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. Before we reply to the specific comments of the referees, we summarize the major changes in the revision.

Generalization of the theoretical results. We have extended the theoretical results as suggested by Referee 1:

(i) We have derived the following result for the asymptotic power of our test:

Let the conditions of Theorem A.1 be satisfied and consider two sequences of functions $\lambda_{i,T}$ and $\lambda_{j,T}$ with the following property: There exists $\mathcal{I}_k \in \mathcal{F}$ such that

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

where $\{c_T\}$ is any sequence of positive numbers with $c_T \to \infty$ faster than $\frac{\sqrt{\log T}\sqrt{\log\log T}}{\log\log\log T}$. We denote the set of triplets $(i, j, k) \in \mathcal{M}$ for which (1) holds true as $\mathcal{M}_1$. Then

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

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

This result is stated as Corollary A.2 on p.?? of the revised paper.

(ii)

Application. We have improved the application section in the following way:

(i) In order to make the data even more comparable across countries, we take the starting date $t = 1$ to be the first Monday after reaching 100 confirmed case in each country. Such alignment of the data by starting on Monday takes into account possible differences in reporting the numbers on a weekly level. As a robustness check, we perform our analysis on the time series without the alignment, where we take the starting date $t = 1$ to be the first day after reaching 100 confirmed case in each country, and we report the results of the robustness check in Section S.3 in the Supplement.

(ii) We have extended the considered time period from $T = 139$ to $T = 150$ and since now the data are available for longer time period, we perform the robustness check for longer time series with $T = 200$ days and report the results in Section S.4 in the Supplement.

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?*

As you point out, our test statistics $\hat{s}_{ijk,T}$ measure (local) mean distances between the functions λ_i and λ_j. More specifically, for each pair of countries (i, j) and each interval $\mathcal{I}_k$, the statistic

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

estimates the mean distance

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

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 $\Delta_{\text{mean}}(\mathcal{I}_k)$ can be used as a distance measure 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 $\Delta_{\text{mean}}(\mathcal{I}) \neq 0$. Consequently, our test procedure, which checks whether $\Delta_{\text{mean}}(\mathcal{I}_k) \neq 0$ for a large number of intervals $\mathcal{I}_k$ simultaneously, should be able to detect the differences between the functions λ_i and λ_i. This is reflected in the good power properties of the test which are stated in the new Proposition ??.

As you point out completely correctly, one may replace the statistics $\hat{s}_{ijk,T}$ that measure the local mean distances $\Delta_{\text{mean}}(\mathcal{I}_k)$ 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 (see e.g. ??, ??, ?? and many others) which work with local mean distances rather than local L_q distances. The reason is mainly technical: The theory in our paper would be very different if we worked with local L_q distances. In particular, we could not make use of the Gaussian approximation results from Chernozhukov ?? as far as we can see.

We have added some sentences to p.?? of the revision which summarize the discussion above.

(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 I_k and draw some meaningful conclusion?*

One way to aggregate the rejected intervals I_k is to consider their union. More specifically, one may consider the following quantity: Using the notation from the paper, we let $\mathcal{F}_{\text{reject}}(i,j)$ be the set of rejected intervals for a given pair of countries (i,j) and additionally define

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

The elements of $\mathcal{F}_{\text{reject}}^{\min}(i,j)$ are called minimal intervals. By definition, there is no other interval $\mathcal{I}_{k'}$ in $\mathcal{F}_{\text{reject}}(i,j)$ which is a proper subset of a minimal interval $\mathcal{I}_k$. We now consider the union of minimal intervals

$$\hat{U}_{ij} = \bigcup_{I \in \mathcal{F}_{\text{reject}}^{\min}(i,j)} I.$$

As shown in the new Corollary ?? in the Supplementary Material, $\hat{U}_{ij}$ is closely related to the set

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

that is, to the set of time points where λ_i and λ_j differ from each other. Under appropriate regularity conditions (as detailed in Corollary ??), one can in particular prove that

$$\mathbb{P}\Big(\Delta(U_{ij}, \hat{U}_{ij}) \leq \nu_T\Big) \geq 1 - \alpha + o(1), \qquad (*)$$

where $\Delta(U_{ij}, \hat{U}_{ij}) = (U_{ij} \setminus \hat{U}_{ij}) \cup (\hat{U}_{ij} \setminus U_{ij})$ is the symmetric difference between the two sets U_{ij} and $\hat{U}_{ij}$ and $\nu_T \to 0$ as $T \to \infty$. This says that the difference between U_{ij} and $\hat{U}_{ij}$ is small ($\leq \nu_T \to 0$) with high probability ($\geq 1 - \alpha + o(1)$). In this sense, $\hat{U}_{ij}$ can be regarded as an approximation of U_{ij}.

We have added a summary of the above discussion to Section ?? of the paper. The new Corollary ?? and its proof can be found in the Supplementary Material.

(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?*

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 I_k under consideration) the differences are. In particular, in the simple case 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 t, 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 $H_0^{(ijk)}$.

Put differently:

> 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 (t_1, t_2).

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

Our approach being an inference procedure, it allows to make confidence statements about where λ_i and λ_j differ, but it does not produce a point estimate of the time interval $[t_1, t_2]$. However, 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

$$\mathbb{P}\Big(\Delta(U_{ij}, \hat{U}_{ij}) \leq \nu_T\Big) \geq \underbrace{1 - \alpha_T - o(1)}_{=1-o(1)}. \tag{2}$$

This says that $\hat{U}_{ij}$ is a consistent estimator of $U_{ij} = [t_1, t_2]$ in the sense that the difference $\Delta(U_{ij}, \hat{U}_{ij})$ goes to zero (as $\rho_T \to 0$) with probability tending to 1.

Since we are primarily interested in inference rather than point estimation in our paper and since the statement (2) 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*

   (i) *Since the result in Chernozukov 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?*

   (ii) *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 properties of the test against a certain class of local alternatives. Please see the new Corollary ?? in the Appendix of the paper. The proof is provided in the Supplementary Material.

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

(5) *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. 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.*

Reply to Referee 2

Thank you very much for the constructive and useful suggestions. 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.*

(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 ?? for the details. 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 ??. 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 (3.2) as in the old version of the paper. Otherwise, an important detail of the test would be missing in the main text / would be hidden in the Appendix. 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): "This 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." or "see".