

# 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

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

where  $\{c_T\}$  is any sequence of positive numbers with  $c_T \rightarrow \infty$  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

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

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  $\sum_t (X_{it} - X_{jt}) \mathbf{1}_{t/T \in \mathcal{I}_k}$ ? 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  $\lambda_i$  and  $\lambda_j$ . More specifically, for each pair of countries  $(i, j)$  and each interval  $\mathcal{I}_k$ , the statistic

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

estimates the mean distance between  $\lambda_i$  and  $\lambda_j$  on the interval  $\mathcal{I}_k$ , which is given by the term

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

with  $h_k$  being the length of the interval  $\mathcal{I}_k$ .

If we only considered one interval  $\mathcal{I} \subseteq [0, 1]$  (or only a small number of intervals), it would of course not make much sense to use  $\Delta_{\text{mean}}$  as a distance measure. Suppose, for example, that we only consider the interval  $\mathcal{I} = [0, 1]$ . In this case, the mean distance

$$\Delta_{\text{mean}}(\mathcal{I}) = \frac{1}{|\mathcal{I}|} \int_{\mathcal{I}} (\lambda_i(u) - \lambda_j(u)) du = \int_0^1 (\lambda_i(u) - \lambda_j(u)) du$$

does not give much information on whether the two functions  $\lambda_i$  and  $\lambda_j$  are different or not: it only allows to say whether they have the same mean. Hence, other distance measures such as the  $L_q$  distance

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

(with, e.g.,  $q = 1$  or  $q = 2$ ) would be more appropriate.

Importantly, however, we do not only consider one interval but a large family of intervals  $\{\mathcal{I}_k : 1 \leq k \leq K\}$ . The local mean differences  $\Delta_{\text{mean}}(\mathcal{I}_k)$  are an

appropriate distance measure in this situation for the following reason: Suppose the two functions  $\lambda_i$  and  $\lambda_j$  are continuous (which is assumed in the paper). Then  $\lambda_i$  and  $\lambda_j$  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, if the family  $\{\mathcal{I}_k : 1 \leq k \leq K\}$  of intervals is rich enough, then we are able to detect the differences between the functions  $\lambda_i$  and  $\lambda_j$ . This is reflected in the good power properties of the test which are stated in the new Proposition ??.

You are of course right that we could replace the local mean distances  $\Delta_{\text{mean}}(\mathcal{I}_k)$  by local  $L_q$  distances  $\Delta_q(\mathcal{I}_k)$  as defined above. However, the theory would be markedly different if we worked with local  $L_q$  distances. In particular, we could not base our proofs on the Gaussian approximation results from Chernozhukov ?. Hence, we stick to the local mean distances  $\Delta_{\text{mean}}(\mathcal{I}_k)$ . **On p.??, we have added some sentences which summarize the discussion from the last few paragraphs.**

- (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?*

Consider a specific pair of countries  $(i, j)$  and let  $\mathcal{F} = \{\mathcal{I}_k : 1 \leq k \leq K\}$  be a large family of intervals. Moreover, for a given significance level  $\alpha \in (0, 1)$ , let  $\mathcal{F}_{\text{reject}}^{\min}(i, j) \subseteq \mathcal{F}$  be the set of minimal intervals as defined in Section ?? of the paper. We could consider the union of intervals in  $\mathcal{F}_{\text{reject}}^{\min}(i, j)$ ,

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

which is a subset of  $[0, 1]$ . According to our theory, we can make the following confidence claim:

- (\*) *With (asymptotic) probability  $\geq 1 - \alpha$ , the functions  $\lambda_i$  and  $\lambda_j$  differ on the interval  $\hat{\mathcal{I}}$ .*

This confidence statement gives a simpler summary of the test results than the statement (3.7) in the paper according to which the following holds:

- (\*\*) *With (asymptotic) probability  $\geq 1 - \alpha$ , the functions  $\lambda_i$  and  $\lambda_j$  differ on any interval  $\mathcal{I} \in \mathcal{F}_{\text{reject}}^{\min}(i, j)$ .*

On the other hand, the statement (\*\*) is much more informative than (\*). In particular, it allows to pin down more precisely where the differences in the functions are. **We have added some discussion on the quantity  $\mathcal{I}$  and the corresponding confidence statement (\*) on p.?? in the revision.**

- (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 suggested by you, we discuss the following simple case: we consider a specific pair of countries  $(i, j)$  and suppose that the functions  $\lambda_i$  and  $\lambda_j$  differ at any time point  $t \in [t_1, t_2]$  but are identical at any other time point  $t$ . Moreover, to simplify the discussion, we assume the following: the interval  $[t_1, t_2]$  is fixed (that is, does not change with the sample size  $n$ ), the functions  $\lambda_i$  and  $\lambda_j$  do not depend on  $n$  (that is, we consider a fixed rather than a local alternative), they are continuous and  $\lambda_i(t) > \lambda_j(t)$  for all  $t \in [t_1, t_2]$ .

Let  $\mathcal{L}(S)$  denote the Lebesgue measure of a set  $S \subset \mathbb{R}$  and let  $\mathcal{F} = \{\mathcal{I}_k : 1 \leq k \leq K\}$  be a large family of intervals, which in particular is so rich that the following holds: the intervals in  $\mathcal{F}$  with minimal length  $h_{\min}$  cover the unit interval  $[0, 1]$ , that is,

$$\bigcup_{\mathcal{I} \in \mathcal{F}_0} \mathcal{I} = [0, 1] \quad \text{with} \quad \mathcal{F}_0 = \{\mathcal{I} \in \mathcal{F} : \mathcal{L}(\mathcal{I}) = h_{\min}\}.$$

Under the technical conditions of the paper, we can prove the following:

- (i) With (asymptotic) probability  $\geq 1 - \alpha$ , the test does not reject the null for any interval with  $\mathcal{I}_k \cap [t_1, t_2] = \emptyset$ .
- (ii) With probability tending to 1, the test rejects the null for any interval  $\mathcal{I}_k$  for which  $\mathcal{L}(\mathcal{I}_k \cap [t_1, t_2]) \geq h_{\min}/2$ .

Now let  $\mathcal{F}_{\text{reject}}^{\min}(i, j)$  be the set of minimal intervals and let  $\hat{\mathcal{I}}$  be their union, that is,  $\hat{\mathcal{I}} = \bigcup_{\mathcal{I} \in \mathcal{F}_{\text{reject}}^{\min}(i, j)} \mathcal{I}$ . Then

$$\mathbb{P}([t_1, t_2] \subseteq \hat{\mathcal{I}}) \geq 1 - \alpha + o(1)$$

and

$$\mathbb{P}(\mathcal{L}([t_1, t_2] \setminus \hat{\mathcal{I}}) \leq h_{\min}) \geq 1 - \alpha + o(1)$$

Hence, with (asymptotic) probability  $\geq 1 - \alpha$ ,  $\mathcal{I}$  covers  $[t_1, t_2]$  and it is not much larger than  $[t_1, t_2]$  in the sense that the symmetric difference  $[t_1, t_2] \setminus \hat{\mathcal{I}} \cup (\hat{\mathcal{I}} \cap [t_1, t_2])$  is smaller than the smallest interval length  $h_{\min}$ .

The set  $\mathcal{I}$  is of course not a consistent estimator of  $[t_1, t_2]$ . It can rather be interpreted as some sort of confidence set: We know with probability  $\geq 1 - \alpha$  that  $[t_1, t_2]$  is contained in  $\mathcal{I}$  and that the set  $\mathcal{I}$  is not much larger than  $[t_1, t_2]$ .

(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.*

It can indeed be proven that our multiscale test has asymptotic power 1 against local alternatives. We have added this result as Corollary A.2 in the Appendix to 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.*
- (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.*

We have deferred the introduction of this test statistics to the Appendix.

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

We have replaced the "cp." abbreviation by "e.g." and in some cases by "see" throughout the paper.