CAGE

Update on estimating and forecasting COVID-19 evolution in the UK

CAGE Policy Briefing no. 23

April 2020

Roberto Pancrazi, University of Warwick and CAGE

WARWICK
THE UNIVERSITY OF WARWICK

UKRI Economic and Social Research Council

Update on estimating and forecasting Covid-19 evolution in the UK

Roberto Pancrazi[1]

University of Warwick

April 26, 2020

Introduction

In the previous briefing,[2] I estimated the Richards' phenological model to forecast the Covid-19 evolution in the UK and in England's NHS regions using data up to April 8th, 2020. Since then, there have been three important sets of information that should be taken into account to predict the trajectory of the epidemic: first, new data are available; second, the number of tests have been increasing; and third, there was quite a significant revision on how the data on both cases and tests were reported.[3][2]

In this briefing I take into account this additional information to further investigate the expected trajectory of the epidemic. There are three main results. First, the model confirms that the peak of new infections was reached on April 8th; second, while robustly in the downward trajectory of the infection curve, currently, on April 24th, the UK is only at roughly one third of the way through the decline. Finally, the end of the epidemic, defined as the date with zero new daily cases, is predicted for August 8th, conditional on the contagion not growing if/when social restrictions are relaxed.

In the first part of the notes, I validate these predictions investigating the ability of the model to correctly forecast the evolution of the epidemic, by comparing

[1] U. Warwick: r.pancrazi@warwick.ac.uk

[2]see Pancrazi (2020)

[3] See https://coronavirus.data.gov.uk/about

the predictions made on April 8th with the observed data up to April 24th. I show that the model I have proposed has had satisfactory out-of-sample predictions. There are two points to make. First, from a more technical point of view, I highlight that including the correction for the variability of the number of daily tests is a crucial component for the success of the model and that, without weighting the observed number of daily cases with the number of tests taken, the raw data have very little information contents. Second, the fact that the recent observations line up well with the model predictions is a signal that the policy restrictions adopted at the end of March are working, as the epidemic seems to follow its natural course, which is what the model is able to forecast.

Because the UK is only on the first part of the downward trajectory, when deciding for a relaxation of the social restrictions, a policy maker should also take into account a possible strengthening of the contagion.

In summary, after showing that the out-of-sample performance of the Richard's model with the correction for the variation in the number of daily tests are quite satisfactory, I re-estimate the model using the most recent observations to provide relevant information to policymakers about the predicted evolution of the epidemic.

1. The out-of-sample model performance

In this section I re-estimate the Richards' phenological model using the sample considered in the previous briefing, which covers the period up to April 8th, 2020. The main goal is to provide evidence of the good forecasting properties of the model, which will then be used to conduct additional predictions about the evolution of the Covid-19 epidemic in the UK.

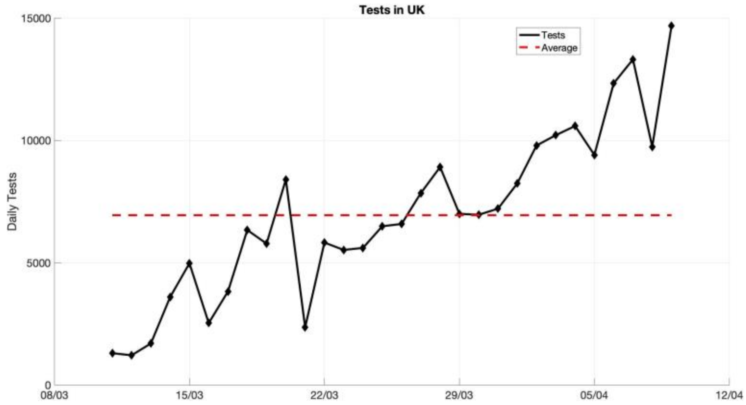
The preferred specification of model is the one that corrects for the variability of daily tests;[4] this is an important factor to consider because the number of daily tests (people tested) have substantially increased from the beginning of March, as displayed in Figure 4. Therefore, an increase in the number of daily cases, as reported by Public Health England (PHE), might be due to more intensive testing, rather than to an increase in spreading of the disease. Hence, I employ the correction I propose in Pancrazi (2020), assuming a daily number of tests equal to its sample mean, 9114. In section 4 I will discuss how the absence of a steady number of daily tests drastically

[4] See Pancrazi (2020) for the details of the model and for the details of the proposed correction.

distorts the ability of the data to give precise information on the evolution of the disease.

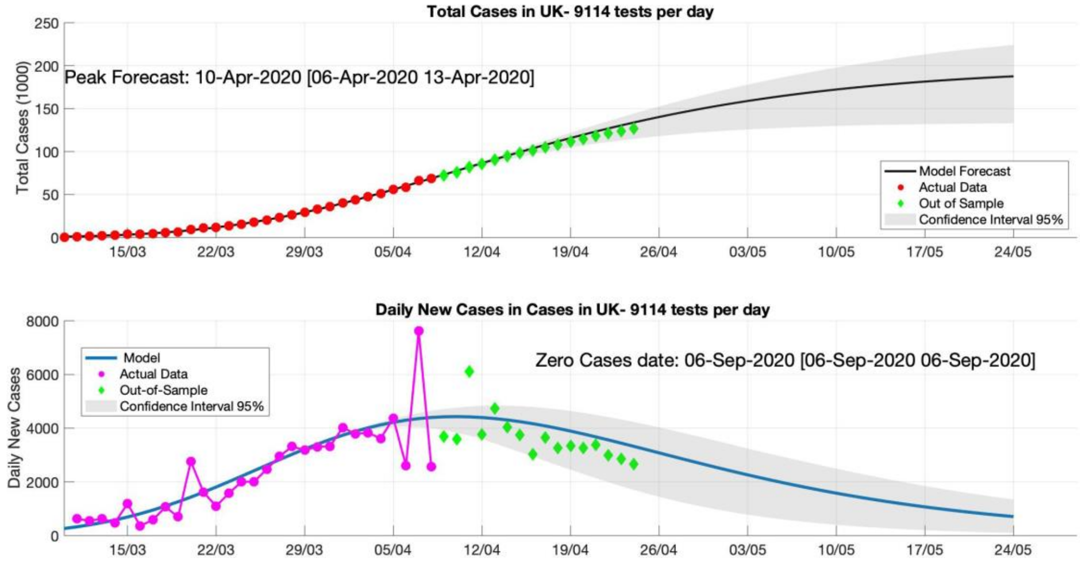
Figure 1: Controlling for number of tests

(a) Tests in UK (people tested)



In Figure 2 I display the following information. In the top panel, the red dots are the official number of total cases, corrected by the variance of daily tests, used to estimate the Richards' model; they cover the sample period 10th March - 8th April. This is the same sample considered in the previous note, Pancrazi (2020). The model fit and predictions are displayed with the black solid line and the shaded grey area represents the 95 percent confidence bands. Finally, the green diamonds display the actual observed daily data that were not included in the estimation process (from April 9th to April 24th). In the bottom panel, the purple dots display the official number of daily cases, corrected by the variance of daily tests, used in the estimation, while the model fit and predictions are displayed with the blue solid line; notice that this model line is simply the rate of change of the estimated curve in the top panel. Again, the green diamonds are the out-of-sample observations.

Figure 2: Out-of-sample model performance



There are several remarks worth making. First, the actual data are slightly different than those used in the previous version of the notes. This is because of a change in the way the PHE has decided to report the data, specifically in the number of tests, which resulted in a revision. In fact, when the original notes were written, the tests reported on 7th April and 8th April, were, respectively 9740 and 14682, while after the revision they are now reported as equal to 4344 and 9527, respectively. This revision moves quite drastically the last couple of datapoints, and it has strongly affected the forecasted trajectory. Notice that the date of the peak estimated using the original sample was April 10th, while in the original notes was April 3th. Again, this discrepancy is due only to the data revision.

Second, the forecasted evolution of the epidemic tracks quite well the actual observed datapoint not used in the estimation. Beside the first three observations, all the observed data from April 12th to April 24th lie within the 95 percent confidence bands, and they have a trajectory that is in line with the one predicted.

Third, if anything, the actual decline in the number of daily cases after having reached the peak was slightly faster than the one predicted, as the last eleven observations lie below the forecast line.

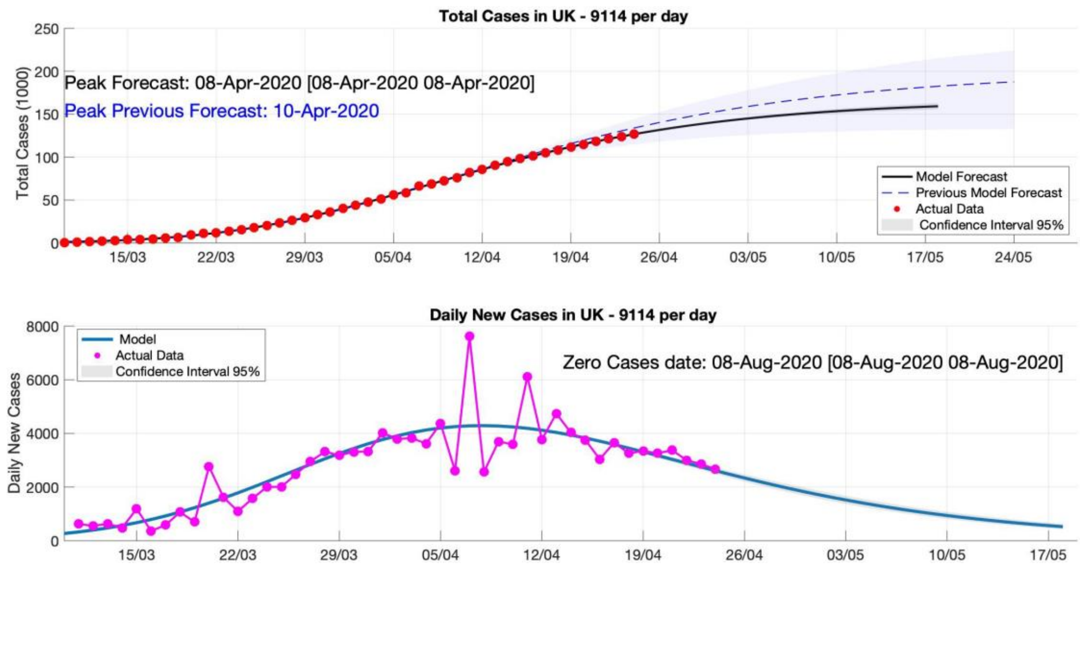
The fact that the recent observations line up well with the model predictions is a signal that the policy restrictions adopted at the end of March are working, as the epidemic seems to follow its natural course, which is what the model is able to forecast. In summary, the out of sample performance of the Richard's model with the correction for the variation in the number of daily tests are quite satisfactory. Therefore, re-estimating the model using the most recent observations, as I present in the next section, might give relevant information for policymakers on the predicted evolution of the epidemic.

2. Updated Forecast

In this section I present the forecast for the evolution of the epidemic when using all the available observations- that is up to April 24th.

In the top panel, the red dots are the official number of total cases, corrected by the variance of daily tests, used to estimate the Richards' model; they cover the sample 10th March - 24th April. The model fit and predictions are displayed with the black solid line and the shaded grey area represents the 95 percent confidence bands. The dashed blue line and the light blue shaded area display the forecast estimated using data only up to April 8th and its confidence band: they are the same as the one presented in Figure 2. In the bottom panel, the purple dots display the official number of daily cases, corrected by the variance of daily tests, used in the estimation, while the model fit and predictions are displayed with the solid blue line.

Figure 3: Updated forecast



There are several important points to highlight. First, the new available data revise downward the forecast with respect to the one made on April 8th; however, the new forecast decisively belongs to the original confidence interval. As I have already mentioned, this fact might be interpreted as a signal that the social restrictions imposed by the government are indeed working.

Second, as a consequence of the downward revision, the estimated date of the peak moves backward two days, from April 10th to April 8th. Hence, we can quite confidently assess that the peak of the daily infections has passed.

Third, while the uncertainty, measured by the width of the confidence band, was quite large in the original estimation, with the new available data the confidence band is extremely tight.

Forth, the data observed in the last week are strongly in line with the path predicted by the model. Notice that, while robustly in the downward trajectory of the infection curve, currently, on April 24th, the UK is only at roughly one third of the overall decline.

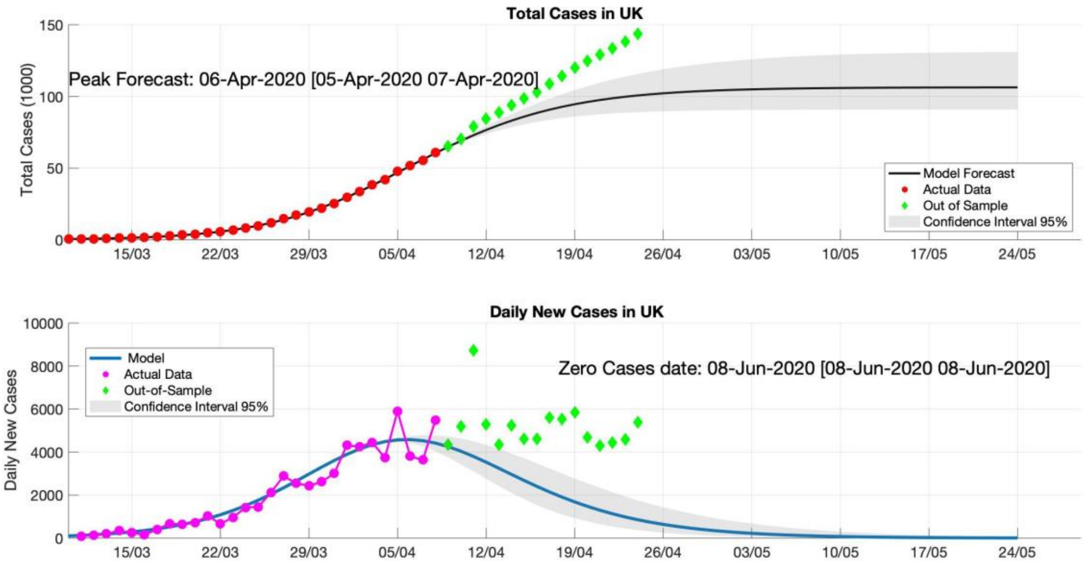
Finally, the end of the epidemic, defined as the date with zero new daily cases, is predicted for August 8th, conditional on the contagion not growing if/when social restrictions are relaxed.

3. The importance of the correction for the volatility of daily tests

As shown in Figure 1, the number of tests almost doubled in the last two weeks. With such a large increase in the number of people tested it should not come as a surprise that more infected people are detected. However, this information is not taken into account in standard epidemiologic models, which assign increasing observed number of cases to a rise in the disease spread rather than to a more expansive testing strategy. Hence, it would be misleading to conduct inference and compute forecasts using raw data that is without the correction for the number of daily tests.

In order to visualize this point, in Figure 4 I display the out-of-sample performance of the model without correction, which, therefore, use the raw data. This figure is, then, the analogous of Figure 2. We can see that the observed data are far from the predicted decline of the epidemic curve, as the number of cases was roughly constant around 4500 in the last 20 days. However, recall that in the same time span the number of people tested almost doubled, which implies that the actual number of infected per tests declined, as predicted by the model.

Figure 4: Out-of-sample model performance



Because, as I just showed, an analysis of the raw data is not informative and because, unfortunately, the number of tests at the level of NHS regions is not available, in this note I do not forecast the epidemic path at a more disaggregated level.

References

Pancrazi, R. (2020): “Estimating and forecasting COVID-19 evolution in the UK and its regions,” CAGE Policy Briefing no. 20.