Short-term forecasting of the Coronavirus Pandemic - 2020-06-05

Jennifer L. Castle, Jurgen A. Doornik and David F. Hendry*
Department of Economics, Nuffield College, Magdalen College, and Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, UK

Abstract

We have been publishing real-time forecasts of confirmed cases and deaths for COVID-19 online at `www.doornik.com/COVID-19` from mid-March 2020. These forecasts are short-term statistical extrapolations of past and current data. They assume that the underlying trend is informative of short term developments, without requiring other assumptions of how the SARS-CoV-2 virus is spreading, or whether preventative policies are effective. As such they are complementary to forecasts from epidemiological models.

The forecasts are based on extracting trends from windows of the data, applying machine learning, and then computing forecasts by applying some constraints to this flexible extracted trend. The methods have previously been applied to various other time series data and have performed well. They are also effective in this setting, providing better forecasts than some epidemiological models.

KEYWORDS: *Autometrics*; Cardt; COVID-19; Epidemiology; Forecasting; Forecast averaging; Machine learning; Smoothing; Trend Indicator Saturation.

1 Introduction

Our aim is to provide short-term forecasts of the number of confirmed cases and deaths attributed to COVID-19. These forecasts may be a useful guide as to what happens in the next few days. For example, the reports on 2020-03-17 that Italian deaths increased by 16% was largely in line with our forecast of 18%, and need not have been the surprise it was at the time.

The target variables consist of cumulative daily counts, which grow exponentially in the initial epidemic phase. For forecasting to be effective, we need to step down, not just to the daily increments, but the change in the daily increments. At some stage in the spread of a virus the counts will settle down: at most the entire susceptible population can be infected, but usually this is well before that happens.

Estimation of the trend is subject to several data challenges. First of all, policy interventions aim to suppress the transmission of the SARS-CoV-2 virus that causes COVID-19. In the UK, an alarming scenario forecast, Ferguson (2020), led to a switch from mitigation to lockdown and isolation. Futhermore, countries have different testing strategies and technologies, and these are occasionally revised. Some

*This is a preliminary draft: please do not quote without permission from the authors. Financial support from the Robertson Foundation (award 9907422), Institute for New Economic Thinking (grant 20029822), and the ERC (grant 694262, DisCont) is gratefully acknowledged. We wish to thank Max Roser and Our World in Data (`ourworldindata.org`) for very helpful information regarding data sources, and their informative COVID-19 page. We wish to thank Johns Hopkins University/CSSE, the New York Times, and ECDC for their data collection. We also wish to thank Neil Ericsson and Andrew Martinez for helpful comments. Disclaimer: these results are solely the work of the authors, and cannot be considered official forecasts from their respective institutions. A preliminary version of this paper was made available at www.doornik.com/COVID-19 on 19 March 2020. Any opinions are those of the authors only. email: jennifer.castle@magd.ox.ac.uk, jurgen.doornik@nuffield.ox.ac.uk and david.hendry@nuffield.ox.ac.uk

countries include asymptomatic cases, others not. Deaths are sometimes recorded a few days late, and it is not always clear whether COVID-19 was the cause. Deaths are counted differently too, e.g., the UK only counts deaths in hospitals, while other countries may also count those in care homes. So the counts will be subject to structural breaks, underreporting, definitional changes, delays, and errors. Nonetheless they are the focus of the media and in the UK, Government briefings, and the target of our forecasts.

Epidemiological models have a sound theoretical basis and a history of useful applications. Nevertheless, novel viruses and the susceptible population may behave in different ways from what models assume. Not only are pandemic data highly non-stationary, but so are the methodologies used for reporting the pandemic data, with stochastic trends and distributional shifts. Thus, there is a compounding effect as the non-stationarity of the underlying data interacts with the non-stationarity of the reporting process. Viable forecasting models must be able to handle this quadruple non-stationarity: two forms (stochastic trends and shifts) from two sources (outcomes and measurements thereof); Clements and Hendry (1999) discuss this issue for economic time series. Epidemiological models can be too highly driven by their assumptions, which, combined with their assumed mathematical processes, can limit their usefulness in forecasting until the epidemic has settled, as they are not empirical enough. As a consequence, there is an important role in short-term forecasting after distributional shifts for adaptive data-based models using a class we call 'robust', namely devices that avoid systematic forecast failure after sudden distributional shifts, also see Castle, Clements, and Hendry (2015). However, when forecasting that adaptability must remain firmly controlled to avoid excess volatility.

The methodology to construct the robust forecasts involves several steps. First, the observed daily time series is decomposed into a trend and a remainder term. The trend is estimated by taking moving windows of the data and saturating these by linear trends. Selection from these trends is made with an econometric machine learning algorithm, and the selected linear trends are then averaged to give the overall flexible trend. Next, the trend and remainder terms are forecast separately using the Cardt method and recombined in a final forecast. Cardt is a somewhat improved version of Doornik, Castle, and Hendry (2020), see Castle, Doornik, and Hendry (2019).

We made our first tentative forecasts using data up to 16 March 2020, starting data collection the day before. We reported our first public forecast on `www.doornik.com/COVID-19` on 20 March 2020 (which we will write in ISO format as 2020-03-20), using data up to 2020-03-18 and forecasting for 2020-03-19 to 2020-03-23. From then, we usually updated the forecasts every other day, mostly late afternoon or early evening (UTC) when the data sources have been updated.

One advantage of presenting real-time forecasts is that these cannot be biased by knowing what has happened – obviously, with such a major event, there is a large amount of information communicated every day. But we consider it acceptable to make small adjustments to our procedures as we learn, and have more time for the implementation.[1] Occasionally we need to correct minor errors in our coding.[2]

To reflect the real-time nature of this paper, we provide dates of when the forecasts were made and make it clear when later information is used for assessment. Publication on our website was usually made on the same day, except for the very beginning, where there was a delay of up to two days.

The outline of the paper is as follows. The underlying data is discussed in §2. The initial forecasts are introduced in §3. §4 explains the methods, followed by a discussion in §5 on how it they are used in the COVID-19 setting. §6 assesses the accuracy of our forecasts, and compares them to forecasts from several epidemiological models. The final section concludes.

[1] These would then be used from that point onwards. A month later we can confirm that many improvements to the presentation of results was made, but only very minor changes to the forecast procedure. We did introduce an additional and different method to complement forecasting when the daily increments stop rising, based on path indicator saturation with scenarios from the Chinese experience. This is documented in the companion paper Castle, Doornik, and Hendry (2020).

[2] All results have been coded in Ox (Doornik, 2018), using OxMetrics 8.2 (Doornik and Hendry, 2018). The HTML pages are also generated using Ox code. Data download is automated using the (prerelease) DataFetch package.

2 Data source

We use the data repository for the 2019 Novel Coronavirus Visual Dashboard operated by the Johns Hopkins University Center for Systems Science and Engineering (JH/CSSE). This is currently updated daily and located at `github.com/CSSEGISandData/COVID-19`. The data consisted of confirmed cases and deaths. Recovered cases are also included, but these have always not been available from this source. Coverage was all countries, Chinese provinces (and similar administrative areas), US states, and some cruise ships (which we ignore).

A dataset for modelling is created from this with minor adjustments. First, observations are put in columns with ISO date labels, and a few countries are renamed to be closer to their ISO name. Next, for France, Denmark, United Kingdom, and Netherlands we only include the mainland tallies. Aggregates for China and EU-27 are constructed, and later also the world.

On 2020-03-23 the Johns Hopkins data omitted US states. This gap was filled later by the New York Times, collecting data from state-level health authorities. We used this from the 27th onwards. Their US state data can be downloaded from Github at `github.com/nytimes/covid-19-data`. The New York Times started redefining the US state data on 2020-05-06. At that stage only the last observation was adjusted (e.g New York deaths jumped from 19645 on 2020-05-05 to 25956 a day later), rendering it useless as a historical data set. Since Johns Hopkins/CSSE had reintroduced US disaggregate data, we switched to using that instead. Note the aggregate US count was always based on JH/CSSE.

All forecasting relates to the cumulative counts of 'confirmed' and 'deaths' separately. The regions and countries for which we publish forecasts changes over time, based on our interests, subject to a minimum amount of 250 confirmed cases or 30 deaths (increased later to 2000/200 respectively), see Appendix A. Appendix B considers data revisions.

3 Initial forecasts for confirmed cases

Forecasts of confirmed cases of COVID-19 and deaths from COVID-19 are obtained, commencing with a few countries only. Data is available from 2020-01-22, although the first confirmed case in many countries is later in the sample, e.g. 2020-01-31 in the UK. The last observation is for 2020-03-16, and the forecasts were finalized by the next day. This section introduces our approach to forecasting, with details of the methodology given in the sections below.

Figure 1 shows the forecasts of confirmed cases for the UK, EU-27, US and China. In each graph we report the observed value in a grey line marked with dots. Forecasts are made from 9 March onwards from estimates up to 8 March. These are the red crosses in the graphs. Next, we make two weeks of daily forecasts from 17 March onwards: the red circles. The thin lines are 60% forecast intervals – these are very large and are reconsidered later.

The UK and EU in Figure 1 seem to be on similar trajectories. The UK had 1543 confirmed cases on 2020-03-16, and we forecast $10\,000$ around 25/26 March.[3] The adopted mitigation policies may reduce this, which then will be reflected in updated forecasts.

The EU had almost $60\,000$ cases at the end of the sample, and is predicted to reach $250\,000$ around 25 March. But again, measures are in place now in an effort to reduce this. The US here is on a more rapid growth path.[4] This could be a reflection of the different approach to handling the spread of SARS-CoV-2.

In each case, the older forecasts are already remarkably effective. Where they overlap, the difference is small, except for the UK where they are starting to fall below the later forecasts.[5] China is included

[3]The subsequent outcomes were 9529 cases on 25 March and 11658 a day later.

[4]The outcome for the EU was 252 770 on 2020-03-26. Even the US forecasts are remarkably accurate, forecasting 79 700 for 2020-03-26 when the actual was 83 836 and starting from 4632 as the last observed value at the time of forecasting.

[5]When the forecasts were made, both were plausible, but subsequently the higher turned out to be the more accurate path.

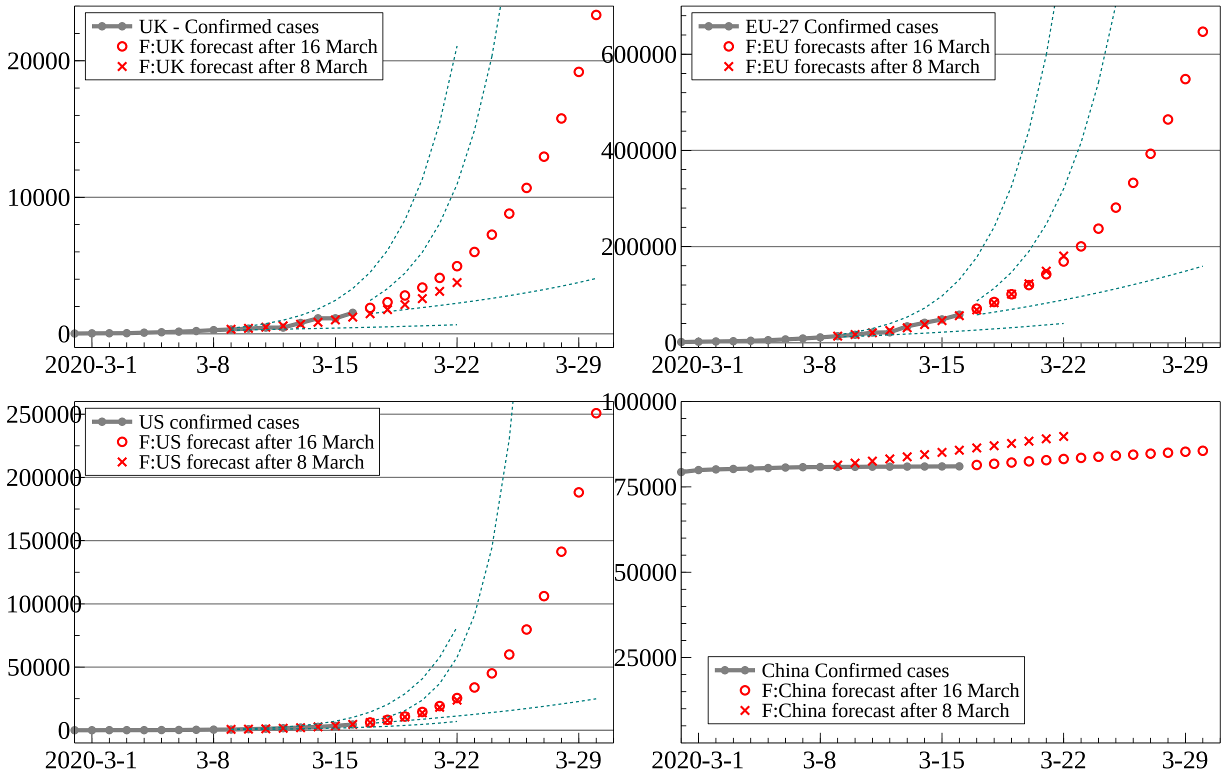

Figure 1: Forecasts of confirmed cases of COVID-19 for UK, EU-27, US, and China. Johns Hopkins/CSSE Data collected and forecast made on 2020-03-17.

for contrast, because there the epidemic has run its course (assuming there is no later recurrence). Our methods clearly over forecast here: in order to capture the exponential growth early on, some trend remains when it is no longer needed. This will need to be addressed later, because it means that we expect to over forecast when the inflection point is reached.

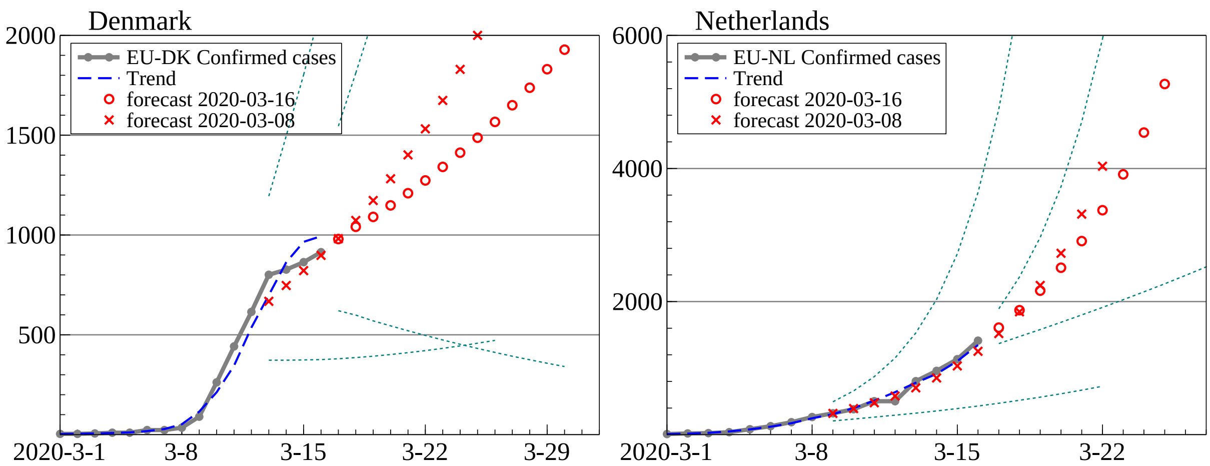

Figure 2: Forecasts of confirmed cases of COVID-19 for mainland Denmark and the Netherlands. Data from 2020-03-17.

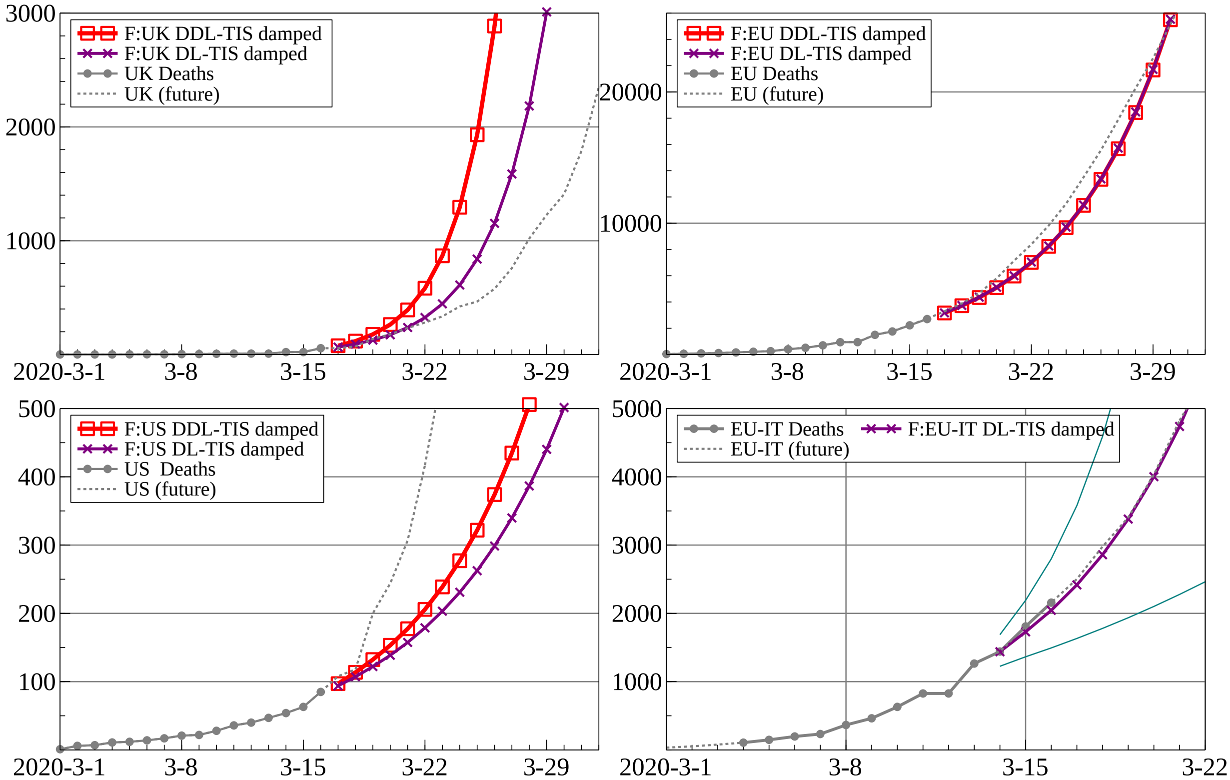

Figure 3: Forecasts of deaths from COVID-19 for UK, EU-27, US, and Italy. Data and forecasts from 2020-03-17, subsequently updated with later data (dotted line).

Figure 2 shows the forecasts of confirmed cases for Denmark on the left and the Netherlands on the right (mainland only for both countries). The surface area of Denmark and the Netherlands is almost identical, but the population of the Netherlands is three times higher. Correspondingly, we set the vertical scale of the graph for the Netherlands to three times that of Denmark. The trajectories are quite different, and, in Denmark's case, show a trend break. The Netherlands has one outlying observation on 12 March.

3.1 Results for death counts

Figure 3 reports preliminary short-term forecasts for the death count. The graphs are for the UK, EU-27, the US, and Italy separately, starting forecasting from the last observation at 2020-03-16, except for Italy where the start is two days earlier.

Two different models are used, DDL-TIS and DL-TIS, as explained below. When these forecasts were made on 17 March, the observations on mortality were still limited, particularly for the US and the UK, making forecasting more difficult. The graphs also show the 'future' in the dotted line. Thus, afterwards we can confirm that the forecasts for Italy were extremely accurate, for the EU fine, but for the UK far over, and the US far under.

4 Methodology

We use local averaged time trend estimation (LATTE, Doornik, 2019) to decompose the dependent variable $y_t, t = 1, ..., T$ in a *trend* term μ_t, and *residual* or irrregular ε_t; seasonality is assumed to be absent (there are likely to be weekend effects in most countries, but these are not modelled). For the logarithmic model:

$$\log y_t = \widehat{\mu}_t + \widehat{\varepsilon}_t,$$

from which we obtain

$$y_t = \exp(\widehat{\mu}_t)\exp(\widehat{\varepsilon}_t).$$

To allow for the counts of zero before the SARS-CoV-2 virus took hold, we replace the specification with

$$x_t \equiv \log(y_t + 1) = \widehat{\mu}_t + \widehat{\varepsilon}_t, \tag{1}$$
$$y_t \approx [\exp(\widehat{\mu}_t) - 1]\exp(\widehat{\varepsilon}_t), \tag{2}$$

with (2) not exactly following from exponentiation of (1).

The sample $x_1, x_2, ..., x_T$ is split in overlapping windows, and for each window a trend indicator saturation (TIS) model is estimated: the model is saturated with linear trends and selection with *Autometrics* (Doornik, 2009) then obtains a sparse linear regression model. For a typical window w:

$$\widehat{x}_{w,t} = \widehat{\alpha}_w^F + \widehat{\beta}_w^F t + \sum_{s\in\mathcal{T}_w} \widehat{\theta}_{w,s}(t-s-1)I(t \le s), \quad t = T_w, ..., T_{w+1} - 1. \tag{3}$$

The (broken) linear trend $(t - s - 1)I(t \le s)$ starts at a negative value, then increases by unity until it hits zero, after which it stays at zero. The superscript F indicates that those terms are always included. $\mathcal{T}_w$ is the subset of trends that were selected for this data window. In contrast to many other trend-cycle decomposition methods, this approach can handle smooth changes as well as abrupt breaks (although that seems less important in the current setting, so other smoothing methods could conceivably be used).

The LATTE estimates of the trend and residual from W windows are:

$$\widehat{\mu}_t = W^{-1}\textstyle\sum_{w=1}^{W} \widehat{x}_{w,t},$$
$$\widehat{\varepsilon}_t = y_t - \widehat{\mu}_t.$$

Model (3) is called L-TIS. The following two variants allow for a quadratic and cubic trend respectively:

$$\text{DL-TIS} \quad \Delta\widehat{x}_{w,t} = \widehat{\alpha}_w^F + \widehat{\beta}_w^F t + \sum_{s\in\mathcal{T}_w} \widehat{\theta}_{w,s}(t-s-1)I(t \le s), \tag{4}$$
$$\text{DDL-TIS} \quad \Delta\Delta\widehat{x}_{w,t} = \widehat{\alpha}_w^F + \widehat{\beta}_w^F t + \sum_{s\in\mathcal{T}_w} \widehat{\theta}_{w,s}(t-s-1)I(t \le s). \tag{5}$$

DL-TIS introduces cumulated differences in the trend, making $\widehat{\mu}_t$ an I(1) variable[6] with quadratic trend, while DDL-TIS makes $\widehat{\mu}_t$ I(2) with up to a cubic trend. For economic data it is common to restrict the model to a linear trend.

We also consider step-indicator saturation (SIS), where the model is saturated with broken intercepts (steps) instead of trends. The corresponding models are DDL-SIS and DL-SIS. See Castle, Doornik, Hendry, and Pretis (2015, 2020) for SIS and TIS respectively, and Walker, Pretis, Powell-Smith, and Goldacre (2019) for another application of TIS.

[6]An I(1) variable is stationary when differenced, an I(2) variable needs differencing twice for stationarity, see Johansen (1995), Doornik and Juselius (2018) *inter alia*.

The generality provided by the I(2) model with up to cubic trend can become a burden for forecasting: too much flexibility can lead to wild forecasts. The adopted forecasting device is Cardt (Castle, Doornik, and Hendry, 2019), which allows for up to I(1) with linear trend, making automatic decisions about whether to use differencing. Cardt takes the average of three forecasting models, two autoregressive and one moving average, followed by calibration where the forecast are treated as pseudo-observed values. This performs very well on the data from the M4 and M3 forecast competitions (Makridakis, Spiliotis, and Assimakopoulos, 2020). The automatic choice between forecasting in differences or levels is too limited here, so several possible extensions are considered beyond the default Cardt forecasts (6):

$$\widehat{\mu}^{(0)}_{T+h} = \mathrm{Cardt}(h \mid \widehat{\mu}_1, ..., \widehat{\mu}_T) \qquad \text{[standard]}, \tag{6}$$

$$\widehat{\mu}^{(1)}_{T+h} = \widehat{\mu}_T + \textstyle\sum_{s=1}^{h} \mathrm{Cardt}(s \mid \Delta\widehat{\mu}_2, ..., \Delta\widehat{\mu}_T) \qquad \text{[up to I(2)]}, \tag{7}$$

$$\widehat{\mu}^{(2)}_{T+h} = (\widehat{\mu}^{(0)}_{T+h} + \widehat{\mu}^{(1)}_{T+h})/2 \qquad \text{[damped I(2)]}, \tag{8}$$

$$\widehat{\mu}^{(3)}_{T+h} = (\mathrm{Cardt}(h \mid \widehat{\mu}_{T-7}, ..., \widehat{\mu}_T) + \widehat{\mu}^{(1)}_{T+h})/2 \qquad \text{[short damped I(2)]}. \tag{9}$$

Forecasts (6) are targeted at economic applications. Forecasts (7) apply Cardt to the differenced trend, which is then reintegrated. The differences can then have a damped trend. Next, (8) is the simple average of the previous two. This may seem ad hoc, but leads to an effective forecasting device; (7) is not used on its own, because it is too strongly trending. Forecasts (9) use a shorter sample for the standard forecasts, and are used when the cumulative counts start to slow.

Cardt forecasts are also made for the residual term $\widehat{\varepsilon}_t$. This is combined into the final forecast, which for the damped I(2) version yields:

$$\widehat{y}_{T+h} = \left[\exp(\widehat{\mu}^{(2)}_{T+h}) - 1\right] \exp(\widehat{\varepsilon}_{T+h}).$$

The forecast intervals in Figures 1 and 2 are based on damped I(2) and deemed too wide by quite a margin. Subsequently, we changed the intervals to those obtained from Cardt applied to the levels rather than the logarithms. Cardt gives upper and lower confidence bounds of the central forecasts for an adopted confidence α. At least in the M4 settings this worked very well, with on average about $100(1-2\alpha)$ of outcomes inside the confidence interval. In practice we used $\alpha = 0.1$, aiming for 80% forecast confidence intervals. Let $\widehat{\mu}^{(2)}_{T+h}(\mathrm{H})$ be the upper bound from the damped I(2) forecasts, and $\widehat{y}_{T+h}(\mathrm{H})$ the upper bound for the levels forecast, then we construct new upper bounds as:

$$\widehat{\mu}_{T+h}(\mathrm{H}) = \min\left\{\widehat{\mu}^{(2)}_{T+h}(\mathrm{H}) - \widehat{\mu}^{(2)}_{T+h}, \log[\widehat{y}_{T+h}(\mathrm{H})] - \log[\widehat{y}_{T+h}]\right\}.$$

The analogous procedure is used for the lower bounds $\widehat{\mu}_{T+h}(\mathrm{L})$.

5 Adaptation to COVID-19

Several different specifications of the model and forecasting approach were tried for the results in §3. The adopted versions are in Table 1. When countries are listed together, a multivariate DDL-TIS model was estimated, but each forecast is made separately. When two methods are given in the table, the forecasts are the equally weighted average of these. In most cases the DDL-TIS model with damped I(2) forecasts was used initially. The X in DDLX-TIS indicates that the data were extended by forecasts prior to trend extraction.

It was not considered practical to find the best formulation for each country separately when forecasts are made every day. Moreover, this may not stay constant over time. Instead, we adopted the alternative where we supply two forecasts, the first based on DDL-TIS, and the second on an average over several forecasts:

Data	*Countries*	LATTE	*Forecasts starting* 2020-03-09	2020-03-17
Confirmed	UK,EU	DDL-TIS	damped I(2)	damped I(2)
Confirmed	US	DDL-TIS	damped I(2)	I(2)
Confirmed	China	DDL-TIS	damped I(2)	damped I(2)
Confirmed	DK	DDLX-TIS	standard	damped I(2)
Confirmed	NL	DDL-TIS	damped I(2)	damped I(2)
Deaths	UK,EU,US	DL-TIS, DDL-TIS		damped I(2)
Deaths	IT	DL-TIS		damped I(2)

Table 1: Model and forecast specifications used in the results reported in §3.

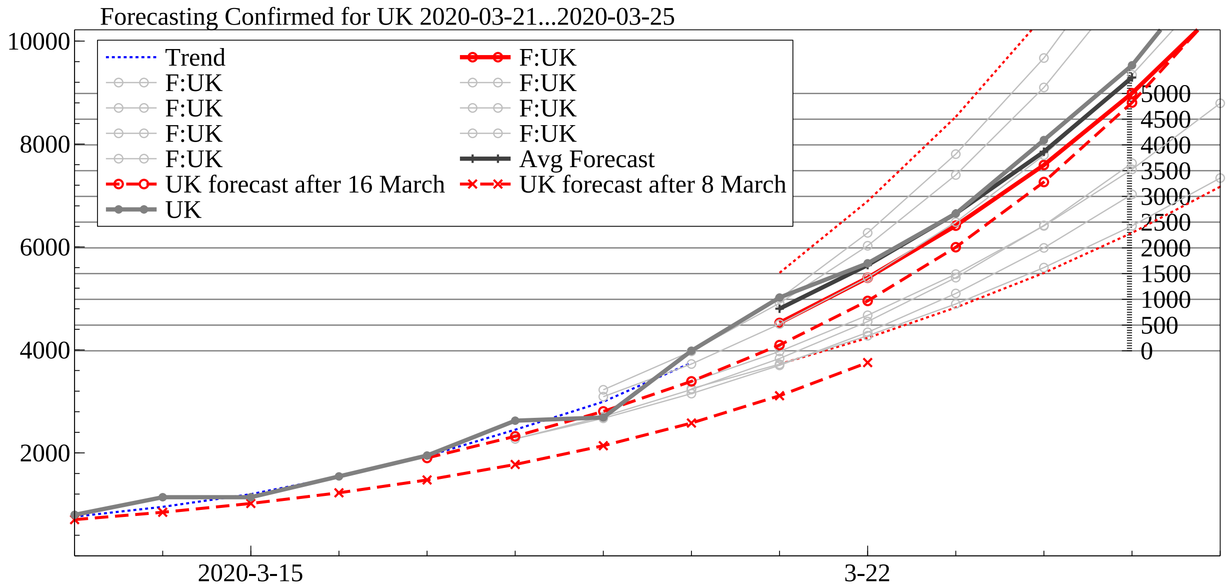

Figure 4: Forecasts of confirmed cases of COVID-19 for UK from 2020-03-17, together with forecasts from 2020-03-21.

F: DDL-TIS with damped I(2), estimated up to T, forecasting $T+1, T+2, \ldots$.

Avg: Average of eight forecasts: DDL-TIS and DL-TIS both with damped I(2), each estimated up to $T, T-1, T-2, T-3$. The first of these is forecast F. The average includes a forecast for observation T, which is already known; the whole average path is shifted to match this last known value exactly.

When the inflection point is reached, the methods change to:

F: DDLX-SIS with short damped I(2), where the X indicates that the raw data are extended by forecasts prior to trend estimation.

Avg: Average of four forecasts: DDLX-SIS with short damped I(2), each estimated up to $T, T-1, T-2, T-3$. The average includes a forecast for T, which is known; the whole average path is shifted to match this last known value exactly.

In both cases, the first forecast path (the solid red line labelled 'F' in Figure 4 and the graphs on our

	Confirmed cases							**Deaths**						
	1 step			2 step		4 step		1 step			2 step		4 step	
Area	*Count*	Avg	F	Avg	F	Avg	F	*Count*	Avg	F	Avg	F	Avg	F
Canada	18	2.4	2.5	4.1	4.3	6.5	6.4	15	6.1	6.4	8.0	9.3	12.3	14.1
EU	18	1.7	1.8	3.2	3.2	8.1	7.5	18	1.1	1.0	2.1	1.7	5.4	3.2
EU-DE	18	1.0	1.1	2.5	2.1	5.9	4.0	18	3.7	4.5	6.9	7.9	11.7	13.5
EU-ES	18	1.6	1.5	3.0	2.3	7.4	5.2	18	1.2	1.9	2.1	2.8	5.0	3.5
EU-IT	18	0.4	0.4	1.0	0.8	2.7	1.7	18	0.8	0.7	1.7	1.3	4.7	3.1
Iran	18	0.8	0.7	2.0	1.5	5.5	3.9	18	0.4	0.4	1.1	0.9	3.2	2.2
Switzerland	18	0.9	0.7	1.9	1.1	5.8	3.3	18	2.3	3.4	4.0	5.6	7.4	8.5
UK	18	1.7	2.1	3.2	3.1	6.5	5.2	18	3.6	4.0	8.2	8.1	16.3	13.6
US	18	1.0	1.3	2.7	2.5	7.6	6.3	18	4.2	5.1	8.8	9.1	15.0	14.0
World	13	0.9	1.3	1.8	2.1	4.7	4.6	13	1.8	1.9	4.2	4.0	10.4	9.3
All	705	1.7	1.9	3.2	3.2	6.6	5.9	506	3.1	3.9	5.7	6.8	9.6	11.0
Coverage %	705	98.7	95.3	96.5	90.2	83.9	79.2	506	95.8	86.6	88.2	77.3	76.7	65.3

Table 2: Forecast accuracy for different geographical areas. MAPE over 2020-03-24 to 2020-04-25 for each area for 1,2,4 step ahead forecasts.

website) is preferred. However, when the average forecast (the solid black line labelled Avg Forecast) deviates considerably, this could mean that there has been a sudden recent change.[7] In that case it can be difficult to decide between the two.

Figure 4 shows one instance of the forecasts for the UK. It is an updated version of the top left plot from Figure 1. In addition to the earlier forecasts, it shows the eight individual forecasts (unshifted), with the first one (DDL-TIS with damped I(2)) in bold red. The bold black line is the forecast average, which is close to the red line in this case. The updated outcomes are shown as well, with the forecasts tracking these very closely.

6 Assessment of forecast accuracy

6.1 Forecast accuracy

The forecast accuracy is measured as the error in percentage of the outcome, computed for almost all the forecasts that we have published. Following Appendix B, we restrict evaluation to the period 2020-03-24 to 2020-04-25, and omit France and Mexico. Reported is the mean absolute percentage forecast error (MAPE), which is, for a forecast $\widehat{y}_{j,T+H}$ from group $j = 1, ..., J$:

$$\text{MAPE} = \frac{100}{J}\sum_{j=1}^{J}\frac{|y_{j,T+H} - \widehat{y}_{j,T+H}|}{y_{j,T+H}}.$$

Table 2 gives the average MAPE for a selection of geographical regions, as well as all together. Count is the number of forecasts in the 1-step ahead percentage errors. The accuracy for deaths is about half that of confirmed, except for Spain, Iran, and the whole EU, where it is reversed. On the whole,

[7]Very occasionally, the end point of the estimated trend seems to be estimated unsatisfactorily. Forecast paths that go down were discarded, but this also only happened occasionally, and only with earlier forecasts.

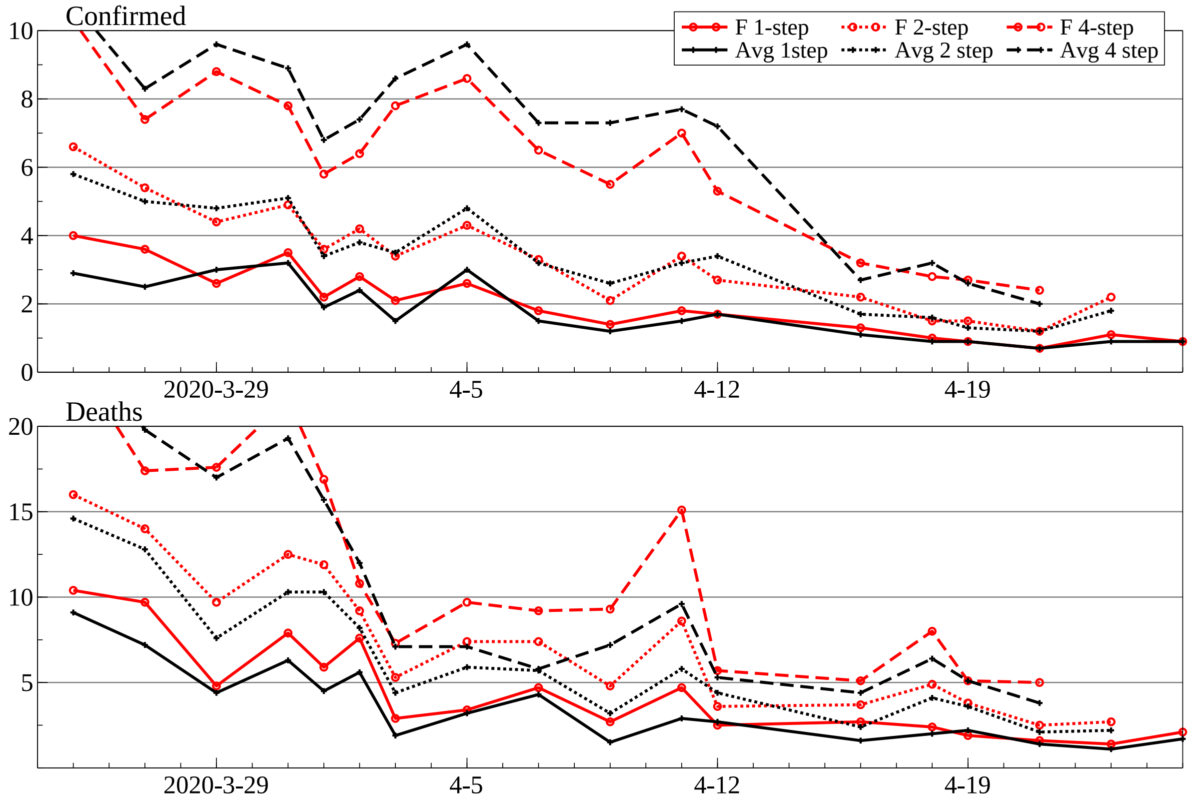

Figure 5: Forecast accuracy over time. MAPE for each target date for 1,2,4 step ahead forecasts. Confirmed at top, deaths at bottom.

our forecasts show good accuracy, giving useful insights to policy makers of developments that can be expected in the coming days.

The coverage gives the percentage of forecasts that are within the quoted forecast interval. As we were aiming for 80% forecast confidence intervals, these are close to the target for 4-steps ahead forecasts. They are too wide for one-step ahead, and will be too narrow beyond 4 steps.

Figure 5 plots the MAPE for 1,2 and 4-step ahead forecasts, for confirmed cases (top) and deaths (bottom). Accuracy is given for the forecast F, as well as the average forecast (Avg). All lines show that accuracy increases over time. This is not so much from the accumulation of information, but more from the changing state of the pandemic, moving through exponential growth to a slow down.

For confirmed cases, the 1-step ahead error ranges from around 3% at the start to below 1% at the end. As expected, accuracy decreases as the forecast horizon grows. For 1 and 2-steps ahead, the average forecast and F are close, but for 4 steps F mostly dominates. This is the other way round for deaths, where the average forecast is mostly better. It is likely that increasing the death threshold for inclusion (see Appendix A) contributed to the improvement from early April. Nonetheless, forecasts of deaths are less accurate than confirmed cases: the former data is more noisy.

6.2 A comparison with MRCGIDA epidemiological models

The MRC Centre for Global Infectious Disease Analysis at Imperial College London (MRCGIDA, but shortened to MCIC below) started publishing weekly forecasts of deaths from the 8th of April onwards

Week ending	*Count*	*MAE*		*MPE(W)*		*MPE(T)*		*MAPE(W)*		*MAPE(T)*	
		MCIC	F	MCIC	F	MCIC	F	MCIC	F	MCIC	F
up to 2020-04-04	29	1068	629	−66	26	−43	21	71	40	47	27
2020-04-11	23	1912	661	−87	−2	−40	2	91	31	42	16
2020-04-18	24	372	372	−31	−12	−12	−3	35	30	13	10
2020-04-25	24	1101	1108	−23	−10	−7	−1	32	29	9	8
2020-05-02	27	388	372	−62	−33	−10	−6	65	43	11	7
2020-05-09	28	161	166	−9	−11	−1	0	22	26	3	3

Table 3: Mean absolute errors (MAE) and mean [absolute] percentage errors (M[A]PE) of MCIC and our forecasts F. (W) is as a percentage of the weekly total, (T) is relative to the cumulative total at the end of the week.

for a selection of countries, Bhatia (2020). Their forecasts are a weighted average of three or four Bayesian epidemiological models.

Using our notation, y_t is the cumulative reported death count and T the last available observation. The MCIC reports forecasts $\widehat{w}_{T+7} = \Delta_7 \widehat{y}_{T+7}$, where the observed value a week later will be $y_{T+7} - y_T$, but only once a week. We forecast $\widehat{y}_{T+1}, ..., \widehat{y}_{T+7}$ roughly every other day, which can be mapped to weekly forecasts $\widehat{y}_{T+7} - y_T$, so directly compared to the MCIC forecasts. One difference is that the MCIC uses European Centre for Disease Prevention and Control (ECDC) data, while we use the Johns Hopkins (JH) data. Because of construction in different time zones, the datasets differ by one day, e.g. ECDC reports a weekly increase in UK deaths of 3294 on 4 April, but JH has this same number for 5 April. There are occasional differences in the reported data, so we compare each forecast to its own data set. The evaluations are aligned by dates, but we only refer to the JH dates.

The MCIC report of 2020-04-15 (UK date) has forecasts for the week starting 2020-04-11 and ending 2020-04-18 (JH dates; 12-19 April in ECDC dates). The MCIC report dated 2020-04-08 has forecasts for the week before, but also contains forecasts for several previous weeks, for which the outcome was already known at the time of publication.

The forecast comparison is restricted to the countries and periods for which we both produced forecasts, resulting in $j = 1, ..., 100$ weekly forecast errors $u_j = \widehat{w}_{j,T+7} - \Delta_7 y_{j,T+7} = \widehat{y}_{j,T+7} - y_{j,T+7}$. We compute the mean absolute error (MAE), as well as two versions of the mean absolute percentage error (MAPE). For J forecast errors:

$$\text{MAPE(W)} = \frac{100}{J}\sum_{j=1}^{J}\frac{|u_j|}{\Delta_7 y_{j,T+7}},$$

$$\text{MAPE(T)} = \frac{100}{J}\sum_{j=1}^{J}\frac{|u_j|}{y_{j,T+7}}.$$

Appendix B shows that data revisions in this period do not change the results.

Table 3 reports the error measures for the two forecasts, for the weeks ending up to 4 April combined, and for each week afterwards. The first column after the counts gives the MAE, which is influenced by several large errors. The largest MCIC error in weekly death counts is for France in the week ending 2020-04-11 at $-17\,500$, so an over forecast, compared to our error of -6300. The second largest error

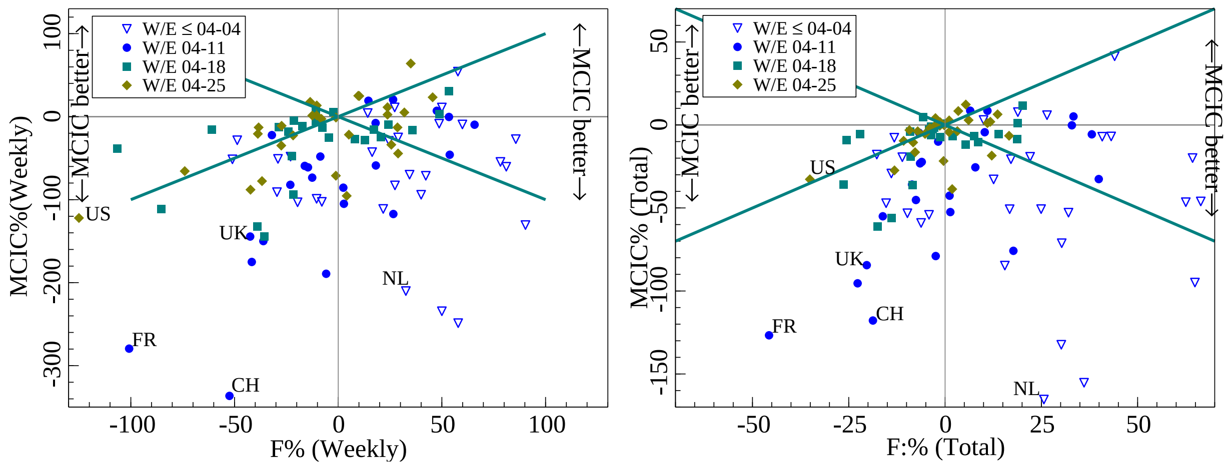

Figure 6: Weekly forecast percentage error of deaths from COVID-19 for MCIC and our forecasts up to week ending 2020-04-25. Relative to weekly count (left) and end-of-week cumulative total (right). Week endings marked by different symbols.

is for the US for the week ending 2020-04-25, where we both over forecast by about 18 000. The next is for the UK, where the MCIC forecast was 13900 (week ending 2020-04-11), our forecast 8457 and the outcome 5562.

The next four columns in Table 3 show the mean percentage error, relative to weekly counts and end of week totals. This shows a negative bias for MCIC throughout, but diminishing as the pandemic progresses. Ours is positive initially, but then close to zero. The final four columns show the two measures of the MAPE. Initially, the MIC errors where about two to three times higher. For the last two weeks, when the daily counts in most countries started to go down, the gap is much smaller.

In both cases the forecast confidence intervals seem too narrow. The MCIC reports 95% intervals, but has only 67% inside (out of the 100 considered forecasts); we report 80% intervals, with just 58% inside (both over all errors up to week ending 2020-04-25).

Figure 6 gives cross plots of the two measures of forecast percentage errors. Some large entries have been marked with the country name, the same for the left and the right panel. For the US, France and UK marks these also correspond to the large absolute errors discussed above.

6.3 A comparison for states of the USA

Two sources of US forecasts are available over a long enough period to allow a comparison with our forecasts. The first is by the Institute for Health Metrics and Evaluation (IHME, `healthdata.org/covid`) starting 2020-03-25; the second source of forecasts is provided by Los Alamos National Laboratory (LANL, `covid-19.bsvgateway.org`, from 2020-04-05 onwards).

Evaluation is against each institute's report of the outcomes, taken as shortly after the forecast date as possible. This reduces the impact of data revisions, and is preferred over using the most recent realization.

As it happens, our respective choice of start dates did not align so well, so we filled in several gaps in our forecast history, using the procedure we would have used at the time. Following the approach of the previous section, the comparisons are bilateral, only including those cases where target date and area match for both sources of forecasts. We restrict the results to states with a large enough count, see Appendix A, which are also the most relevant ones. Only forecasts of up to seven days ahead are

	CA	*CO*	*CT*	*FL*	*GA*	*IL*	*LA*	*MA*	*MI*	*NJ*	*NY*	*WA*
IHME	8.5	9.2	8.2	17.8	16.7	16.9	8.5	18.0	10.8	10.1	8.7	4.4
Avg	4.3	11.9	6.8	4.8	12.1	7.7	4.8	8.8	6.2	5.9	4.1	4.3
F	6.6	11.8	7.7	8.9	16.1	10.1	5.1	9.4	8.9	7.3	4.5	4.4
Count	49	35	49	49	49	49	49	49	49	49	49	49
LANL	5.6	8.9	14.7	5.1	11.6	4.6	11.7	8.4	10.7	5.1	6.9	4.0
Avg	3.9	9.5	6.7	4.5	10.5	4.9	5.2	5.9	5.3	5.6	4.1	4.1
F	6.9	6.7	6.6	9.9	15.2	4.8	7.6	8.9	6.4	5.5	2.7	3.3
Count	49	35	49	49	49	49	49	49	49	49	49	49

Table 4: Forecasts of deaths from 1 to 7 days ahead: mean absolute percentage errors relative to the outcome for US states with substantive cases.

Horizon	Deaths			Deaths			Confirmed		
	IHME	*Avg*	*F*	*LANL*	*Avg*	*F*	*LANL*	*Avg*	*F*
1	6.5	3.6	3.3	3.3	2.3	2.8	2.2	1.4	1.6
2	7.6	4.8	4.8	5.8	3.4	4.5	3.2	2.6	2.7
3	9.1	5.9	7.1	7.2	4.4	6.0	4.1	4.0	4.0
4	11.9	6.4	8.7	8.5	5.8	7.6	5.3	5.5	5.2
5	13.2	7.2	10.2	9.9	6.8	8.9	6.1	7.2	6.6
6	15.0	8.4	11.4	10.6	7.6	9.3	7.0	9.3	8.2
7	17.5	10.5	12.7	11.3	10.0	10.2	8.1	12.2	10.0

Table 5: Forecasts of deaths and confirmed cases for 12 US states: mean absolute percentage errors relative to the outcome for different forecast horizons. There are 82 forecast errors at each horizon for deaths, and 72 for confirmed.

included, in line with our focus on short-term forecasting. For IHME we use forecasts from seven reports from 2020-04-05 to 04-27 with a subsequent one for actual values; for LANL also seven reports from 2020-04-05 to 04-26 with two subsequent reports for actual values.

Table 4 gives the MAPE of the relative forecast errors from one to seven days ahead of cumulative deaths for 12 states. By the end of April these 12 states accounted for just over 70% of confirmed cases and deaths in the United States. The first part of the table is based on forecasts that we have in common with IHME, the second part on those with LANL. Despite the fact that the number of forecasts included is the same for both, it is a different set, so the IHME and LANL forecasts cannot be compared directly to each other.

Table 5 splits the results by forecasts horizon, from one to seven days ahead. The same set is included for deaths as in Table 4. There are 82 forecasts errors at each horizon, but again a different set in the IHME and LANL part of the table. Both our forecasts F and Avg dominate at each horizon for deaths, with the average the best in terms of MAPE, as was the case in Table 2. For confirmed cases there is not much difference at the early horizons, until we get past 4 days ahead when LANL has smaller forecast errors.

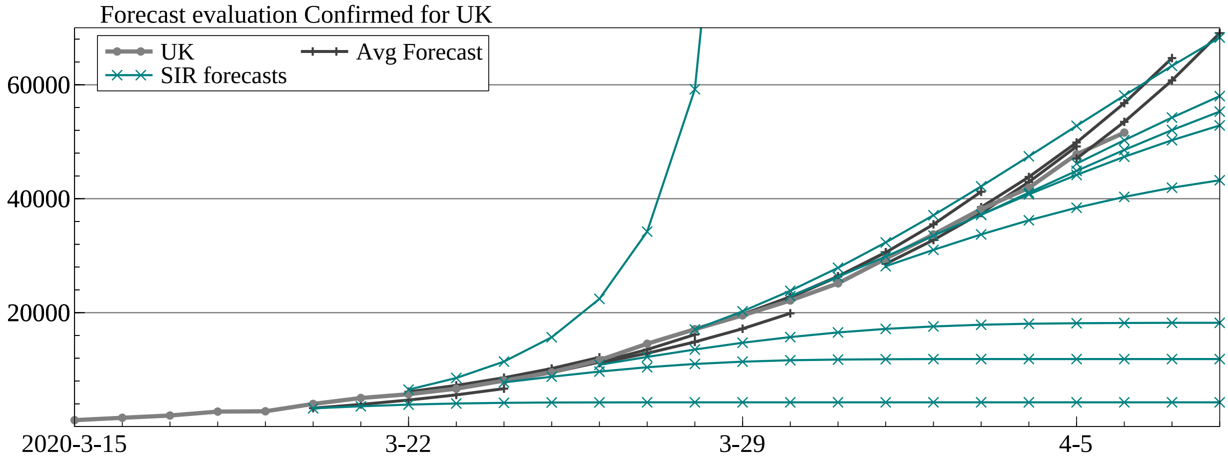

Figure 7: Forecasts of confirmed cases of COVID-19 for UK from 2020-03-20 to 2020-04-05. Average forecasts and SIR forecasts.

6.4 A comparison with simple SIR forecasts

The classic epidemic model is the SIR model, consisting of three compartments: S for the number that are susceptible, I for infectious and R for removed (either recovered or died). See Hethcote (2000) for an overview, and Wikipedia (Compartmental models in epidemiology) for an introduction. The transition rate from S to I is governed by the contact rate β, and removal by the combined recovery and death rate γ of an infected individual; $N = S + I + R$ is the total population. Then $\beta I/N$ is the average number of contacts of a susceptible person with the infectious each time period, and β the average number of potentially transmissive ('adequate') contacts of one person with another person. The simple version leads to a set of ordinary differential equations (ODE):

$$\begin{aligned}
\mathrm{d}S(t)/\mathrm{d}t &= -\beta I(t)S(t)/N,\\
\mathrm{d}I(t)/\mathrm{d}t &= \beta I(t)S(t)/N - \gamma I(t),\\
\mathrm{d}R(t)/\mathrm{d}t &= \gamma I(t).
\end{aligned}$$

These sum to zero so N is constant. Given initial conditions $R(0) = 0, I(0) = 0$, and specific values of $S(0), \beta, \gamma$, the ODE can be integrated to find the time paths of S, I, R. The model can be used to compute the epidemic evolution for assumed parameters, which will usually be done in a richer model. The basic reproduction number [8] in this model is $R_0 = \beta/\gamma$ and the replacement number $R = (\beta/\gamma)S(t)/N$.

Alternatively, we can estimate the parameters from observed data. Given daily observations of $I_t + R_t$ in the form of confirmed cases $y_t, t = 1, ..., T$, we obtain SIR residuals $\varepsilon_t(S_1, \beta, \gamma) = y_t - (I_t + R_t)$. The parameters can be estimated by non-linear least squares.[9]

Figure 7 shows the forecast paths obtained by SIR in comparison to those from our average forecast described above for the recent UK history. While the SIR model may be useful to describe the completed pandemic process, these SIR forecasts offer very little guidance to the short-term movements in UK

[8] R_0 is defined as defined as the average number of secondary infections produced when one infected individual is introduced into a host population where everyone is susceptible (Hethcote, 2000, p.601). The replacement number R is defined to be the average number of secondary infections produced by a typical infective during the entire period of infectiousness.

[9] Batista (2020) provides an overview. We prefer to reparameterize the model in terms of $N_1/S_1, \beta, \gamma$ for a chosen N_1, and maximize $-\log \sum \varepsilon_t^2/T$.

confirmed cases. The average forecasts are unable to provide the sigmoid shape over long horizons, but provide better forecasts in the early stages, and will follow the slow down.

7 Conclusion

While models based on well-established theoretical understanding and available evidence are crucial to viable policy-making in observational-data disciplines, shifts in distributions can lead to systematic mis-forecasting. Consequently, there is an important role for short-term forecasts using adaptive data-based models that are 'robust' after distributional shifts.

Our real-time forecasts of confirmed cases and deaths fulfil this role: they have been timely, relevant, and relatively accurate. Moreover, they outperform several epidemiological models.

References

Batista, M. (2020). Estimation of the final size of the coronavirus epidemic by the SIR model. Online paper, ResearchGate.

Bhatia, S. *et al.* (2020). Short-term forecasts of covid-19 deaths in multiple countries. Online reports 2020-04-15 and 2020-04-08, mrc-ide.github.io/covid19-short-term-forecasts/index.html, MRC Centre for Global Infectious Disease Analysis.

Castle, J. L., M. P. Clements, and D. F. Hendry (2015). Robust approaches to forecasting. *International Journal of Forecasting 31*, 99–112.

Castle, J. L., J. A. Doornik, and D. F. Hendry (2019). Some forecasting principles from the M4 competition. Economics discussion paper 2019-W01, Nuffield College, University of Oxford.

Castle, J. L., J. A. Doornik, and D. F. Hendry (2020). Medium-term forecasting of the coronavirus pandemic. in progress, Nuffield College, University of Oxford.

Castle, J. L., J. A. Doornik, D. F. Hendry, and F. Pretis (2015). Detecting location shifts during model selection by step-indicator saturation. *Econometrics 3(2)*, 240–264.

Castle, J. L., J. A. Doornik, D. F. Hendry, and F. Pretis (2020). Trend-indicator saturation. mimeo, Department of Economics, Oxford University.

Clements, M. P. and D. F. Hendry (1999). *Forecasting Non-stationary Economic Time Series*. Cambridge, Mass.: MIT Press.

Doornik, J. A. (2009). Autometrics. In J. L. Castle and N. Shephard (Eds.), *The Methodology and Practice of Econometrics: Festschrift in Honour of David F. Hendry*. Oxford: Oxford University Press.

Doornik, J. A. (2018). *Object-Oriented Matrix Programming using Ox* (8th ed.). London: Timberlake Consultants Press.

Doornik, J. A. (2019). Locally averaged time trend estimation and forecasting. mimeo, Nuffield College, Oxford.

Doornik, J. A., J. L. Castle, and D. F. Hendry (2020). Card forecasts for M4. *International Journal of Forecasting 36*, 129–134. doi.org/10.1016/j.ijforecast.2019.03.012.

Doornik, J. A. and D. F. Hendry (2018). *OxMetrics: An Interface to Empirical Modelling* (8th ed.). London: Timberlake Consultants Press.

Doornik, J. A. and K. Juselius (2018). *CATS 3: Cointegration Analysis of Time Series in OxMetrics*. London: Timberlake Consultants Press.

Ferguson, N. *et al*. (2020). Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. discussion paper (16 march 2020), MRC Centre for Global Infectious Disease Analysis.

Hethcote, H. W. (2000). The mathematics of infectious diseases. *SIAM Review 42*, 599–653.

Johansen, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

Makridakis, S., E. Spiliotis, and V. Assimakopoulos (2020). The M4 competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting 36*, 54–74. doi.org/10.1016/j.ijforecast.2019.04.014.

Walker, A., F. Pretis, A. Powell-Smith, and B. Goldacre (2019). Variation in responsiveness to warranted behaviour change among NHS clinicians: a novel implementation of change-detection methods in longitudinal prescribing data. *British Medical Journal 367, l5205*.

A Countries included

Forecasts were made for a selection of countries (maximum number in parentheses):

Europe (26)

(1) EU, sum of 27 EU countries.

(22) EU-AT,...EU-SK, 22 EU entries: Estonia, Latvia and Lithuania merged into EU-BS; Cyprus, Luxemburg and Malta omitted.

(3) UK, Norway, Switzerland (CH).

United States (14)

(1) US,

(13) CA,DC,NJ,NY,WA. More US states from 2020-04-01 onwards: CA, CO, CT, DC, FL, GA, IL, LA, MA, MI, NJ, NY, WA. No data available on 2020-04-02, after which we had the New York Times data.

Rest of world (12)

(1) World from 2020-04-01.

(11) Australia (AU), Brazil, Canada, Iran, Malaysia, Philippines, South Africa, with India, Indonesia, Mexico, Turkey added from 2020-04-18.

We only report results when there are sufficient cases available. Initially this was 250 confirmed and 30 deaths, but from 2020-04-01 this was changed to 2000 and 200 respectively.

	2020-03-23	*2020-04-08*	*2020-04-17*	*2020-04-24*	*2020-04-27*
EU			04-04:13 (5.1%)		
EU-AT	03-22 (10.4%)				
EU-BS	03-22 (1.0%)				
EU-FR			04-04:13 (28%)	04-14:22 (0.8%)	
CH	03-22 (3.2%)				
AU	≤03-22 (35%)			04-18:22 (1.6%)	
Brazil	03-22 (3.0%)				
Canada					04-12:04-25 (2.0%)
Mexico					03-13:04-24 (116%)
US		03-19:04-06 (0.9%)			
World			04-04:13 (1.6%)		

Table 6: Revisions in cumulative confirmed count, Johns Hopkins data. Each column is headed by the date on which we detected the change. Each entry lists the date range affected, and the largest absolute percentage change of the revision in parentheses.

B Data revisions

Tables 6 and 7 document the data revisions in the Johns Hopkins/CSSE data set where they exceed 0.5%. This is only for the countries for which we produce any forecasts, and excluding US states. Note that we regularly have a two-day gap between data downloads, so the actual revision date may differ by one day.

For confirmed cases, if we ignore the revisions of 23 March, we see a large change for France. We detect this first on 17 April (when comparing to the data from 15 April), and note that the revision applies to ten observations (04-04 to 04-13, shortened to 04-04:13 in the table). The largest change of 28% is a downward revision by just over 30 000 cases. This is a temporary change of the upward trend: on 14 April the numbers are identical again. For Mexico there was a correction to the dates: 12 March was deleted, and the numbers shifted up, and the duplication of 23 and 24 April was also corrected.

The correction to the death count for Mexico, Table 7, is again a shifting up of the data. For the UK it is the permanent redefinition from only counting deaths in hospitals to including non-hospital deaths (mainly care homes). For the US it is a revision for most of the history, but the observations for 2020-04-25 are identical again.

In order to omit the largest revisions, we restrict assessment of accuracy to the period 2020-03-24 to 2020-04-25, and leaving France and Mexico out. To minimize the impact, the outcome is taken from the first available data set after the forecast.

Table 8 lists the revisions in the ECDC deaths data for the same countries, up to 2004-04-27. These are all corrections for short periods. The forecast evaluation of §6.2 is based on the data from the first available data set that has the outcome. None of these corrections should have affected the MCIC forecasts.

	2020-04-17	*2020-04-18*	*2020-04-27*	*2020-04-29*
EU	04-01 (1.2%)			
EU-FR	04-01 (9.2%)			
UK				03-05:04-27 (105%)
Canada			04-13:04-25 (0.9%)	
Mexico			03-20:04-24 (100%)	
US		04-07:09 (0.6%)	03-12:04-24 (44%)	
World	04-01:15 (0.7%)		03-12:04-25 (7.0%)	03-05:04-27 (1.4%)

Table 7: Revisions in cumulative deaths, Johns Hopkins data. Each column is headed by the date on which we detected the change. Each entry lists the date range affected, and the largest absolute percentage change of the revision in parentheses.

	2020-04-02	*2020-04-07*	*2020-04-08*	*2020-04-09*	*2020-04-18*	*2020-04-21*
EU-DE		04-03:04 (8.4%)				
EU-FI					03-14 (18.1%)	
EU-HU			03-31:03 (11.4%)			
EU-IT						
CH		04-03:05 (3.4%)				
Canada				03-22:28 (5.8%)		
Malaysia						04-01 (5.3%)
Mexico			04-03 (9.6%)			
Turkey	04-01 (12.3%)		03-31 (6.1%)			

Table 8: Revisions in cumulative deaths, ECDC data and dates. Each column is headed by the ECDC date on which we detected the change. Each entry lists the date range affected, and the largest absolute percentage change of the revision in parentheses.

	UK	*AT*	*BE*	*DE*	*DK*	*ES*	*IT*	*NL*	*SE*	*N*	*CH*
				MAPE 2020-04-08 : 2020-05-05							
IHME	14.2	4.8	5.8	5.5	6.9	4.3	4.0	7.4	9.7	11.5	2.9
Avg	6.8	4.0	5.4	6.7	2.3	1.6	1.2	3.8	8.5	3.2	2.2
F	6.2	3.6	5.4	6.1	3.7	0.9	0.6	2.9	15.5	1.9	2.1
				MPE 2020-04-08 : 2020-05-05							
IHME	7.0	−4.8	−5.0	0.6	6.9	−3.4	−4.0	7.4	4.1	11.5	1.6
Avg	1.4	1.3	3.4	5.1	1.0	1.1	0.7	3.0	4.1	3.2	2.0
F	0.1	−0.0	0.5	2.2	−0.0	0.4	−0.2	−0.1	0.4	1.9	1.9
Count	47	47	47	47	42	47	47	47	47	14	42

Table 9: Forecasts of deaths from 1 to 7 days ahead: mean (absolute) percentage errors relative to the outcome for 19 European countries.

Horizon	Deaths 04-08:05-05			Deaths 04-27:05-31			Confirmed 04-27:05-31		
	IHME	*Avg*	*F*	*LANL*	*Avg*	*F*	*LANL*	*Avg*	*F*
1	2.1	1.2	1.3	0.6	0.5	0.5	0.4	0.4	0.4
2	3.6	2.0	2.1	1.0	1.0	0.9	0.7	0.7	0.6
3	5.1	2.9	3.7	1.2	1.2	1.2	0.9	0.9	0.7
4	6.8	3.8	4.6	1.5	1.5	1.5	1.1	1.1	0.9
5	8.1	4.6	5.5	1.8	2.0	2.0	1.4	1.4	1.1
6	9.2	5.9	5.9	2.0	2.4	2.3	1.6	1.6	1.4
7	10.6	7.9	6.7	2.1	3.0	2.6	1.8	1.8	1.6

Table 10: Forecasts of deaths and confirmed cases for 19 European countries: mean absolute percentage errors relative to the outcome for different forecast horizons.

C Further comparisons – European countries

IHME provides forecasts for European countries from 2020-04-08 onwards. We compare the accuracy of forecasts of deaths for 19 European countries for target dates 2020-04-08 to 2020-05-05. Table 9 selects 11 countries, omitting CZ, FI, FR, HU, IE, PL, PT, and RO from the table. The top half of Table 9 gives the MAPE, the bottom half the MPE, and the number of forecasts errors J included in the measures (there are 689 errors for all European countries that the forecasters have in common). The table shows that, for most countries, our forecasts have higher accuracy and lower bias. Table 10 shows that this higher accuracy is present at each forecast horizon.

By the time LANL started forecasting European countries, the pandemic was well past the peak in most countries, which is reflected in the increased forecast accuracy. The comparison with IHME is largely for April, and with LANL for May, which are very different stages of the pandemic in Europe.

	Argentina	*Bolivia*	*Brazil*	*Chile*	*Colombia*	*Mexico*	*Panama*	*Peru*
			MAPE	2020-04-27 : 2020-06-03				
LANL	2.1	7.2	4.8	4.0	2.4	3.8	1.0	3.7
Avg	2.4	5.0	2.9	5.8	2.9	3.3	2.0	2.9
F	2.8	5.6	3.8	4.9	3.4	5.5	1.6	4.7
			MPE	2020-04-27 : 2020-06-03				
LANL	1.9	3.8	−0.6	−2.7	−0.8	−2.6	0.5	0.3
Avg	−0.4	−1.5	−0.2	−5.5	−1.3	−1.4	−1.1	−0.1
F	−0.9	−3.2	−0.0	−4.9	−1.2	−1.6	−0.4	−0.8
Count	59	17	73	59	59	73	52	59

Table 11: Forecasts of deaths from 1 to 7 days ahead: mean (absolute) percentage errors relative to the outcome for Latin American countries.

Horizon	Deaths 04-27:06-03			Confirmed 04-27:06-03		
	LANL	*Avg*	*F*	*LANL*	*Avg*	*F*
1	1.0	1.4	1.2	1.2	1.5	1.5
2	1.9	2.2	2.3	2.3	2.5	2.6
3	2.9	2.9	3.2	3.1	2.9	3.1
4	3.7	3.2	4.1	4.3	3.6	3.7
5	4.4	3.9	5.2	5.1	4.3	4.6
6	5.2	4.6	5.9	6.1	5.0	5.6
7	5.6	5.1	6.1	6.9	6.1	6.7

Table 12: Forecasts of deaths and confirmed cases for Latin American countries: mean absolute percentage errors relative to the outcome for different forecast horizons.

D Further comparisons – Latin American countries

IHME provides the first forecasts for Latin American countries (LAC) for 2020-05-11. The forecast errors compared to JH/CSSE for Chile, Mexico, and Peru are so large that it is likely that there are issues with the supplied information. It also appears that in some cases the forecasts were not updated in a subsequent release.