CAGE

Estimating and forecasting COVID-19 evolution in the UK and its regions

CAGE Policy Briefing no. 20

April 2020

Roberto Pancrazi, University of Warwick

WARWICK
THE UNIVERSITY OF WARWICK

UKRI Economic and Social Research Council

Estimating and forecasting Covid-19 evolution in the UK and its regions: A brief note

Roberto Pancrazi*
University of Warwick

April 8, 2020

1 Introduction

Understanding and predicting the evolution of outbreaks is crucial for policymakers, as they face the trade-off between limiting social interaction and avoiding a further spread of the disease. Currently the UK is in lockdown and it will likely last until the severity of infections is sufficiently reduced. The length of the restrictions in place clearly has implications on total government budget, businesses' ability to stay afloat, consumer sentiment, etc. In addition, social restrictions tend to make the evolution of the epidemic location specific; therefore, some areas might be able to come out of lockdown before others. Hence, a key question for policymakers is when and where the epidemic will sufficiently slow down.

Using official data on new cases for the four nations and England's NHS regions, I estimate a Richards' phenological model to forecast their Covid-19 evolution. As the evolution of official cases is severely affected by the variability of daily testing, I also propose a simple strategy to control for this source of noise. Chowell (2017), Chowell, Tariq, and Hyman (2019), Hsieh, Fisman, and Wu (2010) Hsieh (2009), Hsieh and Chen (2009), Roosa, Lee, Luo, Kirpich, Rothenberg, Hyman, Yan, and Chowell (2020), Viboud, Simonsen, and Chowell (2016), and Wu, Darcet, Wang, and Sornette (2020), among others, have investigated and applied the properties of Richards' model for forecasting purposes to an early stage and real-time evaluation of an epidemic outbreak, such as for H1N1 in Canada in 2009, SARS in the Great Toronto Area in 2003, Dengue in Singapore in 2005, Ebola in the Democratic Republic of Congo, and for the recent Covid-19 in China.

Let us summarise the main results. For the UK, overall it is likely that we are approaching the peak of the epidemic, which the model pinpoints during the current week. This timeline holds even when accounting for the large variability of daily tests. It will then take around one to three weeks

*U. Warwick: r.pancrazi@warwick.ac.uk

before the epidemic shows half-way on the decaying path. The more positive scenario places the day we achieve zero new cases is around May 10th, while the more negative one places it at the end of May. These predictions are conditional on strict social distancing. As these are likely to be imperfect, the day of the peak might be delayed by a few days. In the case of Italy, the final estimated peak was 5 days later than the one originally estimated.

The dynamics of the UK path are driven by England and Wales, the regions of which share a very similar predicted timeline. Scotland and Northern Ireland seem to be on a slightly faster track towards the end of the epidemic. The analysis for NHS regions confirms that this week is likely to see the peak of new cases for all the regions.

2 Forecast

2.1 Method

Let us denote by $C(t)$ the total number of cases in period t. Richards' model implies it follows:

$$C(t) = \frac{K}{\left[1 + S \exp^{-ar(t-t_c)}\right]^{\frac{1}{a}}}. \quad (1)$$

Here *r is* the intrinsic growth rate, *K* is the carrying capacity, *S* is a shifter, and *a* is the exponent of deviation. t_c denotes the turning point (henceforth *peak)* defined as the time when the second derivative of $C(t)$ vanishes, and, consequently, the cases per unit of time reach its maximum level. The basic intuition behind the model is that the rate of growth of new cases is an inverse u-shaped function of the evolution of the total number of cases, and, therefore, of time.

2.2 Data

I consider the reported cases of coronavirus per upper tier local authority (UTLA) in England, up to April 8th, 2020. In this note I will show the prediction for the United Kingdom, the four countries (Scotland, Wales, Northern Ireland, and England) and for the six NHS regions. [1] Daily cumulative case counts are those that were published each day on the Public Health England (PHE) Dashboard and based on cases reported to PHE by diagnostic laboratories. People who have recovered and those who have died are included in the cumulative counts, as they are people who have been infected.

I then use the data to estimate the parameters *{K, a, r, S}* of the equation (1).

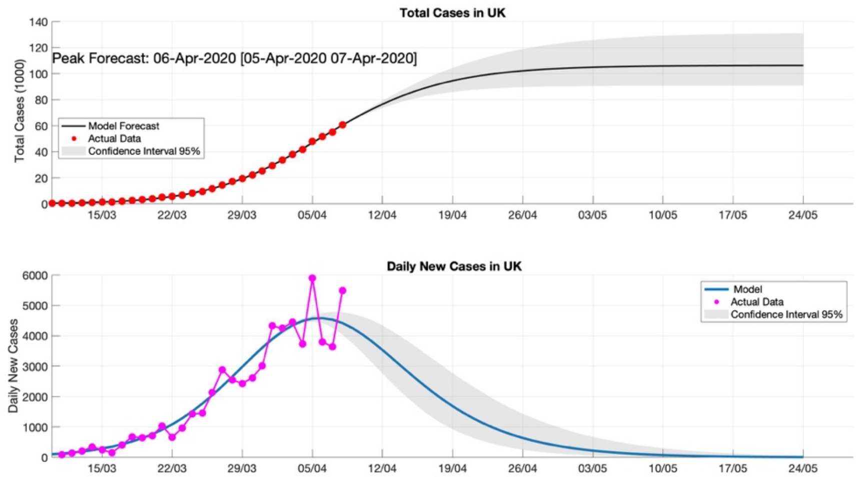
[1] London, Midlands, North West, North East and Yorkshire, South East, East of England, South West

As a result, I estimate the date of the peak. I then show how the actual data compares to the model prediction and I display the forecasted path of the epidemic. I compute 95 percent confidence bands using bootstrap.

2.3 Forecast for the UK

Let's start with the forecast for the United Kingdom, considering Figure 1. In the top panel, the red dots are the official number of total cases used to estimate the Richards' model. The model fit and predictions are displayed with the black solid line. Although the peak is estimated to be on April 6th, there is still quite a bit of uncertainty, as measured by the shaded grey area, which represents the 95 percent confidence bands. The estimation places the peak of the epidemic between April 5th and April 7th. The more positive scenario places the day with zero new cases around May 10th, while the more negative scenario places it at the end of May. Most importantly, recall that the forecast is conditional on the restrictive measures working well: if that is not the case, with new data available we should observe the forecast lines, as well as the date of the peak, move up.

Figure 1: UK: Estimation and Prediction of the Richards' model



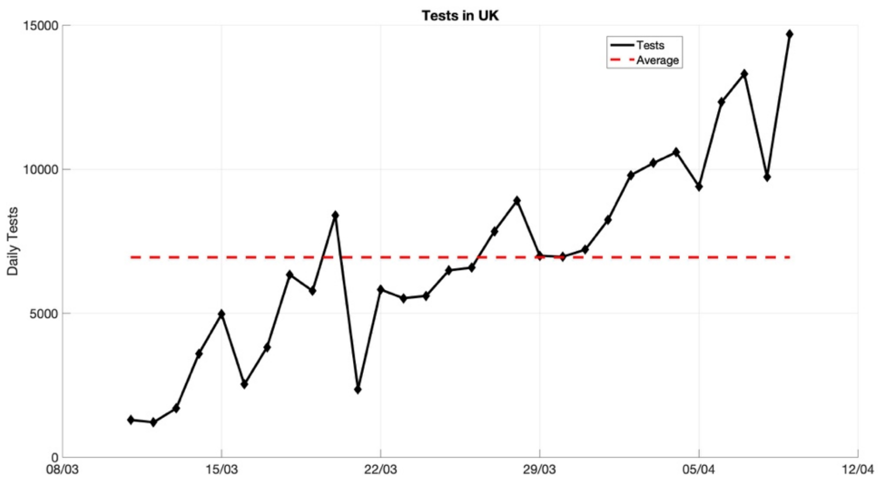
Note: In the top panel, the red dots are the official number of total cases used to estimate the Richards' model. The model fit and predictions are displayed with the black solid line. In the bottom panel, the purple dots display the official number of daily cases. The model fit and predictions are displayed with the blue solid line. The shaded grey area represents the 95 percent confidence bands.

In the bottom panel, the purple dots display the official number of daily cases. 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. The daily new cases are quite noisy. This is a common feature of the data across countries, probably due to measurement disturbances, due, for example, to the highly volatile number of tests conducted on different days. The estimated model attempts to minimize the overall distance between its predicted line and the noisy observed data. Hence, it is important to evaluate the model not so much on its ability to predict daily new cases, day after day, but rather on its ability to capture its underlying trend.

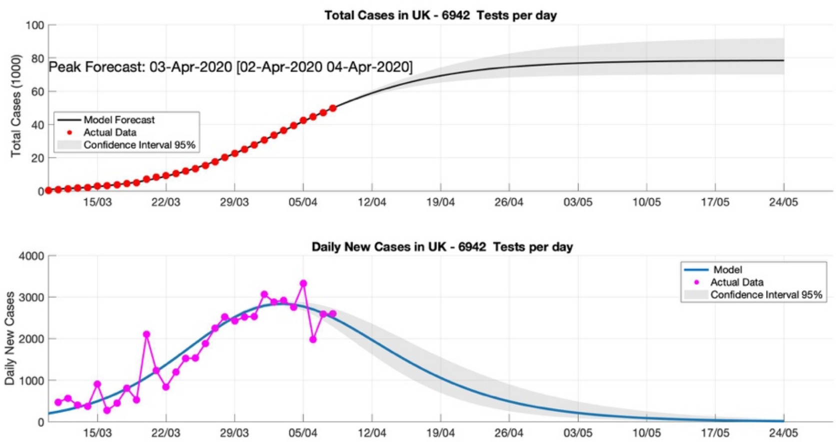
The number of official cases is affected by the number of tests conducted, which varies quite a lot daily. The top panel of Figure 2 displays how it has evolved. To control for the variability of testing, I have constructed a series of cases assuming that 6942 individuals are tested every day; that number is the average of the tests so far.[2] The top panel of Figure 2 displays the forecasts, controlling for the variation in tests. Notice that uncertainty is much lower, and the model places the peak as happening in the past weekend.

Figure 2: Controlling for number of tests

(a) Tests in UK



(b) Forecast fixing tests



3 Model Validity

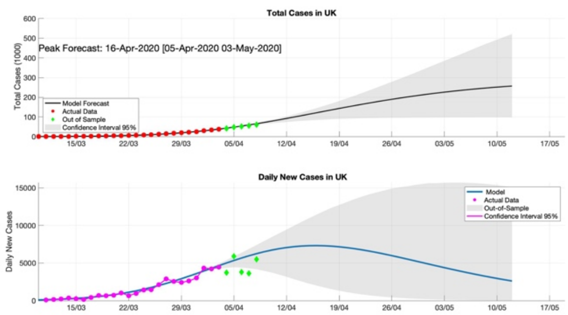
To show evidence of the short run forecast performance of the model, in Figure 3 I have added the prediction of the model using data up to April 3rd (left panel)

[2] The assumed daily number of tests is just a normalisation and does not affect the dynamics of the forecasts, just its level.

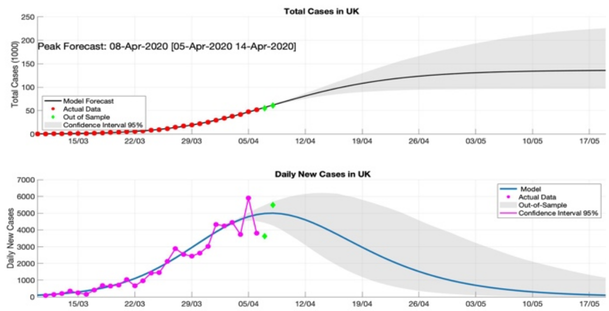
and up to April 6th (right panel). The green dots display the out-of-sample data. The top panel displays the model estimated with the official cases; the bottom panel displays the model estimated with the artificial series that controls for the number of tests. There are two important remarks. First, the overall trend of the epidemic is well forecasted by the model, particularly for the model that controls for the variability in the number of tests. Second, as I mentioned earlier, this class of model gets most of the information when daily cases start rising so it is not surprising that, given the current state of the epidemic in the UK, the model updates sensibly its estimates and forecast with every single day of new data. Notice, in fact, the lower amount of uncertainty in the forecasts when more data points (purple dots) are added.

Figure 3: Out of Sample Forecast

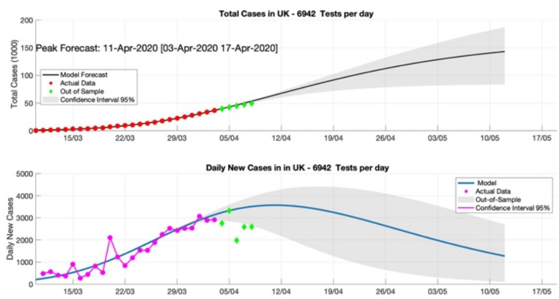
(a) Up to April 2nd



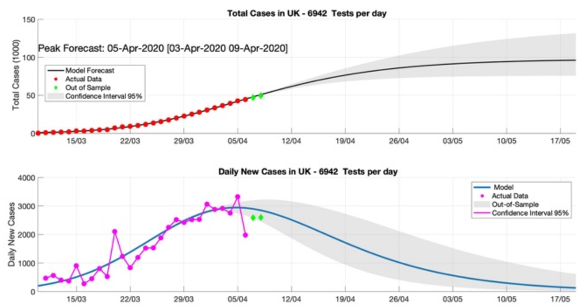
(b) Up to April 4th



(c) Up to April 2nd-Constant Tests



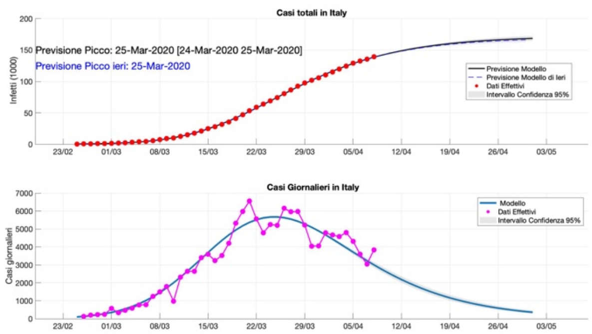
(d) Up to April 2th-Constant Tests



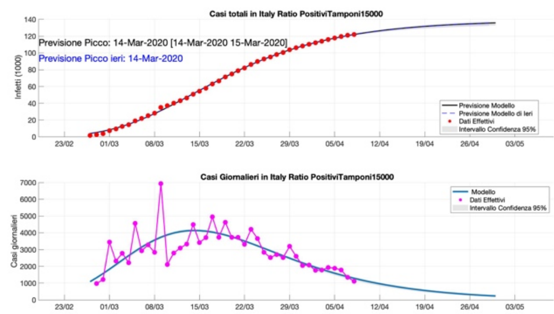
Finally, it is quite important to show how the model behaves when the epidemic is at a more advanced stage. To illustrate this point, I have conducted the same exercise for Italy, which allegedly is a couple of weeks in front of the UK epidemic. I have estimated the model using official cases (Figure 4a), using the artificial series that controls for tests (Figure 4b) and for the two most affected regions, i.e. Lombardia, and Emilia- Romagna. The results are displayed in Figure 4. The model has learnt substantially from the evolution of the cases; hence, the uncertainty bands disappear and we can be quite confident that the policy restrictions imposed on March 8th, March 15th, and March 22nd had an impact and the epidemic is now on its decaying path. This is crucial information for policymakers.

Figure 4: Estimated model for Italy and three regions

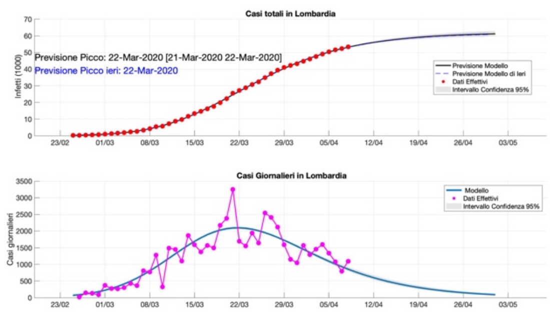
(a) Italy



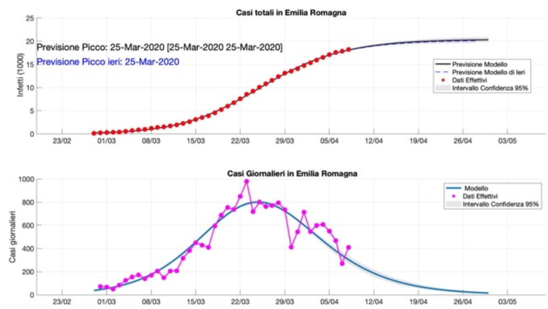
(b) Italy - Constant Tests



(c) Lombardia



(d) Emilia-Romagna



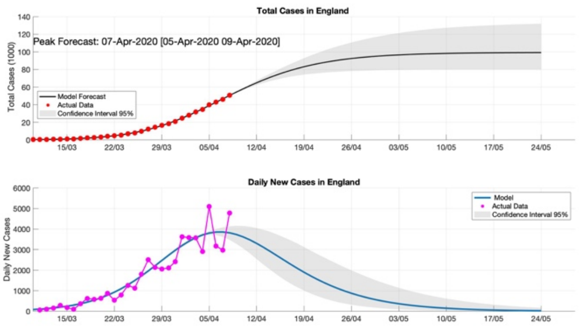
4 Countries and Regions

Here I report the estimates and forecasts for the four countries comprising the UK

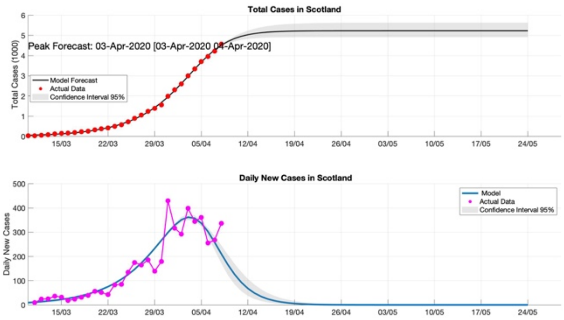
4.1 Countries

Figure 5: Estimated model

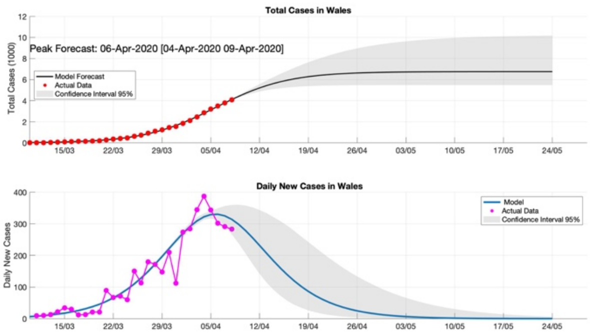
(a) England



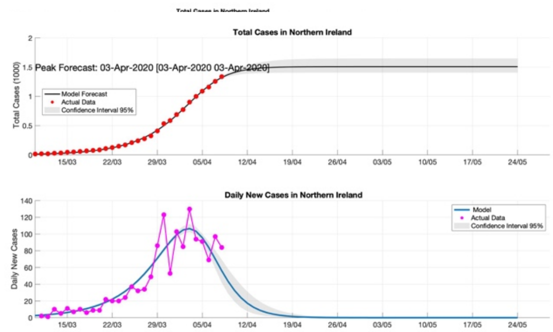
(b) Scotland



(c) Wales



(d) Northern Ireland

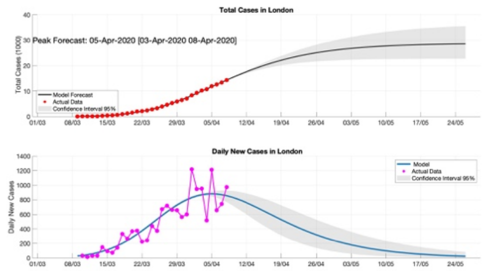


4.2 Regions

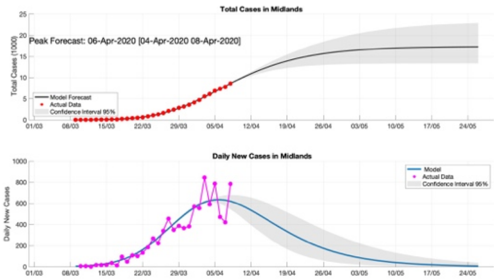
Here I report the estimates and forecasts for the NHS regions in the UK.

Figure 6: Estimated model

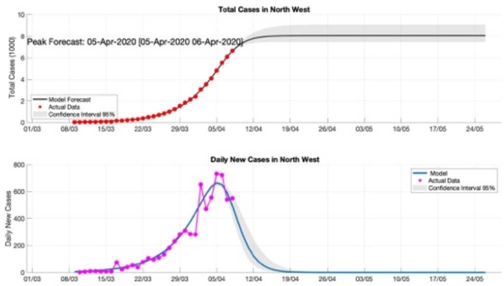
(a) London



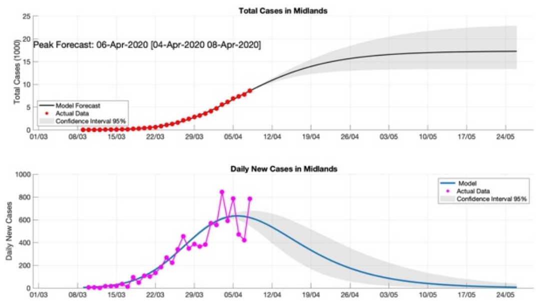
(b) Midlands



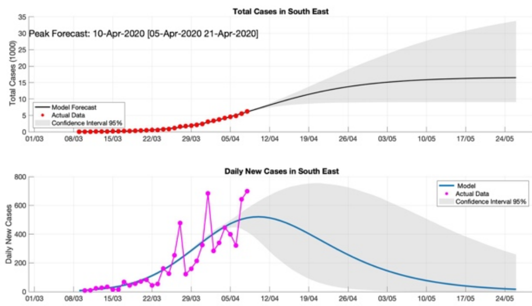
(c) North West



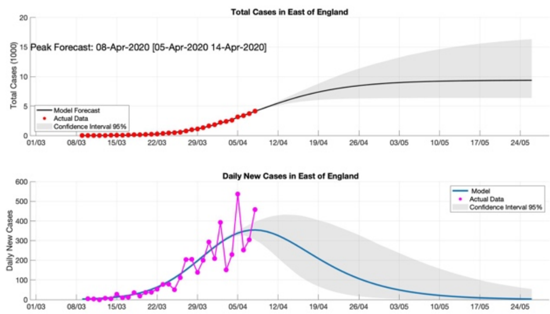
(d) North East and Y.



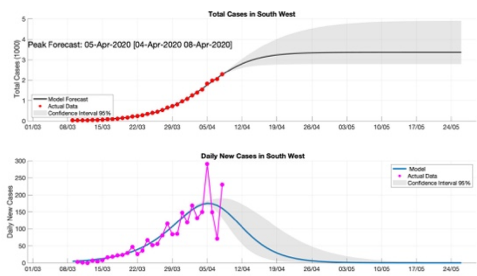
(e) South East



(f) East of England



(g) South West



5 Main Results and Discussion

5.1 Main Results

Let's summarise the main results. For the UK overall it is likely that we are approaching the peak of the epidemic, which the model pinpoints during the current week. This timeline holds even when accounting for the large variability of daily tests. It will take then around one to three weeks before the epidemic is half-way on the decaying path. The more positive scenario places the day with zero new cases around May 10th, while the more negative scenario places it at the end of May. Importantly, these predictions are conditional on strict social distancing. As these are likely to be imperfect, the day of the peak might be delayed by a few days. In the case of Italy, the final estimated peak was 5 days later than the one originally estimated.

The dynamics of the UK path are driven by England and Wales, the regions of which share a very similar predicted timeline. Scotland and Northern Ireland seem to be on a slightly faster track towards the end of the epidemic. The analysis for NHS regions confirms that this week is likely to see the peak of new cases for all the regions.

The analysis for NHS regions confirms that this week is likely to see the peak of new cases for all the regions.

5.2 Discussion

There are several advantages of estimating a Richards' phenomenological model for predicting the pattern of an outbreak. First, it is parsimonious, as only 4 parameters have to be estimated. Second, as shown by Wang, Wu, and Yang (2012) a Richards' phenomenological model has a one-to-one non-linear mapping with a S-I-R epidemiologic model. Third, it allows us to compute, sequentially, a comparison between what the model predicted earlier and the actual realized data; a departure of the realized observation with the model prediction is a measure of how the policies to limit contagion are or are not working.

There are, however, a few caveats to mention. First, the data used in this note for the estimation are total cases officially reported:[3] they are quite noisy and, obviously, they do not track perfectly the actual number of infected, as tested cases are only a small portion of the actual infected.[4] Nevertheless, if the testing policy is consistent over time, in the sense that the selection of tested among

[3] https://www.gov.uk/government/publications/covid-19-track-coronavirus-cases

[4] Flaxman, Mishra, Gandy et al. (2020) estimate that between 1.2 and 5.4 percent of the UK population was infected on March 28th, 2020; only 0.3 percent of the population was tested until April 6th, 2020

infected does not drastically change, the estimated and forecasted dynamic pattern for those tested positive should give relevant information for the unobserved pattern of actual infected. Second, the model gets most of the information from the data when the curve of total cases starts to get steeper; this means that while for regions in which the epidemic is already at high levels, such as the UK overall or the London area, the model is well informed and the forecast more accurate, for regions with low levels of cases, up to today, the model might be still be inaccurate. Third, human behaviour is not present in the model: this means that the forecast should be interpreted as conditional on the assumption that social contacts are very limited so that the epidemic follows its natural course. If that is not the case, one is likely to observe a daily readjustment upward of the forecasts.

References

Chowell, G. (2017): "Fitting dynamic models to epidemic outbreaks with quantified uncertainty: a primer for parameter uncertainty, identifiability, and fore- casts," *Infectious Disease Modelling*, 2, 379–398.

Chowell, G., A. Tariq, and J. M. Hyman (2019): "A novel sub-epidemic modeling framework for short-term forecasting epidemic waves," *BMC medicine*, 17, 164.

Flaxman, S., S. Mishra, A. Gandy, et al. (2020): "Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries," *Imperial College preprint*.

Hsieh, Y.-H. (2009): "Richards model: a simple procedure for real-time prediction of outbreak severity," in *Modeling and dynamics of infectious diseases*, World Scientific, 216–236.

Hsieh, Y.-H. and C. Chen (2009): "Turning points, reproduction number, and impact of climatological events for multi-wave dengue outbreaks," *Tropical Medicine & International Health*, 14, 628–638.

Hsieh, Y.-H., D. N. Fisman, and J. Wu (2010): "On epidemic modeling in real time: An application to the 2009 Novel A (H1N1) influenza outbreak in Canada," *BMC research notes*, 3, 283.

Roosa, K., Y. Lee, R. Luo, A. Kirpich, R. Rothenberg, J. M. Hyman, P. Yan, and G. Chowell (2020): "Short-term Forecasts of the COVID-19 Epidemic in Guangdong and Zhejiang, China: February 13–23, 2020," *Journal of Clinical Medicine*, 9, 596.

Viboud, C., L. Simonsen, and G. Chowell (2016): "A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks," *Epidemics*, 15, 27–37.

Wang, X.-S., J. Wu, and Y. Yang (2012): "Richards model revisited: Validation by and application to infection dynamics," *Journal of Theoretical Biology*, 313, 12–19.

Wu, K., D. Darcet, Q. Wang, and D. Sornette (2020): "Generalized logistic growth modeling of the COVID-19 outbreak in 29 provinces in China and in the rest of the world," *arXiv preprint arXiv:2003.05681*.