Short-term forecasting of the Coronavirus Pandemic - 2020-04-27

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 on-line 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

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 compares our forecasts with several more structural 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. 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, US, and EU-27 are constructed.

At some date the Johns Hopkins data dropped US states from their dataset. This gap was filled later by the New York Times, collecting data from state-level health authorities. Their US state data can be downloaded from Github at `github.com/nytimes/covid-19-data`.

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 2000/100, then 2000/200 respectively).

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 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.

---

[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.
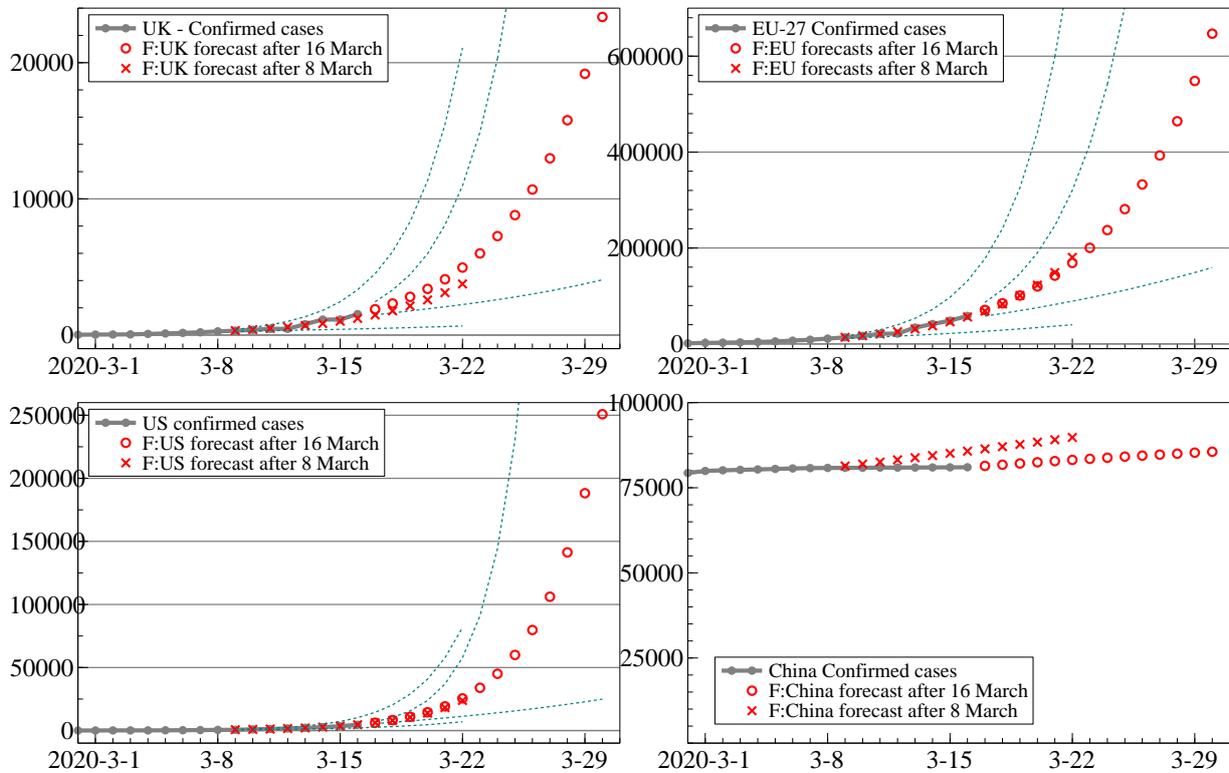
Figure 1: Forecasts of confirmed cases of COVID-19 for UK, EU-27, US, and China. Johns Hopkins/CSSE Data collected and forecast made on 2020-03-17.
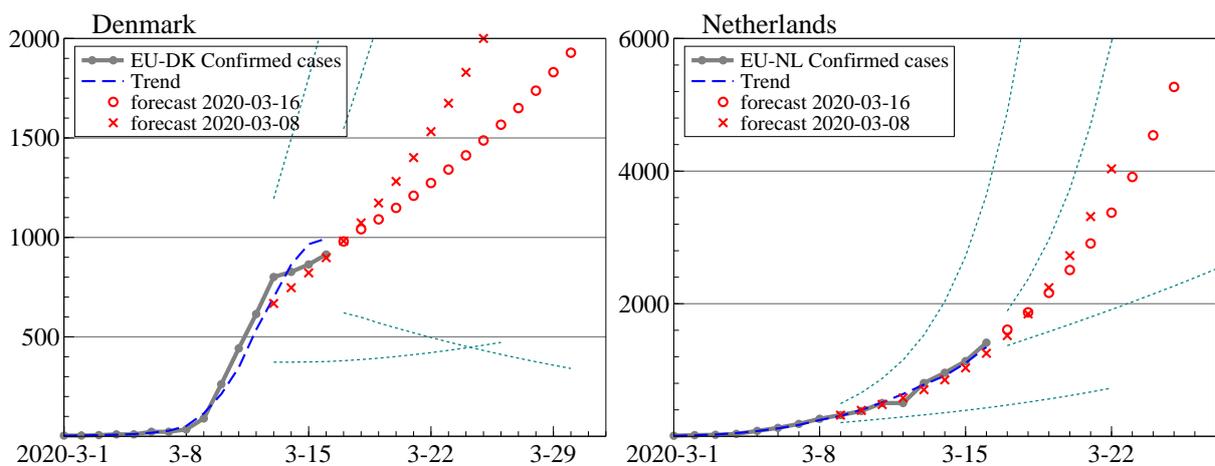
Figure 2: Forecasts of confirmed cases of COVID-19 for mainland Denmark and the Netherlands. Data from 2020-03-17.
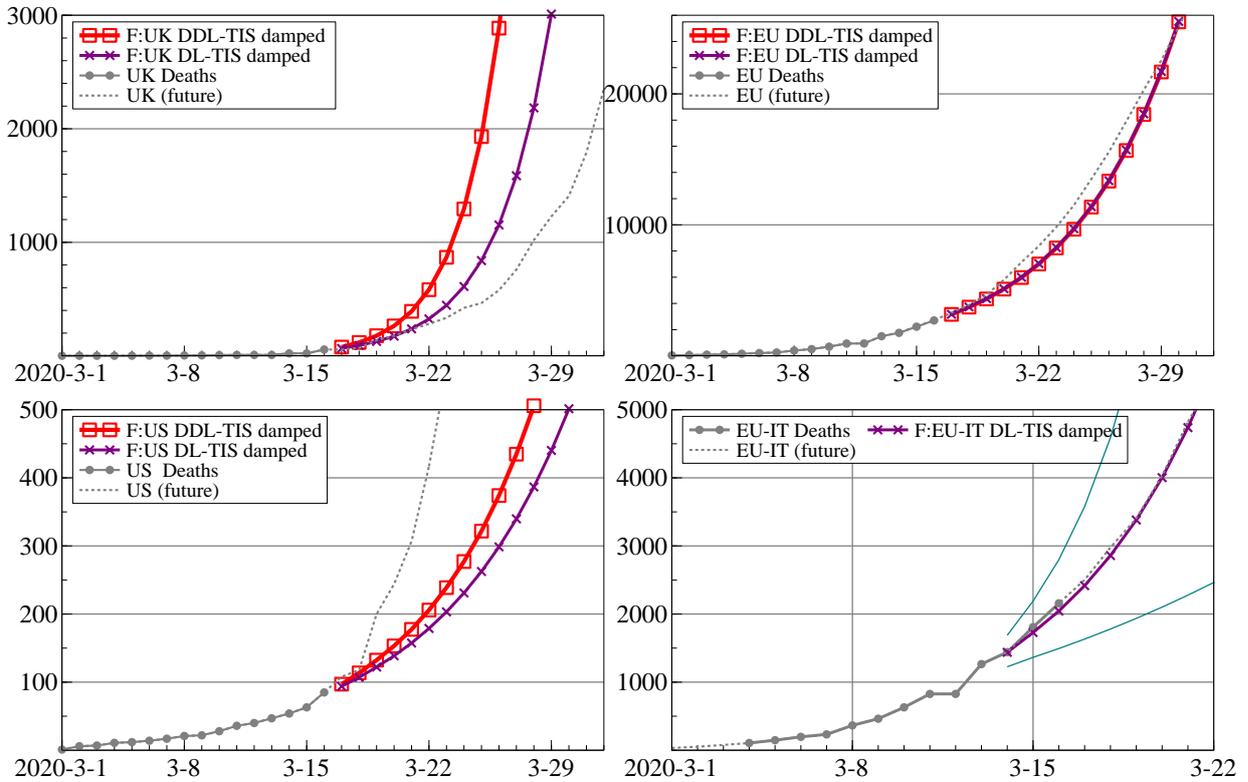
Figure 3: Forecasts of deaths from COVID-19 for UK, EU-27, US, and Italy. Data and forecasts from 2020-03-17, subsequently updated with later data (dotted line).

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:

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

The (broken) linear trend $(t - s - 1)I(t \leq s)$ 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:

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

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

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

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

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):

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

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

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

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

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:

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

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 $\widehat{\mu}_{T+h}^{(2)}(\text{H})$ be the upper bound from the damped I(2) forecasts, and $\widehat{y}_{T+h}(\text{H})$ the upper bound for the levels forecast, then we construct new upper bounds as:

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

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

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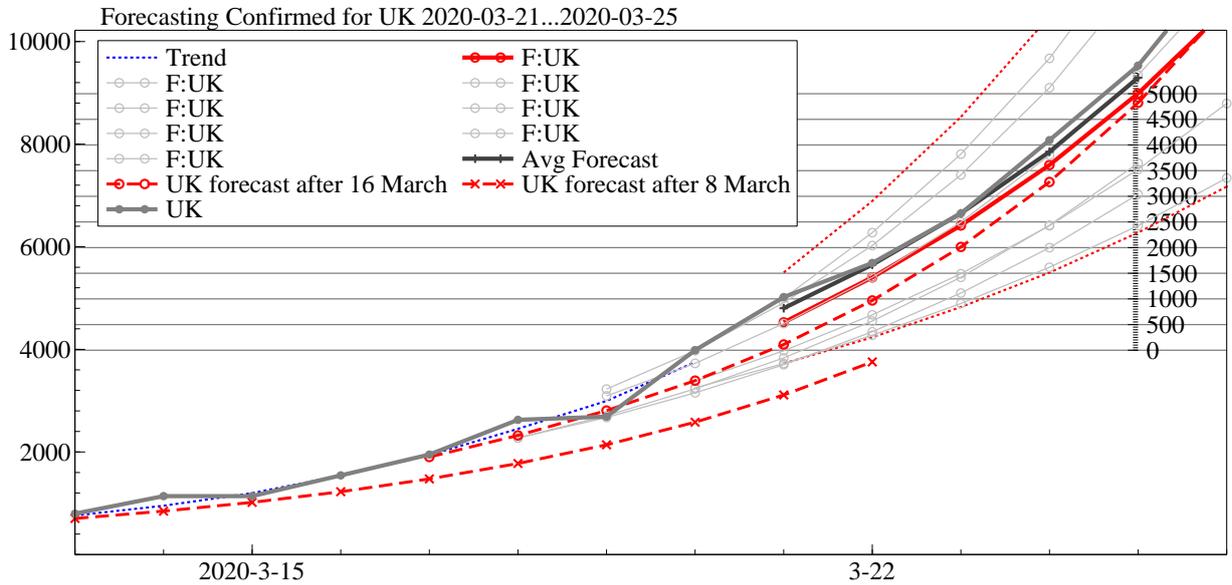
Figure 4: Forecasts of confirmed cases of COVID-19 for UK from 2020-03-17, together with forecasts from 2020-03-21.

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

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 Some forecast comparisons

6.1 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 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 $j = 1, ..., 104$ 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 observations:

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

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

Table 2 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

---

[7]Very occasionally, the end point of the estimated trend seems to be estimated unsatisfactorily.

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	1072	629	−69	26	−44	21	75	41	48	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	28	946	950	−20	−8	−6	−1	33	27	9	7

Table 2: 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.
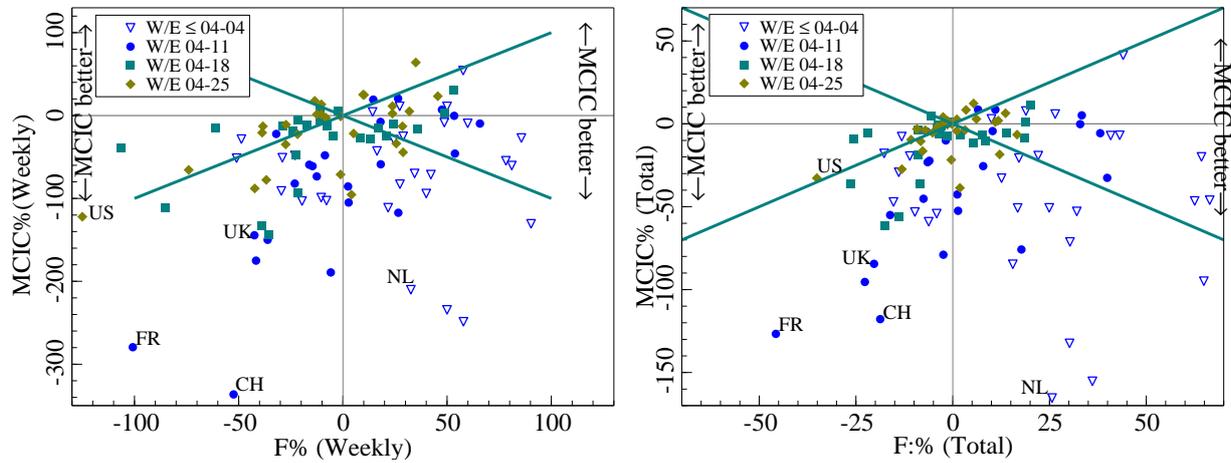
Figure 5: Weekly forecast percentage error of deaths from COVID-19 for MCIC and our forecasts. Relative to weekly count (left) and end-of-week cumulative total (right). Week endings marked by different symbols.

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 2 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 65% inside (out of the 104 considered forecasts). We report 80% intervals, with just 60% inside.

Figure 5 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.
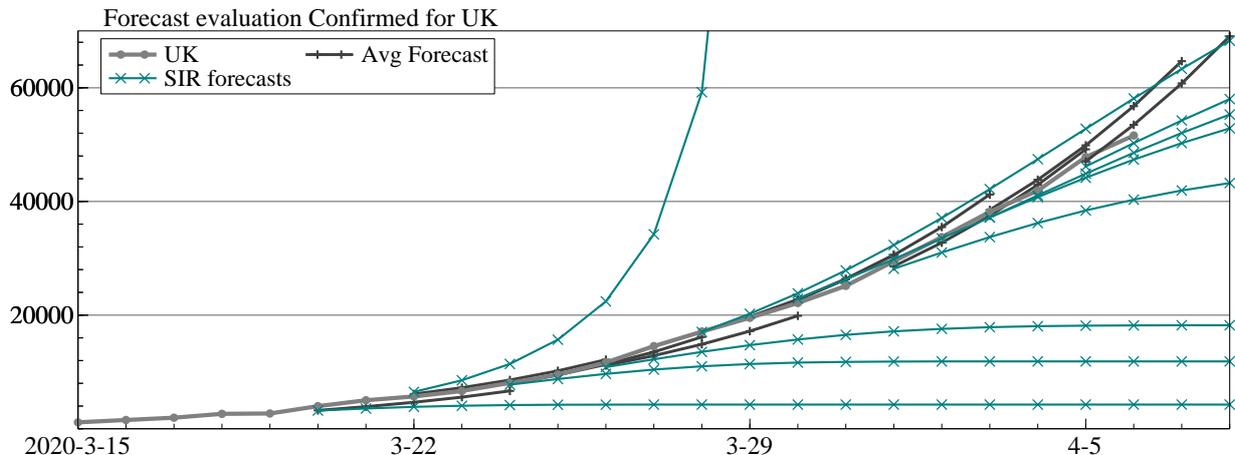
Figure 6: Forecasts of confirmed cases of COVID-19 for UK from 2020-03-20 to 2020-04-05. Average forecasts and SIR forecasts.

6.2 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):

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

These sum to zero so N is constant. Given initial conditions $R_1 = 0, I_1 = y_1$ and specific values of S_1, β, γ, the ODE can be integrated to find the time paths of S, I, R, denoted S_t, I_t, R_t. The model can be used to compute the epidemic evolution for assumed parameters, which will usually be done in a richer model.

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, 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.[8]

Figure 6 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 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.

---

[8]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 $-\log \sum \varepsilon_t^2/T$.

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*.