailed analysis is required.

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* $\alpha \in (0,1)$*,*

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

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.*

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,*

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

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

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

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

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

To start with, we prove that

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

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, $\rho_T/r_T \to 0$ as $T \to \infty$. 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

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

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

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

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 $\psi^0_{ijk,T} = T^{-1/2}\sum_{t=1}^{T} V_t^{(ijk)}$ and thus

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

Analogously, we define

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

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

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

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

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

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 A.1. *Assume that*

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

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

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

Then

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

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

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

which in turn implies (A.2).

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

$$\sup_{q\in\mathbb{R}} \Big| \mathbb{P}\big(\hat{\Psi}_T \le q \big) - \mathbb{P}\big(\Phi_T \le q \big) \Big| = o(1). \tag{A.4}$$

To start with, the above supremum can be bounded by

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

Moreover,

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

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),

$$\begin{aligned}
&\sup_{q\in\mathbb{R}} \Big| \mathbb{P}\Big(\Phi_T \le q \pm r_T \Big) - \mathbb{P}\big(\Phi_T \le q \big) \Big| \\
&= \sup_{\boldsymbol{q}\in\mathbb{R}^p} \Big| \mathbb{P}\Big(\Big| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{W}_t \Big| \le \boldsymbol{q} \pm r_T \Big) - \mathbb{P}\Big(\Big| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{W}_t \Big| \le \boldsymbol{q} \Big) \Big| \le C r_T \sqrt{\log(2p)},
\end{aligned}$$

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

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

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

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

which in turn yields the statement of Theorem A.1. □

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

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

which gives the statement of Corollary A.1. □

References

Chen, L. and Wu, W. B. (2019). Testing for trends in high-dimensional time series. *Journal of the American Statistical Association*, **114** 869–881.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2017). Central limit theorems and bootstrap in high dimensions. *Annals of Probability*, **45** 2309–2352.

Cox, D. R. (1983). Some remarks on overdispersion. *Biometrika*, **70** 269–274.

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.

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.

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.

Härdle, W. and Marron, J. S. (1990). Semiparametric comparison of regression curves. *Annals of Statistics*, **18** 63–89.

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.

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.

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.

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

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 preprint arXiv:2004.00117*.

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.

Tobías, A., Valls, J., Satorra, P. and Tebé, C. (2020). Covid19-tracker: A shiny app to produce to produce comprehensive data visualization for sars-cov-2 epidemic in spain.

Young, S. G. and Bowman, A. W. (1995). Nonparametric analysis of covariance. *Biometrics*, **51** 920–931.

Zhang, Y., Su, L. and Phillips, P. C. B. (2012). Testing for common trends in semiparametric panel data models with fixed effects. *The Econometrics Journal*, **15** 56–100.

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*

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

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

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

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

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

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

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

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 (S.1), we can infer that

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

By similar but simpler arguments, we additionally get that

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

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. □

Proof of (A.1). Since

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

it suffices to prove that

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

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

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

where

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

With this notation, we can establish the bound

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

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

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

$$\max_{(i,j,k)\in\mathcal{M}} \big|\psi^{0}_{ijk,T}\big| = O_p\big(\sqrt{\log p}\big) \tag{S.6}$$

$$\big|\hat{\sigma}^2 - \sigma^2\big| = O_p\Big(\sqrt{\frac{\log p}{T}}\Big). \tag{S.7}$$

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

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

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 (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

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

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

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

and

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

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

Proof of (A.5). The proof is by contradiction. Suppose that (A.5) does not hold true, that is, $\mathbb{P}(\Phi_T \leq q_{T,\mathrm{Gauss}}(\alpha)) = 1 - \alpha + \xi$ for some $\xi > 0$. By Nazarov's inequality,

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

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

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

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

Table 2: Size of the test for different number of time series $n \in \{5, 10, 50\}$ for $\sigma = 10$.

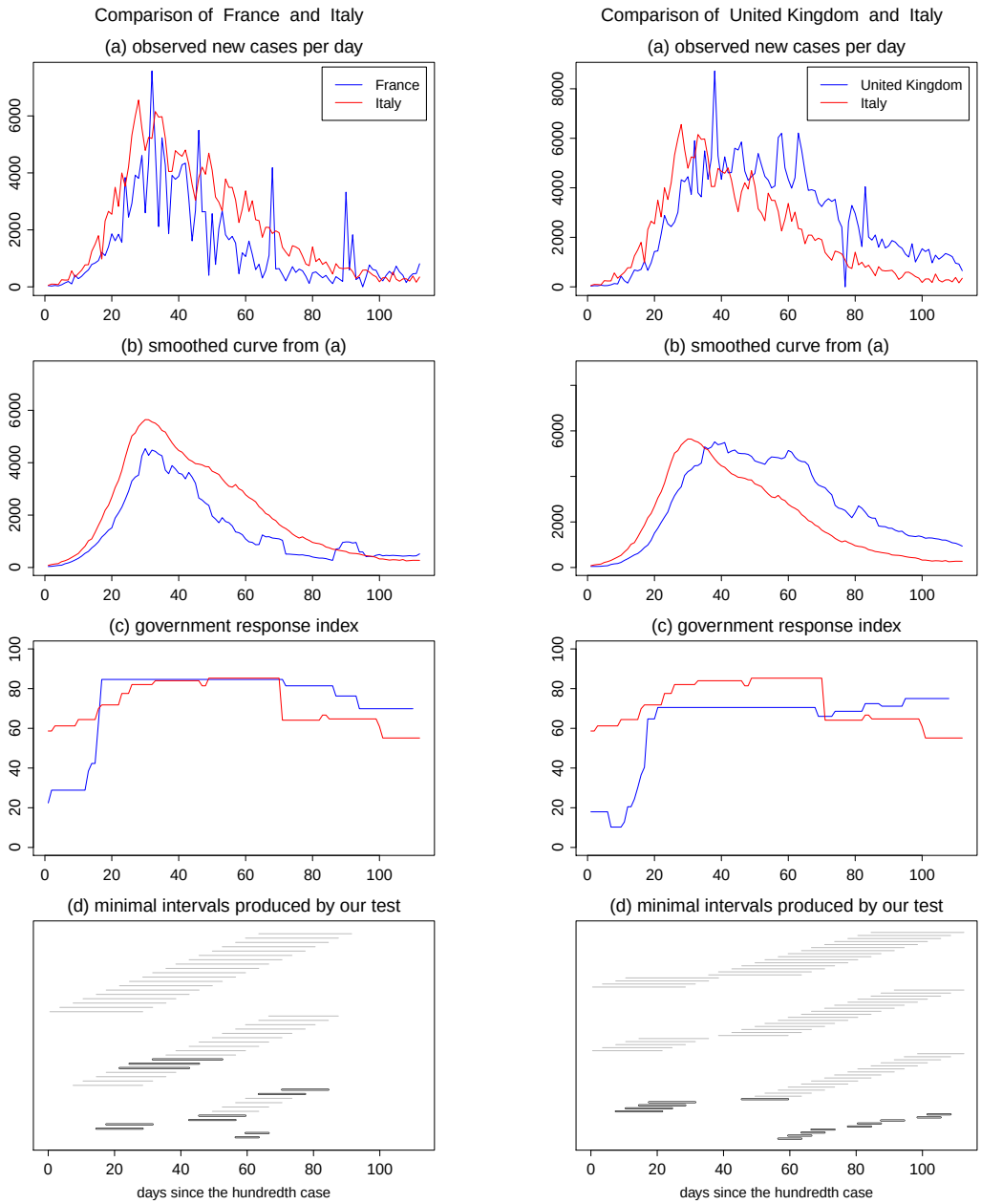
	$n = 5$			$n = 10$			$n = 50$		
	nominal size α			nominal size α			nominal size α		
	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
$T = 100$	0.009	0.034	0.076	0.008	0.029	0.063	0.003	0.019	0.044
$T = 250$	0.008	0.046	0.085	0.008	0.041	0.078	0.006	0.030	0.072
$T = 500$	0.010	0.045	0.091	0.005	0.044	0.085	0.011	0.042	0.075

Table 3: Size of the test for different number of time series $n \in \{5, 10, 50\}$ for $\sigma = 20$.

	$n = 5$			$n = 10$			$n = 50$		
	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
$T = 100$	0.010	0.045	0.087	0.010	0.037	0.074	0.005	0.033	0.060
$T = 250$	0.007	0.036	0.079	0.004	0.032	0.068	0.005	0.024	0.055
$T = 500$	0.007	0.037	0.072	0.007	0.033	0.067	0.004	0.019	0.048

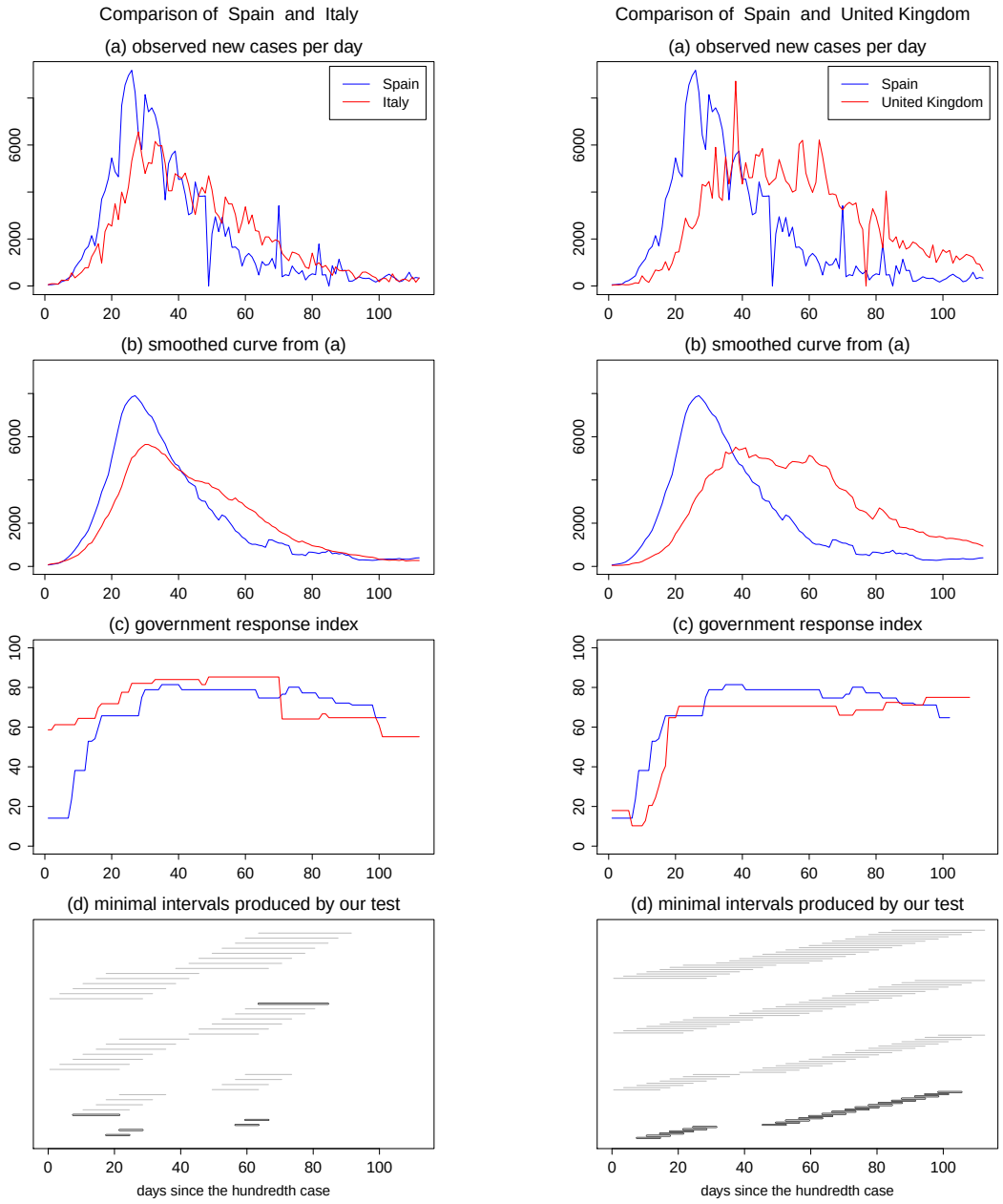
Robustness checks for Section 4.1

In what follows, we carry out some robustness checks to assess how sensitive the multiscale test is to the overdispersion parameter σ. To do so, we repeat the simulation exercises of Section 4.1 for $\sigma = 10$ and $\sigma = 20$. The results are presented in Tables 2 and 3.



Additional graphs for Section 4.2

In what follows, we provide the results of the analysis of COVID-19 data that was omitted in Section 4.2.


Comparison of Spain and Italy
(a) observed new cases per day
Spain
Italy
(b) smoothed curve from (a)
(c) government response index
(d) minimal intervals produced by our test
days since the hundredth case
Comparison of Spain and United Kingdom
(a) observed new cases per day
Spain
United Kingdom
(b) smoothed curve from (a)
(c) government response index
(d) minimal intervals produced by our test
days since the hundredth case

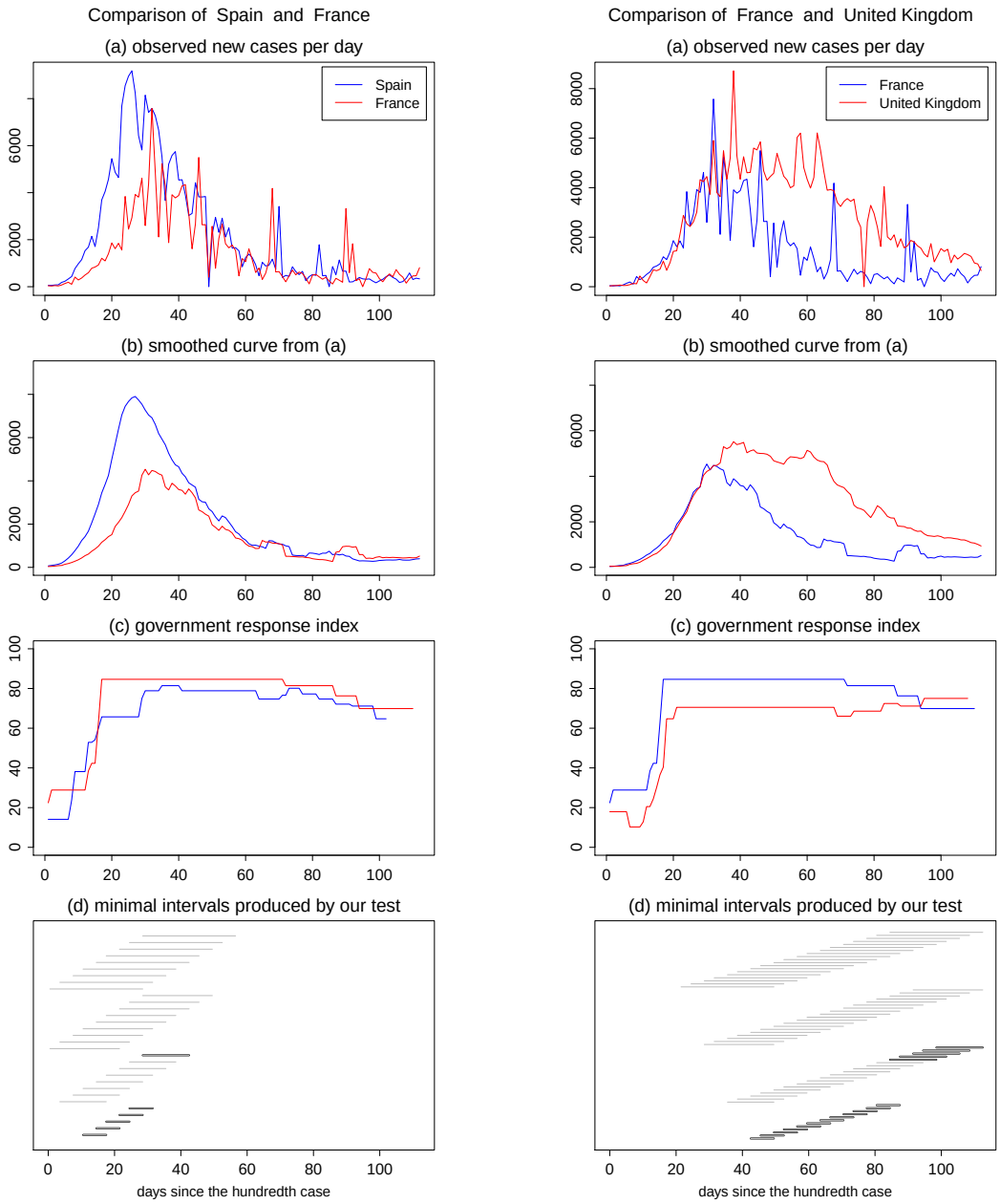


Figure 3: Summary of the results for the daily numbers of the new cases of COVID-19 in Germany, Spain, France, Italy and The United Kingdom with one subfigure corresponding to the results of one pairwise comparison of these countries (excluding comparison between Italy and other countries). Panel (a) 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 shows the corresponding time series of the Government Response Index (solid line) against the lagged version of itself (dashed line), with lag being equal to 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.