Possible questions for the JMP

1. Examples of quasi-Poisson models (other possible applications).

   Since this is a count data, it has a lot of applications in ecology such as species richness or parasitism. There are also studies that are concerned with traffic safety evaluation (accident frequency, Accident Hazard Index) that use quasi-Poisson models. Shopping behaviour can also be accommodated in this model.

   Moreover, since our proof strategy does not rely on the fact that we observe count data, we are also able to work with continuous data. For example, simple volatility models where we are looking at the square of the returns.

2. How many hypotheses overall?

   For $T = 150$, $n = 5$ and the family of the intervals that we consider in the paper, we test 1560 hypotheses simultaneously.

3. Is the overdispersion parameter the same for all countries? Why? Is it possible to accommodate country-specific overdispersion parameter?

   In our model, overdispersion is defined as the structural parameter that is characterised by the nature of the pandemic. Hence, it is the same for all of the countries. Currently, we are working on expanding the model further by introducing the scaling parameter, so that we can accommodate differences in the scale of the epidemic. This scaling parameter would be country-specific, hence, the differences in the overdispersion may be

4. What is the intuition behind the bound on the minimal length $h_{\min}$?

   The idea of our testing procedure is that we approximate the weighted averages of the observed data points by the weighted averages of the standard normal random variables on intervals of different lengths. If we allow the minimal length to go to zero faster than 1/T, we would have less and less data points in our interval, hence, the approximation wouldn't work properly.

5. How to choose a_k and b_k? Are we free to choose whatever we want? Are there any bounds?

   It is possible to choose a_k and b_k in many different ways. For example, $a_k = 1$ and $b_k = 0$ give us the traditional critical values that were just discussed before. However, our choice of these constants allows us to balance the significance of the hypotheses between the intervals of different lengths.

6. Do you allow the number of countries to grow?

   Yes, we allow that. The only restriction is that the overall number of o hypotheses should not grow faster than some polynomial of the sample size.

7. What happens if we have zero cases?

   In our application we do have days with zero cases. One such day does not matter to us. We approximate the weighted averages on the intervals of 7 days and more, so we need the data not to have seven days of zero cases in a row.

8. Why do you need assumption that the mean functions are bounded away from zero?

   Since the variance in our model is proportional to the mean, we can't have the mean equal to zero. Otherwise the variance would be zero as well.

9. What are the examples of a distribution from quasi-Poisson family?

   Binomial distribution has variance proportional to the mean. Negative binomial has the variance that is a quadratic function of the mean. This is actually a mixture model of Poisson-gamma distribution: in this case the parameter of a Poisson distribution can itself be thought of as a random variable drawn from the gamma distribution. Two-parameter type A Neyman distribution which is in fact a type of Poisson-Poisson mixture.

10. Why do you have independence across countries in the error term?

    In the application, we consider roughly the first five months of the pandemic, and during this time most countries introduced severe travel restrictions and even full closure of the borders. So this assumption seems reasonable in the current situation.

11. Why do you have independence across time in the error term?

    For each country, corresponding time series process is non-stationary. Specifically, both the mean and the variance are time-varying. A well-known fact is that such non-stationarities may produce spurious sample autocorrelations. Hence, the observed persistence of a time series captured by the sample autocorrelation function may be due to non-stationarities rather than real autocorrelations. So we opted for a simple non-stationary model over an intricate stationary time series model.

12. You have conservative bounds for FWER. Why? Are there any special cases such that the bounds are not conservative?

    In our case, the bounds are conservative because we can approximate the test statistic by the Gaussian version only under the null. Since the set of intervals where the null hypothesis holds is not known in advance, but we use the critical values from the approximation calculated over the whole set of intervals, the bound is conservative. In the case where the null hypothesis holds everywhere, FWER will be asymptotically tending to the significance level alpha. In the other case, where there is at least one region where two functions differ, there are some iterating procedures that allow us to first determine the set $\mathcal{M}_0$ and then calculate the critical values, but it gets much more complicated. We were aiming for a simple procedure that controls FWER and has asymptotic power of 1 against local alternatives, and we have it.

13. How do you account for seasonality?

    First, we look at the intervals with the lengths that are multiples of 7 days: one week, two weeks, three weeks. This allows us to take into account the reporting procedures, when during the weekend the number of cases may be smaller than over the week.

Second, we align the data taking the first Monday after reaching 100 cases as the starting date of the time series. This way we can have comparable patterns of the reporting in different countries. We also make robustness checks in the paper that perform the test without such alignment.

We do not introduce seasonality specifically to the model but we also do not exclude the possibility of it.

14. Does the size of the country matter? Why do not you divide by the population?

    Obviously, the size of the country matters. However, it is not straightforward to normalise the data properly. Simply dividing by the population of the given country would not be sufficient in our case, because the outbreak of the virus in a specific country is a local event (Wuhan in China, Lombardy in Italy). Clearly, this locality does not play any role in the later stages of the epidemic because additional local outbreaks occur over time and their number will presumably be larger in countries with a larger population. So, overall, the relationship between the population size and the epidemic curve can be very complicated. As for our paper, in the application sections we consider the European countries that are fairly similar (in population size but also in other characteristics), hence, we opted for not normalising the data by population size at all rather than "wrongly normalising".

15. What about different country-specific testing strategies?

    It is very probable that our test results are not driven by different test regimes in the countries under consideration. Take Germany vs. Italy, for example. Germany is often cited as the country that employed early, widespread testing with more than 100 000 tests per week even in the beginning of the pandemic, while testing in Italy became widespread only in the late stages of the pandemic. Nevertheless, even visual inspection of the data suggests that the underlying time trends are very similar in the beginning, and this is confirmed by our multiscale test.

16. Can you incorporate change points? Under the null, the approximation of the test statistics by the Gaussian version of it is still valid. If we are not under the null, but we know the dates where these change points occur, then most probably the critical values are still valid. However, we will need to check it.

Unanswered questions:

- Why don't you normalise by the number of tests performed?

- Why do you look at the number of cases and number of deaths or case-fatality rates?

- What about policy implications?

- What steps can be taken further?

  "A promising direction is to connect the models with data-driven techniques, particularly machine learning. The work by Yang et al. (1) applied a machine learning approach based on a recurrent neural network that is trained by utilising a 2003 SARS epidemic dataset as well as incorporating the COVID-19 epidemiological parameters. They found consistent patterns in the predictions from the SEIR model and from the machine learning. These results are encouraging for wider applications of data analysis and computing approaches to study epidemics and pandemics, particularly COVID-19. Machine learning and other artificial intelligence techniques can complement and improve mathematical epidemic models by taking advantage of the large data sets currently available, including epidemic, genetic, demographic, geospatial and mobility data, the scale of which is typically far beyond the applicability of a standard mathematical model. On the other hand, mathematical modelling can provide a meaningful way to validate machine learning predictions and to guide the development of more efficient and robust algorithms in machine learning and data analytics. Thus, the development and advancement of these two different quantitative approaches could be mutually beneficial, and their integration could lead to potentially transformative progress in the study of COVID-19 and beyond."

- Why rescaled time? What does it mean?

- What is the intuition behind the bound on the maximal length $h_{\max}$?

- Is the smoothness of the data a reasonable assumption? What if some government policies affects the epidemic such that there is a sudden change in the trend function?

- The simulation study is not really consistent with the asymptotic theory. Why?