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 extended 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 additionally run our test on longer time series of $T = 200$ days as a robustness check. (Why $T = 200$? Explanation needed?) The results are reported in Section S.4 of the Supplementary Material. (We have also changed the time period from $T = 139$ to $T = 150$ in the application. The results are virtually the same, right? Maybe we should stick to $T = 139$ after all. I guess we shouldn't change too many things that are not explicitly asked for.)

(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 strong weekly cycle in the data (presumably due to many unreported cases over the weekend which are reported with delay on Monday). To eliminate possible effects of this weekly cycle, we have aligned the days of the week across countries by starting each time series on a Monday.

If one compares the old and new data analysis in Section ??, one can see that the test results are essentially unchanged for most countries. Only the pairwise comparisons including the UK are somewhat different. Consider e.g. Figures S.2 in the old and the revised supplement: Without the new Monday alignment (old version of Fig. S.2), the raw data in the UK and Italy up to approx. day 40 appear to have a somewhat different trending behaviour which is picked up by our test. With the Monday alignment (new version of Fig. S.2), a different picture arise. Now the raw data look much more similar over the first 40 days and our test does not find any deviation from the null up to day 40. 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 is very natural and makes a lot of sense. If not, we are of course happy to switch back to the previous version of the data analysis without this alignment.

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?*

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 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.

(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 define the set of minimal intervals by

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

As shown in the new Lemma ?? in the Supplementary Material, the union of minimal intervals

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

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 Lemma ?? 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?

As you suggest, we discuss the simple setting 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 I_k under consideration) the differences are. In particular, 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 in $\mathcal{F}_{\text{reject}}(i,j)$ for which our test rejects 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 $\mathcal{F}_{\text{reject}}(i,j)$.

(ii) Our approach does not identify 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

$$\mathbb{P}\Big(\Delta(U_{ij}, \hat{U}_{ij}) \le \nu_T\Big) \ge \underbrace{1 - \alpha_T - o(1)}_{=1-o(1)}. \qquad (**)$$

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 (note that $\nu_T \to 0$) 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 of 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 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?*

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 goes 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 dependence.

It is important to note that we would not only have to adapt the theory 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

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

where the term

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

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 in turn 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 footnote to p.?? of the paper which briefly mentions 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 properties of the test against a certain class of local alternatives. Please see the new Proposition ?? in the Appendix of the paper. The proof is provided in the Supplementary Material.

(c) *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.*

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.

Specifically, let $x_t = (x_{t,1}, \dots, x_{t,p})^\top$ be p-dimensional random vectors for $t = 1, \dots, T$ with $\mathbb{E}[x_t] = 0$ and covariance matrix $\Sigma = \mathbb{E}[x_t x_t^\top]$ and consider the statistics $X_t = (X_{t,1}, \dots, X_{t,p})^\top$ with

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

Moreover, let $y_t = (y_{t,1}, \dots, y_{t,p})^\top$ be Gaussian versions of x_t in the sense that $y_t \sim N(0, \Sigma)$ and define $Y_t = (Y_{t,1}, \dots, Y_{t,p})^\top$ with

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

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

$$\sup_{u \in \mathbb{R}} \Big| \mathbb{P}\big(\max_{1 \le k \le p} X_{t,k} \le u \big) - \mathbb{P}\big(\max_{1 \le k \le p} Y_{t,k} \le u \big) \Big|$$

in the case that the variables X_t are independent across t. Generalizations to the case that the X_t's are dependent across t 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

$$\sup_{A \in \mathcal{A}} \Big| \mathbb{P}\big(\max_{1 \le k \le p} X_{t,k} \in A \big) - \mathbb{P}\big(\max_{1 \le k \le p} Y_{t,k} \in A \big) \Big|$$

in the case 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 could most probably generalize our results to the time dependent case.)

(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.

No idea what's meant here. Do you have any idea?

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.*

Check Chapter 12.4.1 on "Testing Independence between Two Random Variables" in Li & Racine, Nonparametric Econometrics for tests on independence. Marina, can you add the references for the tests discussed there? I don't have the book ...

(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".

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.

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.