Presentation

Introduction

0 - 8 min

Thank you very much for having me here. Today I will be speaking about multiscale inference for nonparametric time trends. This is a joint project together with Michael Vogt, who is a professor at the Institute of Statistics at the University of Ulm. This is very much work in progress, so all suggestions, comments and questions are definitely welcome. You can either ask me during the talk or after it, I will try leave around ten minutes for the questions.

The structure of the talk is as follows. First, I will give a brief overview about the research question and our method related to the literature. Then I will introduce the model. Since in the project we propose a new testing method, I will then describe the testing procedure itself. After that, I will outline the main theoretical properties of the method. And in the end, I will show you a couple of pictures in order to illustrate the possible application of our method.

The overall goal of this project is to develop new inference methods that will allow us to identify and locate the differences between nonparametric trend curves with dependent errors.

In order to better illustrate the method and its applications, let's first look at a simple example. Here is the data on the exchange rates daily number of new cases of infection in Germany and in Italy for the first five months of the pandemic. Time series for Germany is depicted in orange, and time series for Italy is depicted in black. Since these types of comparison were very common in the media for the last year, I think most of you saw (or even asked yourself) the following questions at least once: how do the outbreak patterns of COVID-19 compare across different countries? Are there significant differences? And if there are significant differences between the underlying trends, are we able to pinpoint them?

Based on the visual inspection, we can say that both time series are quite similar but we can still find the differences. Let's look closely at this time region. It roughly corresponds to the end of the first month of the pandemic. As you see, time series in Germany exhibits much more volatility than in Italy, specifically, there is a local peak with almost twice as many cases as in Italy on that day. So the question we can ask is whether this peak comes from the differences in the trend or is it just an artefact of the sampling noise? Or consider this time period, roughly corresponding to days 45 to 90 of the pandemic. We can see that during the whole time the Italian data lies almost always above Germany, but can we state this with any kind confidence?

And these are only two intervals, what happens if we look at many intervals at the same

time and multiple pairs of countries? With the help of the method that we propose, even in this complicated setting we are able to make simultaneous confidence statements about the time regions where the trends differ.

The research question that we are aiming to answer in this project is a very relevant question in our current situation. Identifying the systematic differences in the development of the pandemic between countries provides the basis for further research, which goes presumably beyond pure statistics. For instance, it may help to better understand which government policies have been more effective in containing the virus than others.

However, this is not a trivial task. The method that we propose is a multiscale test. The underlying idea of a multiscale test is to consider simultaneously a number of test statistics that are usually not independent. Each of these statistics corresponds to different values of parameters (in our case, to the different time regions of interest) which leads to the well-known multiple testing problem. If the method does not take this issue into account, then the probability that some of the true null hypotheses are rejected by chance alone may be very large.

And the main contribution of the paper is the method with the applications that are not limited to the current COVID-19 crisis. I would like to stress out that it is a general method to compare nonparametric trends. Thus, we advance the research on statistical tests for equality of nonparametric regression and trend curves.

This project contributes to at least three different strands of the literature. First, even though our multiscale test is motivated by the current COVID-19 crisis, I have already said that this is a quite general method. Even though tests for equality of the trends have been developed already for quite a while, most existing procedures allow only to test whether the trend curves are all the same or not, but they almost never allow to infer which curves are different and where. To the best of my knowledge, the only other test for comparing trend curves with similar properties has been developed in Park, Hannig and Kang in 2009. However, their analysis is mainly methodological and is not really backed up by theory whereas we propose an easy-to-use testing procedure with proven theoretical properties.

The second strand of the literature that this project is related to is papers on multiscale testing. Multsicale procedures are developed for different purposes, and the only one that has the similar goal is the one from Park, Hannig and Kang, 2009, which I have just mentioned.

And of course, there is an exponentially growing body of literature on COVID-19. In terms of the research question, our paper is closest in spirit to Dong, Gao, Linton and Peng (2020), who study country-specific trending behaviour of the COVID-19 data while taking into account very pronounced deterministic trends. Specifically, they compare of the overall efforts of different countries to "flatten the curve". We, in contrast, while

also taking into account very pronounced deterministic trends, are able to pinpointing the regions where this "flattening of the curve" was more effective in one country relative to the others.

Model

8 - min

Testing procedure

- min

Theoretical properties

- min

Application

- min