Revision of the paper
"Nonparametric comparison of epidemic time trends: the case of COVID-19"

First of all, we would like to thank the editor, the associate editor and the reviewers for their comments and suggestions which were very helpful in improving the paper. In the revision, we have addressed all comments and have rewritten the paper accordingly. Please find our point-by-point responses below. In addition to the improvements proposed by the referees, we have updated the application section in two ways:

(i) As more data on the corona pandemic have become available since the initial submission of the paper, we have slightly increased the time series length from $T = 139$ to the round number $T = 150$, which covers the first wave of the pandemic in all of the considered countries. As a robustness check, we have additionally run the test on longer time series of $T = 200$ days. The results of this robustness check are reported in Section S.4 of the Supplementary Material.

(ii) In the previous version of the paper, we took the starting date $t = 1$ to be the day of the 100th confirmed case in each country. In the revision, we have slightly changed the definition of the starting date $t = 1$ to be the first Monday after reaching 100 confirmed cases in each country. We have made this change (which was suggested to us by a seminar participant) since there is a weekly cycle in the data (presumably due to many unreported cases over the weekend which are reported with delay during the week). To eliminate possible effects of this weekly cycle on our test results, we have aligned the days of the week across countries by starting each time series on a Monday.

Comparing the old and the new data analysis in Section 4.2, one can see that the results are virtually the same for most countries. Only the pairwise comparisons including the UK are somewhat different. Consider e.g. Figure 6 in the previous and the revised version of the paper: Without the new Monday alignment (old version of Fig. 6), the test finds differences between the trends in Germany and the UK before day 40, whereas the test does not find any differences before that day with the new Monday alignment (new version of Fig. 6). This suggests that not aligning the data by the day of the week may produce some spurious differences across countries. Setting the starting date to be a Monday in each country should take care of this issue.

We hope you agree with us that this additional alignment makes a lot of sense. However, if you want us to, we are of course happy to go back to the previous version of the data analysis without this alignment.

Finally, we would like to mention that we have corrected the following minor mistake in the Appendix: we have replaced the technical condition $h_{\max} = o(1/\log T)$ in Theorem A.1 by the corrected version $h_{\max} = o(1/\{\log T\}^2)$. We now come to our point-by-point responses to the referees.

Reply to Referee 1

Thank you very much for the constructive and helpful comments. In our revision, we have addressed all of them. Please see our replies to your comments below.

(1) *Since the difference could be canceled out, why should one consider $\sum_t (X_{it} - X_{jt})\mathbf{1}_{t/T \in I_k}$? Isn't it more appropriate to use $|X_{it} - X_{jt}|$ or $|X_{it} - X_{jt}|^2$ to capture the distance?*

Consider the statistics $\hat{s}_{ijk,T}$ which underlie the test statistics $\hat{\psi}_{ijk,T}$ of our procedure. As argued in Section 3.1, the statistics $\hat{s}_{ijk,T}$ measure (local) mean distances between the trends λ_i and λ_j. More specifically, for each pair of countries (i,j) and each interval $\mathcal{I}_k$, the statistic

$$
\begin{aligned}
\hat{s}_{ijk,T} &= \frac{1}{Th_k} \sum_{t=1}^{T} \mathbf{1}\left(\frac{t}{T} \in \mathcal{I}_k\right)(X_{it} - X_{jt}) \\
&= \frac{1}{Th_k} \sum_{t=1}^{T} \mathbf{1}\left(\frac{t}{T} \in \mathcal{I}_k\right)\left(\lambda_i\left(\frac{t}{T}\right) - \lambda_j\left(\frac{t}{T}\right)\right) + o_p(1) \\
&= \frac{1}{h_k} \int_{\mathcal{I}_k} (\lambda_i(u) - \lambda_j(u))du + o_p(1)
\end{aligned}
$$

estimates the mean distance

$$
\Delta_{\mathrm{mean}}(\mathcal{I}_k) := \frac{1}{h_k} \int_{\mathcal{I}_k} (\lambda_i(u) - \lambda_j(u))du
$$

between λ_i and λ_j on the interval $\mathcal{I}_k$, where h_k is the length of the interval $\mathcal{I}_k$. The local mean differences $\Delta_{\mathrm{mean}}(\mathcal{I}_k)$ can be used as a distance measure in our test procedure for the following reason: Suppose the two functions λ_i and λ_j are continuous (as assumed in the paper). Then λ_i and λ_j differ from each other on their support $[0, 1]$ if and only if there exists a subinterval $\mathcal{I} \subseteq [0, 1]$ with $\Delta_{\mathrm{mean}}(\mathcal{I}) \neq 0$. Consequently, our test procedure, which checks whether $\Delta_{\mathrm{mean}}(\mathcal{I}_k) \neq 0$ for a large number of intervals $\mathcal{I}_k$ simultaneously, should be able to detect the differences between the functions λ_i and λ_j. This is reflected by the fact that our test has power against a large class of local alternatives. Please have a look at the new Proposition A.1 in the Appendix and its discussion in Section 3.3 for the details on the power of the test.

As you point out completely correctly, one may replace the statistics $\hat{s}_{ijk,T}$ that measure the local mean distances $\Delta_{\mathrm{mean}}(\mathcal{I}_k)$ by statistics that measure local L_q distances $\Delta_q(\mathcal{I}_k)$ of the form

$$\Delta_q(\mathcal{I}_k) = \frac{1}{h_k} \int_{\mathcal{I}_k} |\lambda_i(u) - \lambda_j(u)|^q du.$$

Even though this is possible in principle, we follow most other multiscale approaches in the literature which work with local (weighted) mean distances rather than local L_q distances. The reason is mainly technical: The theory in our paper would be quite different if we worked with local L_q distances. In particular, we could not make use of the Gaussian approximation results from Chernozhukov et al. (2017) as far as we can see.

We would finally like to note that local mean distances are used not only in the context of multiscale tests. There are also more classic test procedures which are based on local mean distances. As a simple example, suppose we want to test whether the regression function m in the model $Y_t = m(t/T) + \varepsilon_t$ is equal to zero. This can be achieved by a partial sum statistic which estimates local averages of the regression function m on a number of subintervals of the support of m; see e.g. Stute (1997) for the details.

(2) *Now the conclusion will be largely interfered by the choice of interval sets. I am wondering whether we can reach some unified result without the influence of such selection. That is whether we can aggregate the rejected intervals $\mathcal{I}_k$ and draw some meaningful conclusion?*

A simple way to aggregate the rejected intervals $\mathcal{I}_k$ for a given pair of countries (i, j) is to consider their union

$$V_{ij} = \bigcup_{\mathcal{I} \in \mathcal{F}_{\mathrm{reject}}(i,j)} \mathcal{I},$$

where, as in the paper, we let $\mathcal{F}_{\mathrm{reject}}(i, j)$ denote the set of rejected intervals for (i, j). The union V_{ij} is the time span where our test finds differences between the functions λ_i and λ_j. It thus gives a first rough idea about where the trends λ_i and λ_j differ from each other. There is however one issue with the union V_{ij}: If the family $\mathcal{F}$ contains very long intervals and the test finds a deviation from the null on such a long interval, the union V_{ij} will be a very large interval itself and will thus not be very informative about where exactly λ_i and λ_j differ. Hence, it may be better to consider the union

$$U_{ij} = \bigcup_{\mathcal{I} \in \mathcal{F}_{\mathrm{reject}}^{\mathrm{min}}(i,j)} \mathcal{I}$$

3

of minimal intervals, where $\mathcal{F}_{\text{reject}}^{\min}(i,j) = \{\mathcal{I}_k \in \mathcal{F}_{\text{reject}}(i,j) : \text{there exists no } \mathcal{I}_{k'} \in \mathcal{F}_{\text{reject}}(i,j) \text{ with } \mathcal{I}_{k'} \subset \mathcal{I}_k\}$ denotes the set of minimal intervals.

As shown in the new Lemma S.2 in the Supplementary Material, one can formally relate U_{ij} to the set of time points

$$U_{ij}^* = \{u \in [0,1] : \lambda_i(u) \neq \lambda_j(u)\}$$

where λ_i and λ_j differ from each other. In particular, under certain regularity conditions, we can prove the following result for a given pair of functions λ_i and λ_j:

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

where $\Delta(U_{ij}, U_{ij}^*) = \mathcal{L}\{(U_{ij} \setminus U_{ij}^*) \cup (U_{ij}^* \setminus U_{ij})\}$ is the Lebesgue measure $\mathcal{L}$ of the symmetric difference between the two sets U_{ij} and U_{ij}^* and ρ_T converges to 0 as $T \to \infty$. According to $(*)$, the difference between U_{ij} and U_{ij}^* is small $(\leq C\rho_T = o(1))$ with high probability $(\geq 1 - \alpha + o(1))$. In this sense, U_{ij} can be regarded as an approximation of U_{ij}^*.

We have added a brief summary of the above discussion to the end of Section 3.4 of the paper. The new Lemma S.2 and its proof can be found in the Supplement.

(3) *The author mentioned this method can be used to identify locations of changes in the trends. But the detail is not very clear to me. For example consider a very simple case: if the two series i, j differ from time t_1 to t_2 and are the same before and after this interval, where $t_1, t_2, t_2 - t_1$ are all unknown. Can we somehow able to identify this interval $[t_1, t_2]$ using our method and how well can we estimate t_1 and t_2? If one takes difference of each pair (i, j), and then the trends are zero except some unknown intervals. Then the task is to detect those unknown intervals. Such problem can be possibly solved by for example MOSUM. Can author comments about this?*

In what follows, we discuss (i) in which sense our method allows to identify and locate differences between the trends λ_i and λ_j and (ii) in which sense it does not allow to do so. In our discussion, we consider the simple setting described in your comment where λ_i and λ_j differ at any time point $t \in (t_1, t_2)$ but are identical at any other time point.

(i) Our method allows to identify differences between the trends λ_i and λ_j in the sense that we can make confidence statements about where (that is, in which time intervals $\mathcal{I}_k$ under consideration) the differences are. In particular, as discussed in more detail in Section 3.3 of the paper, we can make the following confidence statement:

With (asymptotic) probability at least $1 - \alpha$, the trends λ_i and λ_j are different on each interval $\mathcal{I}_k$ for which our test rejects the null $H_0^{(ijk)}$.

In the simple setting under consideration, this can be rephrased as follows:

With (asymptotic) probability at least $1 - \alpha$, each interval $\mathcal{I}_k$ for which our test rejects the null $H_0^{(ijk)}$ has some overlap with the interval (t_1, t_2).

Hence, the intervals in $\mathcal{F}_{\text{reject}}(i, j)$ for which our test rejects the null give information about the interval (t_1, t_2) where λ_i and λ_j differ from each other. To summarize the test results, we thus propose to plot the family of intervals $\mathcal{F}_{\text{reject}}(i, j)$.

(ii) Our approach does not identify and locate differences between λ_i and λ_j in the sense that it produces a consistent point estimate of the interval (t_1, t_2). Nevertheless, we conjecture that it is possible to construct a point estimate of (t_1, t_2) based on our approach. More specifically, it should be possible to extend the result $(*)$ discussed in our reply to your previous comment to the case where $\alpha = \alpha_T \to 0$ sufficiently slowly. Hence, we should be able to show that

$$\mathbb{P}\Big(\Delta(U_{ij}, U_{ij}^*) \leq C\rho_T\Big) \geq \underbrace{1 - \alpha_T - o(1)}_{=1-o(1)}. \qquad (**)$$

This says that U_{ij} is a consistent estimator of $U_{ij}^* = (t_1, t_2)$ in the sense that the difference $\Delta(U_{ij}, U_{ij}^*)$ goes to 0 (note that $\rho_T \to 0$) with probability tending to 1.

Since we are primarily interested in inference rather than point estimation in the paper and since the statement $(**)$ is merely a conjecture not covered by our theory, we have decided not to discuss this extension in the paper. However, we are happy to do so if you think this is needed.

(4) *Some theory question*

(a) *Since the result in Chernozhukov et al's Gaussian approximation (GA) does not require the series to be independent cross sectionally, I wonder does that mean the current result can be extended to data with cross-sectional dependence?*

In principle, it should be possible to extend our theory to the case of cross-sectional dependence. As far as we can see, our proof strategy should still go through in this case (even though the technical arguments would of course become more complicated and would need some tweaks here and there). In particular, as you point out, we could still use the Gaussian approximation results of Chernozhukov and coauthors since these do not require cross-sectional independence.

It is important to note that we would not only have to adapt the theory in order to deal with cross-sectional dependence. We would also have to adjust the test procedure itself: Assume that the error terms η_{it} in our model are dependent across i (i.e. cross-sectionally dependent) but independent across t. In this case, the variance of the statistic $\hat{s}_{ijk,T}$ is given by

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

where the term

$$C_{ij,T} = 2\sqrt{\lambda_i\left(\frac{t}{T}\right)}\sqrt{\lambda_j\left(\frac{t}{T}\right)}\mathbb{E}[\eta_{it}\eta_{jt}]$$

reflects the cross-sectional dependence in the data. (In the case of no cross-sectional dependence, $C_{ij,T} = 0$.) To construct the test statistic $\hat{\psi}_{ijk,T}$ of the null hypothesis $H_0^{(ijk)}$, we require an estimator of $\text{Var}(\hat{s}_{ijk,T})$. In order to obtain such an estimator in the case of cross-sectional dependence, we need to estimate the additional term $C_{ij,T}$, which requires us to come up with an estimator of the covariance $\mathbb{E}[\eta_{it}\eta_{jt}] = \text{Cov}(\eta_{it}, \eta_{jt})$.

We have added a remark to Section 2 of the paper (see the second bullet point on p.5) which states that it is in principle possible to allow for cross-sectional dependencies in the data.

(b) *It would be better if the author can derive power under certain alternatives, so that one can get a better idea as how different the trends needs to be in order to be detected.*

As suggested, we have derived the power of the test against a large class of local alternatives. Please see the new Proposition A.1 and its proof in the Appendix as well as the discussion of Proposition A.1 in Section 3.3.

(c) *The argument about no need for time dependent data is reasonable, just a short comment: there already exists result extending Chernozhukov et al's GA to time dependent case, maybe this paper can be further extended to time dependent data as well.*

We are aware of extensions of the Gaussian approximation results in Chernozhukov et al. (2013) to the time-dependent case. However, we are not aware of any such extensions of the more general Gaussian approximation results in Chernozhukov et al. (2017), on which our theory is based.

More specifically, let $x_t = (x_{t,1}, \ldots, x_{t,p})^\top$ be p-dimensional random vectors for $t = 1, \ldots, T$ with $\mathbb{E}[x_t] = 0$ and covariance matrix $\mathbb{E}[x_t x_t^\top]$ and consider the

statistics $X = (X_1, \ldots, X_p)^\top$ with

$$X_k = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_{t,k}.$$

Moreover, let $y_t = (y_{t,1}, \ldots, y_{t,p})^\top$ be Gaussian versions of x_t in the sense that $y_t \sim N(0, \mathbb{E}[x_t x_t^\top])$ and define $Y = (Y_1, \ldots, Y_p)^\top$ with

$$Y_k = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_{t,k}.$$

The results in Chernozhukov et al. (2013) allow to bound the distance

$$\sup_{u \in \mathbb{R}} \left| \mathbb{P}\left(\max_{1 \leq k \leq p} X_k \leq u \right) - \mathbb{P}\left(\max_{1 \leq k \leq p} Y_k \leq u \right) \right|$$

under the assumption that the variables x_t are independent across t. Generalizations to the case of dependent x_t's can be found e.g. in Zhang and Wu (2017) and Zhang and Cheng (2018). The more general results in Chernozhukov et al. (2017) allow to bound the distance

$$\sup_{A \in \mathcal{A}} \left| \mathbb{P}\left(\max_{1 \leq k \leq p} X_k \in A \right) - \mathbb{P}\left(\max_{1 \leq k \leq p} Y_k \in A \right) \right|$$

under the assumption that the variables x_t are independent across t, where $\mathcal{A}$ is the class of hyperrectangles in $\mathbb{R}^p$. This is the type of approximation result on which our proofs are based. As already mentioned above, we are not aware of any extension of this type of result to the time-dependent case. (If you know of any such extension, we'd be grateful for the exact reference. With such an extension, we may be able to generalize our results to the time-dependent case.)

(5) (a) *The specific allowance of $p = |W|$, which is essential in high dimensional analysis, is not mentioned until appendix. Please put them forward in the main context to provide some guidance in application.*

As you propose, we have added more details on the dimension p to the beginning of Section 3 (see the last few sentences before Subsection 3.1). In particular, we briefly outline the conditions on the growth of p in terms of the time series length T. However, since the precise conditions on the growth of p cannot be stated without reference to the technical assumptions in the Appendix, we have deferred the mathematically exact formulation of the conditions on p to the Appendix (see Theorem A.1 and the discussion thereafter on p.24/5).

(b) *Also since the convergence speed of Gaussian approximation depends on T, p, it would be better to keep the bound in terms of those parameters, so that we know how large the sample size we need in order to obtain the desired accuracy.*

We are not completely sure whether we fully understand your comment. Do you suggest formulating the bounds in the Appendix in terms of p and T only? That is, do you suggest replacing the quantities $h_{\min}$ and $h_{\max}$ in some of the bounds by p and T? Even though this is of course possible, we are a bit reluctant to do so: The bounds are formulated such that they explicitly give the interplay between the time series length T, the number of test problems p, the minimal interval length $h_{\min}$ and the maximal interval length $h_{\max}$. By imposing restrictions on $h_{\min}$ and $h_{\max}$ in terms of T (in particular, the restrictions from Theorem A.1 that $h_{\max} = o(1/\{\log T\}^2)$ and $h_{\min} \geq CT^{-b}$ for some $b \in (0, 1)$), we ensure that the bounds converge to 0 sufficiently fast. We could of course plug the restrictions on $h_{\min}$ and $h_{\max}$ directly into the bounds. However, this would make them less informative in our opinion. For this reason, we have decided not to rewrite the bounds. We hope you are fine with this decision.

Reply to Referee 2

Thank you very much for the constructive and useful comments. In our revision, we have addressed all of them. Here are our point-by-point responses to your comments.

(1) *The assumption of independence across countries may be debatable, but it seems that in the context of the model, this could be tested, so this may be worth mentioning.*

As you propose, we briefly mention that it is in principle possible to test this assumption and we give some references to tests for independence. Please see the third bullet point on p.5 in Section 2 of the revised paper.

(2) *Some arguments may be worth further details in the text. For instance, the equation involving $\hat{s}_{ijk,T}/\sqrt{Th_k}$ on Page 2 Line 7, or the bound for $|r_{it}|$ on Page 2 Line -5.*

As suggested, we have added further details on the statistic $\hat{s}_{ijk,T}$ and the bound for $|r_{it}|$ to the text. Please see the revised Section 3.1. We hope you find the changes appropriate.

(3) *I am unsure why the statistic in (3.2) is introduced, I feel the discussion in Pages 8-9 could be done without referring to it.*

The statistic $\hat{\psi}^0_{ijk,T}$ introduced in (3.2) is a modification of the test statistic $\hat{\psi}_{ijk,T}$. It is needed to define the critical values $c_T(\alpha, h_k) = b_k + q_T(\alpha)/a_k$ of the multiscale

test in Section 3.2. Specifically, $\hat{\psi}^0_{ijk,T}$ is required to define the $(1-\alpha)$-quantile $q_T(\alpha)$ of

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

For this reason, we have decided not to defer its definition to the Appendix but to introduce it in Section 3 as in the old version of the paper. Otherwise, an important detail of the test would be missing in the main text. We hope you are fine with this. To emphasize that the statistic $\hat{\psi}^0_{ijk,T}$ is required for the definition of the critical values of the multiscale test, we have added the following sentence after its introduction in equation (3.2) (which is equation (3.4) in the revised paper): "This auxiliary statistic is needed to define the critical values of our multiscale test in what follows".

(4) *The "cp." abbreviation is uncommon, I feel it should be replaced by "see" or "e.g."*

As suggested, we have replaced the "cp." abbreviation by "e.g." / "see".

References

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Annals of Statistics*, **41** 2786–2819.

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

STUTE, W. (1997). Nonparametric model checks for regression. *Annals of Statistics*, **25** 613–641.

ZHANG, D. and WU, W. B. (2017). Gaussian approximation for high dimensional time series. *Annals of Statistics*, **45** 1895–1919.

ZHANG, X. and CHENG, G. (2018). Gaussian approximation for high dimensional vector under physical dependence. *Bernoulli*, **24** 2640–2675.