



Short-term forecasting of the coronavirus pandemic

Jurgen A. Doornik^{a,b,*}, Jennifer L. Castle^{c,b}, David F. Hendry^{a,b}

^a Nuffield College, Oxford, UK

^b Climate Econometrics and Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, UK

^c Magdalen College, Oxford, UK

ARTICLE INFO

Keywords:

Automatic forecasting
COVID-19
Epidemiology
Forecasting
Forecast averaging
Machine learning
Smoothing
Time series
Trend indicator saturation

ABSTRACT

We have been publishing real-time forecasts of confirmed cases and deaths from coronavirus disease 2019 (COVID-19) since mid-March 2020 (published at www.doornik.com/COVID-19). These forecasts are short-term statistical extrapolations of past and current data. They assume that the underlying trend is informative regarding short-term developments but without requiring other assumptions about how the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus is spreading, or whether preventative policies are effective. Thus, they are complementary to the forecasts obtained from epidemiological models.

The forecasts are based on extracting trends from windows of data using machine learning and then computing the forecasts by applying some constraints to the flexible extracted trend. These methods have been applied previously to various other time series data and they performed well. They have also proved effective in the COVID-19 setting where they provided better forecasts than some epidemiological models in the earlier stages of the pandemic.

© 2020 The Authors. Published by Elsevier B.V. on behalf of International Institute of Forecasters. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Given its massive impacts on lives globally, the coronavirus disease 2019 (COVID-19) pandemic is a major focus of interest at present. Having a good estimate of the number of cases and deaths in the coming days and weeks can help health authorities plan and governments to decide about whether to enforce or ease lockdowns. In this study, we outline how we produced real-time forecasts for confirmed COVID-19 cases and deaths for many parts of the world on an almost daily basis from March 20 onward. These forecasts were largely reliable indicators of what to be expect in the following week.

Models based on well-established theoretical understanding supported by available evidence are crucial for

viable policy-making in disciplines based on observational data, such as economics, other social sciences, and epidemiology. However, economics has a long history of employing relatively simple data-based devices for out-forecasting formal structural models. This phenomenon also affects other subjects. This is because shifts in the distributions of variables from their past behavior lead to systematic mis-forecasting in all models in the equilibrium-correction class. The latter class comprises most of the widely used models ranging from regressions, scalar and vector autoregressions, and cointegrated systems to volatility models such as ARCH and GARCH.

Models of pandemics such as the present coronavirus pandemics are affected by this problem of nonstationarity. Epidemiological models have a sound theoretical basis and a history of useful applications. Nevertheless, novel viruses may exhibit different behaviors to those assumed by models, and policy reactions to early predictions of mass deaths (e.g., mandatory lockdowns) can suddenly

* Correspondence to: Climate Econometrics, Nuffield College, Oxford, OX1 1 NF, UK.

E-mail address: jurgen.doornik@nuffield.ox.ac.uk (J.A. Doornik).

shift distributions in directions that can be difficult to formally model. Due to their slow starts and then exponential increases, before gradually slowing, pandemic data are highly nonstationary. Indeed, the methods used for reporting pandemic data also exhibit stochastic trends, such as due to the ramping up of testing and distributional shifts, including the sudden inclusion of care home cases. Thus, there is a compounding effect because the nonstationarity of the underlying data interacts with the nonstationarity of the reporting process. (Clements & Hendry, 1999) discussed these issues for economic time series.

Viable forecasting models must be able to handle this quadruple nonstationarity with two forms (stochastic trends and shifts) from two sources (outcomes and their measurements). In the earlier stages of an epidemic, epidemiological models can be excessively highly driven by their assumptions, which together with the assumed mathematical processes may limit their usefulness in forecasting because they are not sufficiently empirical. As a consequence, adaptive data-based models from a class that we call robust, i.e., devices that avoid systematic forecast failure after sudden distributional shifts, have an important role in short-term forecasting after distributional shifts (also see Castle, Clements, and Hendry (2015)). However, during forecasting, this adaptability must remain firmly under control to avoid excess volatility. In addition, a noticeable drop in outcomes relative to baseline extrapolations according to these models can provide an indication that policies are having a positive impact.

The remainder of this paper is organized as follows. In Section 2, we outline the general methodology used to produce the statistical forecasts and the forecast evaluation metrics. In Section 3, we discuss the COVID-19 data and highlight problems with the data measurements and errors. In Section 4, we explain how the forecasting methodology was applied to these data and the judgmental decisions involved in the forecast specification. The forecasts are introduced in Section 5. In Section 6, we present assessments of the performance of our forecasts, which we compared with the forecasts obtained from several other structural epidemiological models. In the final section, we give our conclusions.

2. Modeling and forecasting methodology

The methodology employed to construct the robust forecasts involves several steps. First, the observed daily time series is decomposed into an underlying flexible trend and a remainder term, where it is assumed that there is no seasonality. This trend is estimated by taking moving windows of the data and saturating them with segments of the linear trends. Selections are made from these linear trends with an econometric machine learning algorithm, and the selected subset estimates 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.

This approach has several similarities to that adopted by Petropoulos and Makridakis (2020) to obtain their

forecasts of global counts, i.e., estimating a smooth trend in the absence of seasonality and assuming that the trend continues, at least in the short run. By contrast, we use linear models to determine the trend (instead of exponential smoothing) and obtain the forecasts with a device that is separate from the model.

2.1. Estimating the unobserved trend

Let y_t denote the dependent variable that needs to be decomposed in an unobserved trend term μ_t , and residual or irregular ε_t . For the logarithmic model, we have:

$$\log y_t = \hat{\mu}_t + \hat{\varepsilon}_t, \quad t = 1, \dots, T, \quad (1)$$

from which we obtain: $y_t = \exp(\hat{\mu}_t) \exp(\hat{\varepsilon}_t)$. A new technique for local averaged time trend estimation (LATTE, Doornik (2019)) is used to obtain this decomposition. First, the sample $1, 2, \dots, T$ is split into overlapping windows. For example, with a window size of 10 and by moving one observation at a time, we obtain sliding spans of $\{1, \dots, 10\}, \{2, \dots, 11\}, \dots, \{T-9, \dots, T\}$. Next, for each window, a trend indicator saturation (TIS) model is estimated, where saturation is achieved by broken linear trends that are zero in the future. The first broken trend is $-3, -2, -1, 0, \dots, 0$, the second is $-4, -3, -2, -1, 0, \dots, 0$, and the last is $-T+2, \dots, -1, 0, 0$. By introducing the indicator function $I\{z\}$, which is one when z holds and zero otherwise, the s th trend term can be expressed as $(t-s-1)I\{t \leq s\}$. The initial unrestricted model for the first window of size 10 with $t = 1, \dots, 10$ has candidate regressors comprising the intercept, the full trend, and six broken trends, where the first and last two are omitted:

$$1, t, (t-4)I\{t \leq 3\}, \dots, (t-9)I\{t \leq 8\}.$$

In this example, the window has eight regressors for 10 observations, and model selection is used to obtain sparsity.

Let $x_t = \log y_t$ denote the dependent variable ($x_t = y_t$ if the model is in levels). Then, a typical window w runs from observation i_w to j_w and provides estimates for $x_{w,t}$, $t = i_w, \dots, j_w$. After selection for this window, the final model is:

$$\hat{x}_{w,t} = \hat{\alpha}_w^F + \hat{\beta}_w^F t + \sum_{s \in \mathcal{T}_w} \hat{\theta}_{w,s} (t-s-1)I\{t \leq s\}, \quad t = i_w, \dots, j_w. \quad (2)$$

The coefficients for this window, $\alpha_w, \beta_w, \theta_{w,s}$, are estimated by ordinary least squares. The superscript F indicates that terms are always in the model, so selection is conducted only over the broken trend terms, and \mathcal{T}_w denotes the selected set of broken trends. *Autometrics* (Doornik (2009)) is used for selection with the algorithm for more variables than observations.

The LATTE estimate of the unobserved trend at time t is the average over $w = 1, \dots, W_t$ windows that contain a fitted value $\hat{x}_{w,t}$ for this observation:

$$\begin{aligned} \hat{\mu}_t &= W_t^{-1} \sum_{w=1}^{W_t} \hat{x}_{w,t}, \\ \hat{\varepsilon}_t &= x_t - \hat{\mu}_t. \end{aligned}$$

The number of estimates employed for averaging, W_t , is smaller at the extremes of the full estimation sample. In contrast to other trend cycle decomposition methods, this approach can handle smooth changes as well as abrupt breaks when using only linear models. For a smooth change, the average is taken over gradually shifting trends with a smooth transition, whereas for a large abrupt change, the appropriate combination of trends is likely to be selected from the windows within which they are contained.

Model (2) is called L-TIS, where the L denotes logarithms. The following two variants allow quadratic and cubic trends, respectively:

$$\text{DL-TIS} \quad \Delta \hat{x}_{w,t} = \hat{\alpha}_w^F + \hat{\beta}_w^F t + \sum_{s \in T_w} \hat{\theta}_{w,s} (t-s-1) I\{t \leq s\}, \quad (3)$$

$$\text{DDL-TIS} \quad \Delta \Delta \hat{x}_{w,t} = \hat{\alpha}_w^F + \hat{\beta}_w^F t + \sum_{s \in T_w} \hat{\theta}_{w,s} (t-s-1) I\{t \leq s\}. \quad (4)$$

DL-TIS introduces cumulated differences into the unobserved trend, thereby making $\hat{\mu}_t$ an $I(1)$ variable¹ with a quadratic trend, whereas DDL-TIS makes $\hat{\mu}_t$ $I(2)$ 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 L-SIS model is given by:

$$\hat{x}_{w,t} = \hat{\alpha}_w^F + \hat{\beta}_w^F t + \sum_{s \in T_w} \hat{\theta}_{w,s} I\{t \leq s\}, \quad t = i_w, \dots, j_w. \quad (5)$$

This model omits the first and last steps from the initial candidate set, whereas TIS omits the first two and last two broken trends. The models corresponding to (3) and (4) are DDL-SIS and DL-SIS, respectively (see Castle, Doornik, Hendry, and Pretis (2015, 2020) for SIS and TIS, respectively, and Walker, Pretis, Powell-Smith, and Goldacre (2019) for a medical application of TIS).

2.2. Forecasting

Forecasts are not based on extrapolating the linear trends in the LATTE method because these are only estimated over short samples, thereby leading to excessively volatile forecasts, which will be amplified when the model is estimated based on first or second differences. Instead, we use an improved version of the calibrated average of rho and delta methods (Card; Doornik, Castle, and Hendry (2020)) by adding a third method in the average, thereby resulting in Cardt, as described by Castle, Doornik, and Hendry (2019).

Cardt takes the average of two autoregressive models and one moving average model, before calibration where the forecasts are treated as pseudo-observed values. Cardt allows for up to $I(1)$ data with a linear trend, where it makes automatic decisions about whether to use differencing. This method performs very well with the data from the M4 and M3 forecast competitions (Makridakis, Spiliotis, & Assimakopoulos, 2020).

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

Epidemiological data exhibit exponential growth in the early stage, so we consider the following possible extensions beyond the default Cardt forecasts (6):

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

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

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

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

Procedure (6) gives the standard Cardt forecasts targeted at economic applications. Forecast (7) applies Cardt to the differenced trend, which is then reintegrated. Next, (8) is the simple average of the previous two, which may seem ad hoc but this leads to an effective forecasting device. (7) is not used on its own because the forecasts are excessively strongly trending. Forecast (9) is based on a shorter sample than (8).

Cardt forecasts are also made for the residual term $\hat{\varepsilon}_t$. This unusual aspect is motivated by the observation that the residual may still contain some short-term dynamics that are not captured in the slowly moving estimated trend. This approach led to improved forecasts in our experiments. The final forecast based on the damped $I(2)$ version for the model in logarithms is:

$$\hat{y}_{T+h} = \exp(\hat{\mu}_{T+h}^{(2)}) \exp(\hat{\varepsilon}_{T+h}). \quad (10)$$

3. Data sources

We used the data repository for the 2019 Novel Coronavirus Visual Dashboard operated by the Johns Hopkins University Center for Systems Science and Engineering (JH/CSSE), which is currently updated daily and available at: github.com/CSSEGISandData/COVID-19. These data comprise confirmed cases and deaths. Recovered cases are also included but not throughout the period. The coverage comprises all countries, Chinese provinces (and similar administrative areas), US states, and some cruise ships (which we ignored).

A data set for modeling was created from this data repository with minor adjustments. First, observations were placed in columns with ISO date labels (yyyy-mm-dd) and some countries were renamed to make them closer to their ISO names. Next, for France, Denmark, United Kingdom (UK), and Netherlands, we only included the mainland tallies. Aggregates were constructed for China and EU-27.

At one point during our forecasting process, JH/CSSE stopped reporting data for US states. This gap was subsequently filled by the New York Times, which collected data from state-level health authorities. These US state data can be downloaded from Github at: github.com/nytimes/covid-19-data. Later, the New York Times data became inconsistent due to partially implemented revisions but JH/CSSE had reinstated the US state data by this point, so we reverted to this data source.

All forecasting was related to the cumulative “confirmed” and “deaths” counts separately. The regions and

countries for which we published forecasts changed over time according to our interests but subject to a minimum amount of 250 confirmed cases or 30 deaths (later increased to 2000 and 200, respectively) (see [Appendix A](#); [Appendix B](#) considers the data revisions).

4. Adaptation of the methodology to COVID-19

The target variables comprised cumulative daily counts of confirmed COVID-19 cases and deaths for countries around the world, which grew exponentially in the initial epidemic phase. To ensure effective forecasting, we needed to step down not just to the daily increments but to the change in the daily increments, as reflected by the differencing in models (3) and (4).

Our primary aim was to provide short-term forecasts that might aid policy makers, where these forecasts could serve as a useful guide to what might happen in the week ahead. For example, the report on 2020-03-17 that Italian deaths increased by 16% in a day was largely in line with our forecast of 18% and it should not have been as surprising as it was perceived at the time. Forecasts were restricted to seven days ahead, which is slightly shorter than the 10 days argued for by [Petropoulos and Makridakis \(2020\)](#).

Models estimation was affected by several data challenges. First, policy interventions aimed to suppress the transmission of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) that causes COVID-19. In the UK, an alarming but realistic scenario ([Ferguson et al., 2020](#)) led to a switch from mitigation to lockdown and isolation. Furthermore, countries employed different testing strategies and technologies, and these improved over time. Some countries included asymptomatic cases, whereas others did not. Deaths were sometimes recorded a few days late and it was not always clear whether COVID-19 was the cause. Deaths were also counted differently, e.g., the UK initially only counted deaths in hospitals, whereas some other countries included those in care homes. Thus, the counts were subject to structural breaks, underreporting, definitional changes, delays, and errors. Nonetheless, they were the focus of the media and government briefings, and the target of our forecasts.

Plots of the data indicated occasional sudden jumps and even negative counts. The latter should not be possible but several examples were caused by a redefinition, which was not (or could not be) applied backward. Models need to be robust to this problem, so it was not necessary to guess the actual observations. This issue motivated the LATTE approach discussed in Section 2. However, a few adaptations were needed for the COVID-19 data.

The first adaptation was related to the use of logarithms. Experiments conducted with and without logarithms suggested that the model in logs obtained slightly better forecasts. Counts were zero before the SARS-CoV-2 virus emerged so to allow the use of logs, we replaced (1) with:

$$x_t \equiv \log(y_t + 1) = \hat{\mu}_t + \hat{\varepsilon}_t, \quad (11)$$

$$y_t \approx [\exp(\hat{\mu}_t) - 1] \exp(\hat{\varepsilon}_t). \quad (12)$$

Eq. (12) does not exactly follow from the exponentiation of (11), but the difference was very small after the disease emerged and the numbers became substantial. Forecast (10) was adjusted accordingly.

Next, we found that the interval forecasts based on the damped I(2) methods (8) and (9) were excessively wide by quite a large margin (examples are shown in [Fig. 2](#)). To address this problem, we used the intervals obtained by applying Cardt to the levels rather than the logarithms. Cardt gives upper and lower confidence bounds of the central forecasts for an applied confidence α . This worked very well in the M4 settings with an average of about $100(1 - 2\alpha)$ of the outcomes inside the confidence interval. In our forecast graphs, we employed $\alpha = 0.1$ with the aim of 80% forecast confidence intervals. If we let $\hat{\mu}_{T+h}^{(2)}(H)$ be the upper bound from the damped I(2) forecasts, and $\hat{y}_{T+h}(H)$ the upper bound for the levels forecast, then we constructed the new upper bounds as:

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

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

Third, the forecasts of cumulative counts should not go down and not decrease below the last observation (although that could be revised downward). We did not impose this constraint initially, but we later adjusted the forecasts and intervals so a negative increment was replaced by the previously highest value.

Finally, there was a choice of LATTE models that could be used. It was not considered practical to find the best formulation for each country separately when forecasts were made every day, and the set of countries of interest could change. Instead, we provided two forecasts. The first denoted as F: was based on DDL-TIS and the second was based on an average of several forecasts, as follows.

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

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

[Fig. 1](#) illustrates how the procedure worked for the UK. Using data up to 2020-03-21, the left panel shows the eight individual forecasts (unshifted), with the first in bold red (DDL-TIS with damped I(2), labeled with the prefix “F:”), and the remainder in light grey (F:2, ..., F:8). The 80% forecast interval for F: is also given. The bold black line is the average of the eight forecasts adapted to the last known observation. The average is close to the red line in this case. The trend line in the graph is the estimated trend from the DDL-TIS model.

The right panel in [Fig. 1](#) shows the updated outcomes (solid line with dots) on an extended time scale. The forecasts tracked these very closely. Furthermore, two older forecasts are shown in the graph, which indicate the improved curvature of the forecasts over time.

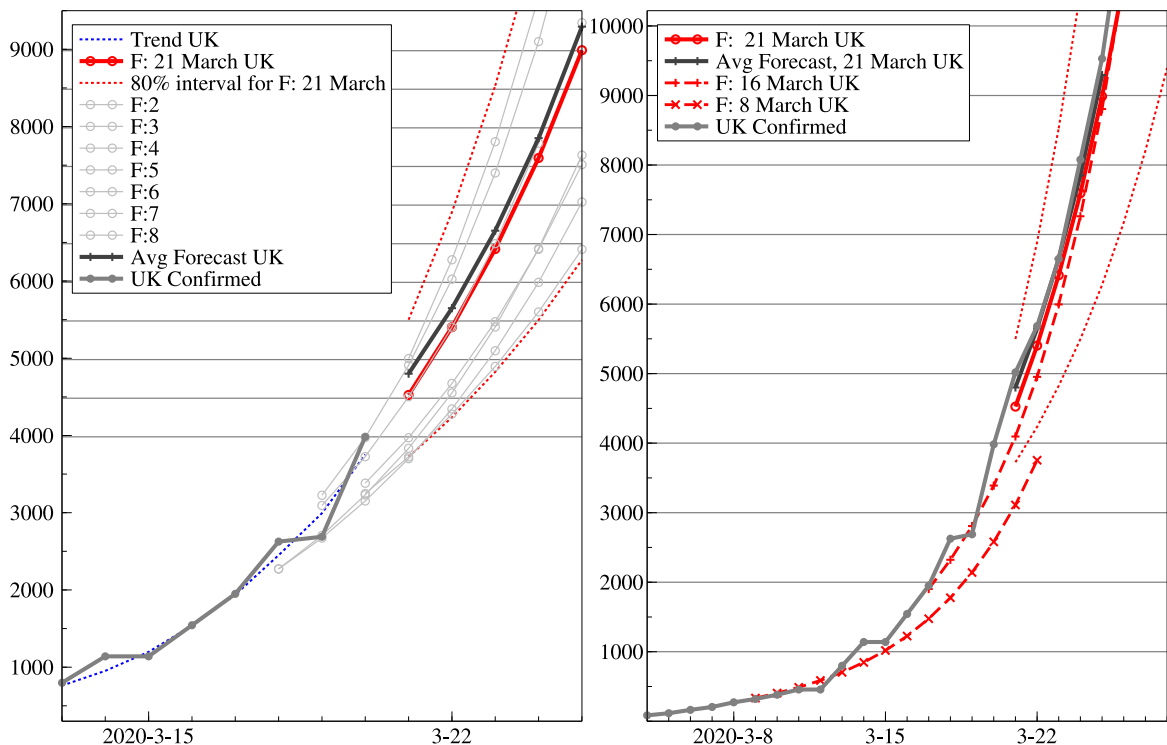


Fig. 1. Forecasts of confirmed cases of COVID-19 in UK from 2020-03-21 (left panel), together with forecasts from 2020-03-16 and 2020-03-08 (right panel). The seven thin grey lines and the red forecast (marked with a circle and prefixed with F:) were used together for the average forecasts (solid black line with + symbols). The thick line labeled UK Confirmed represents the observed counts.

The first forecast path (the solid red line labeled as “F” in Fig. 1 and the graphs on our website) was usually preferred. However, when the average forecast (the solid black line labeled Avg Forecast) deviated considerably, this could indicate a sudden recent acceleration or deceleration, and it might be difficult to decide between the two in that case.

When the inflection point (i.e., the “peak” defined in Appendix C) was reached, the methods change to the following.

F: DDLX-SIS with short damped $I(2)$, where X indicates that the raw data were extended by default Cardt forecasts prior to trend estimation. When the trend was established, these initial forecasts were ignored when making the final forecasts.

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

5. Initial forecast experiments

Initially, we consider the results obtained when we commenced COVID-19 forecasting. Our first tentative forecasts used data up to March 16, 2020, where data collection started the day before. Our first forecasts went live on the website www.doornik.com/COVID-19 on

March 20, 2020 using data up to 2020-03-18 and forecasting for 2020-03-19 to 2020-03-23 (only five days ahead initially but we later changed to seven). Subsequently, we usually updated the forecasts about four times a week, but occasionally more often.

One advantage of presenting real-time forecasts is that they cannot be biased by knowing what has happened. Obviously, for such a major event, a large amount of information was communicated every day. However, we considered it acceptable to make small adjustments to our procedures as we learned and had more time for the implementation.² Occasionally, we needed to correct minor errors in our coding.³

Several different specifications of the model and forecasting approach were considered initially after some trials, as shown in Table 1. A multivariate DDL-TIS model was estimated in the cases where countries are listed

² These procedures were then used from that point onward. A month later, many improvements were made in terms of the presentation of the results, but only minor changes to the forecast procedure. We introduced an additional and different method to complement forecasting when the daily increments stopped rising, which was based on path indicator saturation with scenarios from the Chinese experience. This method was documented in a companion study (Castle, Doornik, & Hendry, 2020).

³ All results were coded in Ox (Doornik (2018)) using OxMetrics 8.2 (Doornik and Hendry (2018)). The HTML pages were also generated using Ox code. Data download was automated using the (prerelease) DataFetch package.

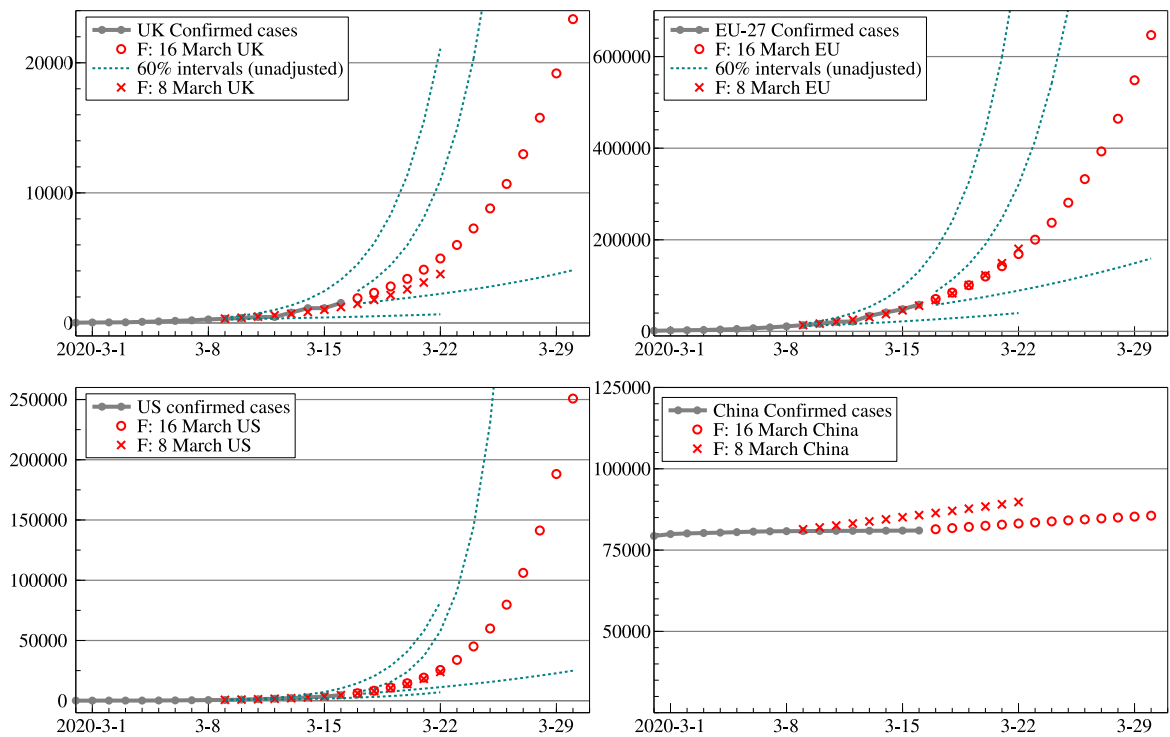


Fig. 2. Forecasts of confirmed cases of COVID-19 for UK, EU-27, USA, and China. JH/CSSE data collected and forecast on 2020-03-17. Model specifications are shown in Table 1.

Table 1
Model and forecast specifications used for the results reported in Section 5.

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		damped I(2)
Deaths	IT	DL-TIS		damped I(2)

together, but each forecast was made separately. In the cases were two methods are given in the table, the forecasts comprised the equally weighted average of both methods. The X in DDLX-TIS indicates that the data were extended by default Cardt forecasts prior to trend extraction.

5.1. Confirmed cases

Fig. 2 shows the first round of forecasts of confirmed cases for the UK, EU-27, USA, and China using the specifications in Table 1. Each graph shows the observed value as a grey line marked with dots. We started with two weeks of daily forecasts from March 17 onward, which are denoted by the red circles. The thin lines are the 60% forecast intervals obtained by Cardt without further

adjustments. As noted above, these intervals appeared excessively large and they were subsequently changed. Retrospective forecasts were made from March 9 onward using data up to March 8, which are denoted by the red crosses in the graphs. This initial attempt at forecasting appeared encouraging on 2020-03-17 and it was confirmed by later data.

According to Fig. 2, the UK and EU seemed to be on similar trajectories by mid-March. The UK had 1543 confirmed cases on 2020-03-16, and we forecast 10 000 around March 25/26. The subsequent outcomes comprised 9529 cases on March 25 and 11658 one day later. The EU had almost 60 000 cases at the end of the sample, and it was predicted to reach 250 000 around March 25. The outcome for the EU was 252 770 on 2020-03-26. The USA was on a more rapid growth path, but even these forecasts were remarkably accurate, where we forecast 79 700 for 2020-03-26 when the actual amount was 83 836 by starting from 4632 as the last observed value at the time of forecasting.

In each case, the older forecasts were already effective. The difference was small when they overlapped, except in the UK where they started to fall below the later forecasts. Both forecasts were plausible when they were made but the higher forecast was subsequently found to be the more accurate path. China was included for contrast because the epidemic had largely run its course (assuming that there is no “second wave”). Our methods clearly over-forecast for China, where some trend remained when it was no longer needed in order to

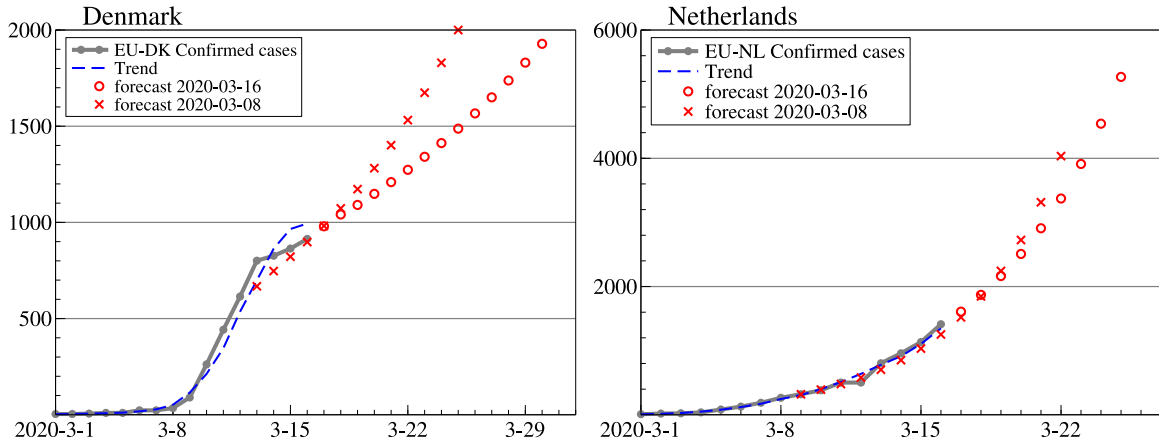


Fig. 3. Forecasts of confirmed cases of COVID-19 for mainland Denmark and the Netherlands. Data from 2020-03-17; model specifications in Table 1.

capture the early exponential growth. This problem was addressed later by switching to the model with step saturation after the peak was reached, as described in Section 4.

Fig. 3 shows the forecasts of confirmed cases for Denmark on the left and the Netherlands on the right (mainland only for both countries). The surface areas of Denmark and the Netherlands are almost identical, but the population of the Netherlands is three times higher. Thus, we set the vertical scale of the graph for the Netherlands to three times that for Denmark. The trajectories were quite different and a trend break occurred in Denmark. The Netherlands had one outlying observation on March 12.

5.2. Results for death counts

Fig. 4 shows the preliminary short-term forecasts for the death counts. Separate graphs are shown for the UK, EU-27, USA, and Italy, where forecasting started from the last observation at 2020-03-16, except the start was two days earlier for Italy.

The two different models used were DDL-TIS and DL-TIS. When these forecasts were made on March 17, the mortality observations were still limited, particularly for the US and UK, which made forecasting more difficult. The graphs also show the “future” as the dotted line. Thus, we subsequently confirmed that the forecasts were extremely accurate for Italy, adequate for the EU, excessively high for the UK, and excessively low for the USA.

5.3. Published forecasts

After the initial investigation reported in Figs. 2 and 4, we stopped using different specifications for each country. Instead, all of the published forecasts were made using procedures “F” and “Avg” described in Section 4 with the amended interval forecasts.

From 2020-04-04 onward, we added peak detection and switched to the DDLX-SIS model after the peak. The range of countries included was expanded over time, as documented in Appendix A.

6. Assessment of forecast performance

A large number of forecasts were produced and it was useful to consider their performance in some detail. The results presented in the following show that the accuracy varied considerably over time and among countries.

The evaluation was based on the mean absolute percentage forecast error (MAPE). A forecast $\hat{y}_{j,T+h}$ at horizon h from group $j = 1, \dots, J$ has the forecast error $u_{j,T+h} = y_{j,T+h} - \hat{y}_{j,T+h}$, and:

$$\text{MAPE} = \frac{100}{J} \sum_{j=1}^J \frac{|u_{j,T+h}|}{y_{j,T+h}}. \quad (13)$$

The mean percentage error (MPE) omits the absolute value from (13) (note that $y_{j,T+h} > 0$ was always the cases in our application). The mean absolute error (MAE) is $(100/J) \sum_j |u_{j,T+h}|$. For one week-ahead forecasts, we also reported a differently scaled MAPE:

$$\text{MAPE(W)} = \frac{100}{J} \sum_{j=1}^J \frac{|u_{j,T+7}|}{y_{j,T+7} - y_{j,T}}. \quad (14)$$

6.1. Forecast accuracy from 2020-03-24 to 2020-04-25

The forecast accuracy was measured as the error in the percentage of the outcome computed for almost all of the forecasts that we published. Following Appendix B, we restricted this first evaluation to the period from 2020-03-24 to 2020-04-25, and omitted France and Mexico.

Table 2 shows the average MAPE values for a selection of geographical regions and all combined. The Count column reports the number of one step-ahead percentage errors used to compute the mean (one fewer for two steps, etc.). The accuracy for deaths was about half of that confirmed, except for Spain, Iran, and the whole EU, where the reverse was the case. Overall, our forecasts exhibited good accuracy, thereby providing useful insights into the developments that could be expected in the coming days at a time of rapid change.

The coverage represents the percentage of outcomes within the quoted forecast interval. We aimed for 80%

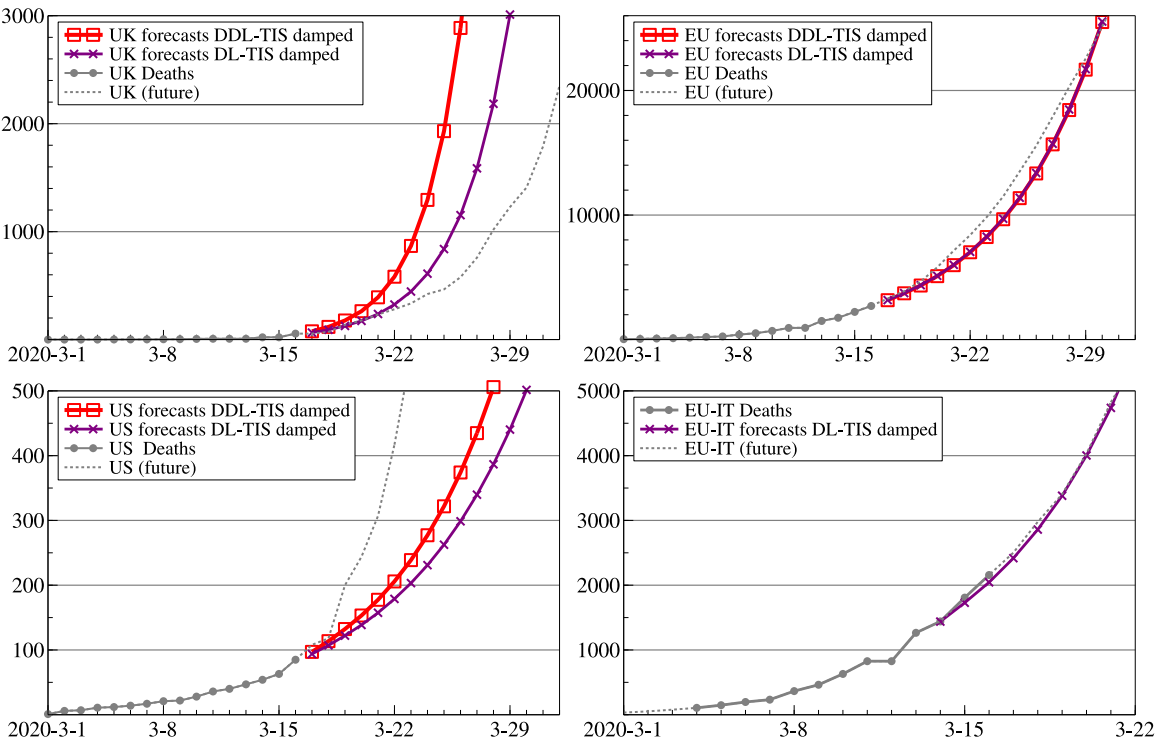


Fig. 4. Forecasts of deaths from COVID-19 for UK, EU-27, US, and Italy. Data and forecasts from 2020-03-17 and subsequently updated with later data (dotted line). Model specifications are shown in Table 1.

Table 2
Forecast accuracy for different geographical areas. MAPE over 2020-03-24 to 2020-04-25 for each area for one, two, and four step-ahead forecasts.

Area	Confirmed cases							Deaths						
	1 step			2 step		4 step		1 step			2 step		4 step	
	Count	Avg	F	Avg	F	Avg	F	Count	Avg	F	Avg	F	Avg	F
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	19	0.4	0.4	1.1	0.9	3.1	2.1
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	4.1	4.9	8.4	8.8	14.6	14.0
All	692	1.7	1.9	3.2	3.2	6.6	5.9	510	3.2	3.9	5.7	6.9	9.6	11.0
Coverage %	692	99	95	96	90	84	79	510	96	86	89	77	78	65

forecast confidence intervals, so they were close to the target for four step-ahead forecasts. They were excessively wide for one step ahead and excessively narrow beyond four steps. The forecast intervals for the average were the averages of the individual forecast upper and lower bounds adapted to the last observation, and these are not reported in the graphs.

Fig. 5 plots the MAPE values for the one, two, and four step-ahead forecasts for confirmed cases (top) and deaths (bottom). The accuracy is given for the forecast F and the average forecast (Avg). All of the lines show that the accuracy increased over time, which was more attributable to the change of the pandemic from exponential growth to a slow down than to the accumulation of information. When the reproduction number was low,

the counts changed more slowly and they were easier to forecast relative to the accumulated totals.

For confirmed cases, the one step-ahead error ranged from around 3% at the start to below 1% at the end. As expected, the accuracy decreased as the forecast horizon increased. For one and two steps ahead, the average forecast and F were close, but F mostly dominated for four steps. The reverse was true for deaths, where the average forecast was mostly better. The increasing death threshold probably determined the inclusion of countries (see Appendix A) that contributed to the improvement from early April. Nonetheless, the forecasts of deaths were less accurate than those of the confirmed cases, and the former data were noisier.

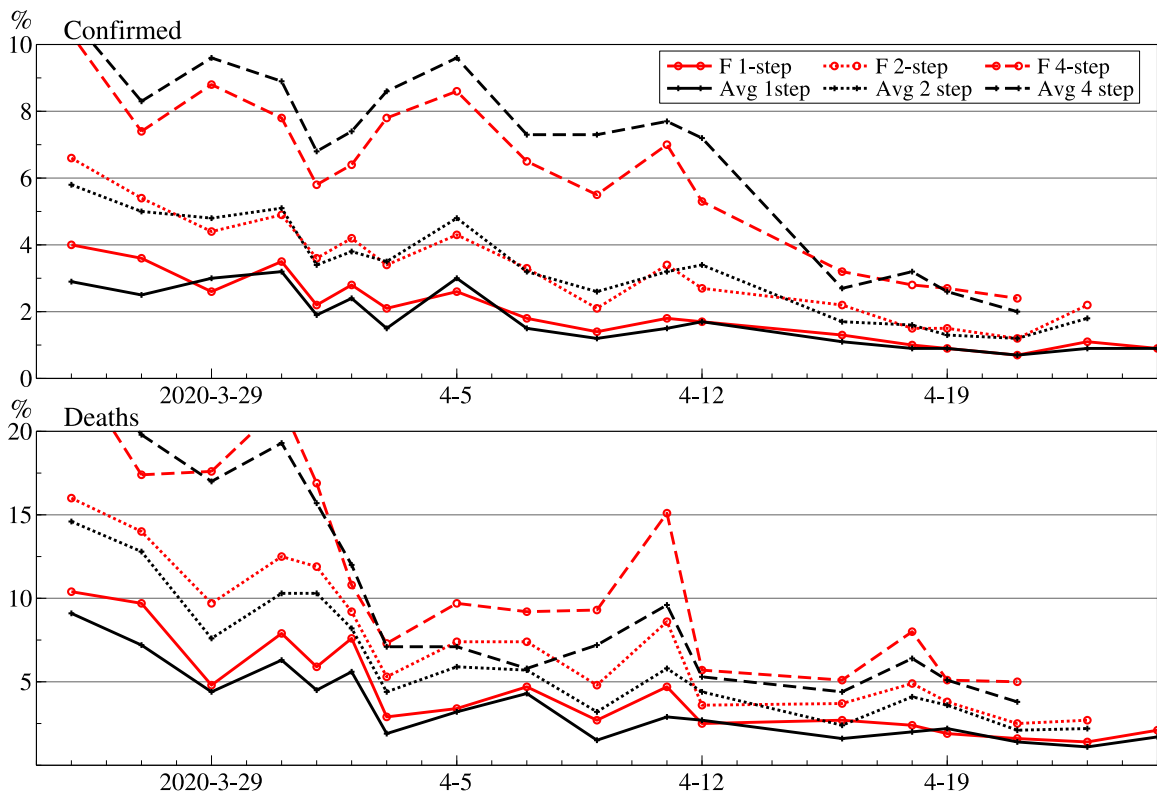


Fig. 5. Forecast accuracy over time. MAPE for each target date for one, two, and four step-ahead forecasts. Confirmed cases are shown at the top and deaths at the bottom.

6.2. Comparison with medical research council centre for global infectious disease analysis epidemiological models

The Medical Research Council Centre for Global Infectious Disease Analysis at Imperial College London (abbreviated as MCIC in the following) started publishing weekly forecasts of deaths from April 8 for a selection of countries (Bhatia et al., 2020). Their forecasts comprised a weighted average of three or four Bayesian epidemiological models.

Using our notation, y_t is the cumulative reported death count and T is the last available observation. The MCIC reported forecasts of $\hat{w}_{T+7} = \Delta_7 \hat{y}_{T+7} = \hat{y}_{T+7} - y_T$, where the observed value a week later is $y_{T+7} - y_T$, but only once each week. We forecast $\hat{y}_{T+1}, \dots, \hat{y}_{T+7}$ roughly every other day, which could be mapped to the weekly forecasts $\hat{y}_{T+7} - y_T$, and compared directly with the MCIC forecasts. The MCIC used European Centre for Disease Prevention and Control (ECDC) data, whereas we used JH/CSSE data. Due to their construction in different time zones, the data sets differed by one day, e.g., ECDC reported a weekly increase in UK deaths of 3294 on April 5, but JH/CSSE had this same number for April 4. There were occasional differences in the reported data, so we compared each forecast with its own data set. The evaluations were aligned by date, but we only refer to the JH/CSSE dates.

The MCIC report from 2020-04-15 (UK date) presented forecasts for the week starting 2020-04-11 and ending

2020-04-18 (JH/CSSE dates; April 12–19 in ECDC dates). The MCIC report dated 2020-04-08 presented forecasts for the week before, but it also contained forecasts for several previous weeks where the outcome was already known at the time of publication.

The forecast comparison was restricted to the countries and periods for which we both produced forecasts, thereby resulting in weekly forecast errors of $u_{j,T+7} = \hat{w}_{j,T+7} - \Delta_7 y_{j,T+7} = \hat{y}_{j,T+7} - y_{j,T+7}$. The evaluation assessed the MAPE as well as the differently scaled MAPE(W) and MPE(W) (see (14)). Table 3 shows the error measures for the two forecasts in the weeks ending up to April 4 combined and for each subsequent week until the end of May. The first column after the counts shows the MAE values, which were influenced by several large errors in April. The next two columns in Table 3 show the mean percentage error relative to the end of week totals. The results indicated a negative bias for MCIC throughout, but it decreased as the pandemic progressed. Our bias was positive initially, but it then became close to zero. The final four columns show the two MAPE measures (see (13),(14)). Initially, the MCIC errors were about two to three times higher. There was not much difference in May when the daily counts started to decrease in most of the countries included.

The largest MCIC error in the early weekly death counts was determined for France in the week ending 2020-04-11 at -17500 , and thus it was over-forecast compared were our error of -6300 . The second largest error was

Table 3

Mean absolute errors (MAE) and mean absolute percentage errors (MAPE) for MCIC and our forecasts *F*. (*W*) denotes a percentage of the weekly total; otherwise, the results are expressed relative to the cumulative total at the end of the week. *Count* lists the number of countries involved, which typically comprised Brazil, Colombia, Mexico, Peru, Canada, India, Indonesia, Iran, Philippines, Turkey, and the remaining European countries.

Week ending	Count	MAE			MPE			MAPE(W)			MAPE		
		MCIC	Avg	F	MCIC	Avg	F	MCIC	Avg	F	MCIC	Avg	F
Up to 2020-04-04	29	1068	790	629	−43	16	21	71	41	41	47	26	27
2020-04-11	23	1912	780	678	−40	−7	3	91	38	32	42	18	16
2020-04-18	24	372	449	226	−12	−7	−1	35	31	29	13	11	10
2020-04-25	24	1101	1209	1108	−7	−3	−0	32	33	31	9	8	8
2020-05-02	27	388	388	372	−10	−7	−6	65	58	43	11	9	7
2020-05-09	28	161	194	166	−1	−1	0	22	31	26	3	3	3
2020-05-16	28	339	240	246	−3	−2	−2	33	33	25	3	3	3
2020-05-23	28	124	183	184	−1	−0	−0	42	49	44	3	4	4
2020-05-30	27	221	296	340	−0	−0	−0	26	26	28	2	2	2

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

In both cases, the interval forecasts appeared to be excessively narrow. MCIC reported 95% intervals but only 64% were inside (out of the 238 forecasts considered). We reported 80% intervals with 74% inside (for the same 238 dates and countries).

6.3. Comparison for USA states

Two sources of USA forecasts were available over a sufficiently long period to allow a comparison with our forecasts. The first was provided by the Institute for Health Metrics and Evaluation (IHME, healthdata.org/covid) starting 2020-03-25 and the second provided by Los Alamos National Laboratory (LANL, covid-19.bsvgateway.org, from 2020-04-05 onward).

Evaluations were conducted against the outcomes reported by each institute as soon after the forecast date as possible to reduce the impact of data revisions, and this approach was preferred to using the most recent realization.

Our choice of start dates did not align well, so we filled in several gaps in our forecast history using the same procedure that we employed at the time. We followed the approach described in the previous section where the comparisons were bilateral, and we only included those cases with target dates and areas that matched for both forecast sources. We restricted the results to states with a sufficiently large count (see [Appendix A](#)), which were also the most relevant, and stopped before we switched back to JH/CSSE data. Only forecasts of up to seven days ahead were included, in order to conform to our focus on short-term forecasting. For IHME, we used forecasts in seven reports from 2020-04-05 to 04-27 and a subsequent one of the actual values. For LANL, we also used seven reports from 2020-04-05 to 04-26 and two subsequent reports of the actual values.

[Table 4](#) show the MAPE values for the relative forecast errors from one to seven days ahead for the cumulative deaths in 12 states. By the end of April, these 12 states accounted for just over 70% of the confirmed cases and deaths in the USA. The first part of the table is based on

Table 4

MAPE values for forecasts of deaths from one to seven days ahead for USA states. Each average is based on 49 periods/horizons, but 35 for Colorado. The period ranged from 2020-04-05 to 2020-04-27 for IHME and from 2020-04-05 to 2020-04-26 for LANL. The lowest values are shown in *italics*.

	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
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

Table 5

MAPE values for forecasts of deaths and confirmed cases in 12 USA states at different forecast horizons. There were 82 forecast errors at each horizon for deaths and 72 for confirmed cases. The period ranged from 2020-04-05 to 2020-04-27 for IHME and from 2020-04-05 to 2020-04-26 for LANL. The lowest values are shown in *italics*.

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

forecasts that we shared in common with IHME, and the second part is based on those shared in common with LANL. The number of forecasts included was the same for both comparisons but the sets were different, so the IHME and LANL forecasts could not be compared directly with each other.

[Table 5](#) splits the results by forecast horizon from one to seven days ahead. The same set is included for deaths as that shown in [Table 4](#). There were 82 forecasts errors at each horizon but again the sets differed in the IHME and LANL parts of the table. Both our F and Avg forecasts dominated at each horizon for deaths, where the average was the best in terms of MAPE, as also shown in [Table 2](#). The differences were not great for confirmed cases in the early horizons, but LANL had smaller forecast errors for horizons longer than four days.

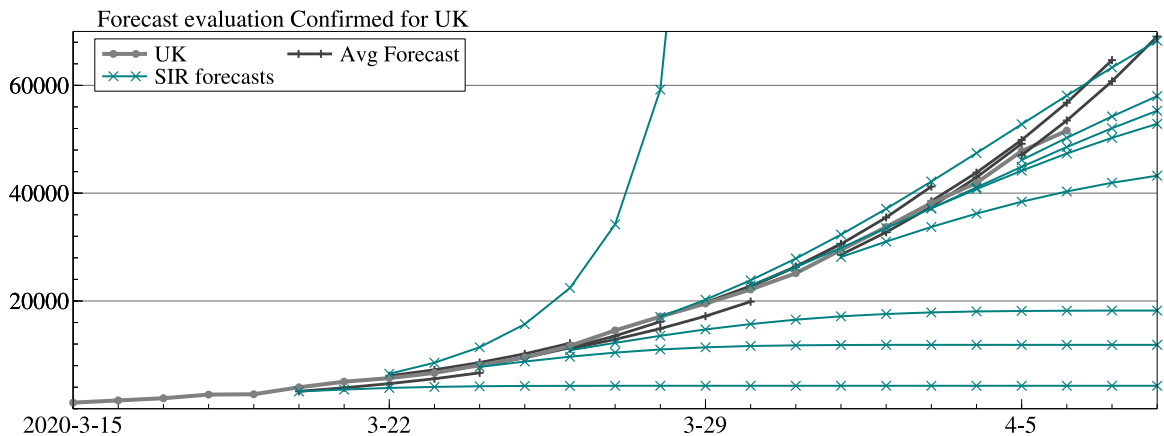


Fig. 6. Forecasts of confirmed cases of COVID-19 in UK from 2020-03-20 to 2020-04-05. Average forecasts and SIR forecasts using nonlinear least squares.

These results and those presented in the previous section suggest that the epidemiological models were less effective prior to the peak and that a few new observations could result in a very different forecast trajectory. The theoretical model could be established only later in the pandemic. Our procedures are more data driven where the aim is robustness, so they are less sensitive, but at the expense of theoretical insights. In the next section, we illustrate this effect with a simple SIR model.

6.4. Comparison with simple SIR forecasts for the UK

The classic epidemic model is the SIR model with three compartments comprising S for the number that are susceptible, I for infectious, and R for removed (either recovered or died). The reader may refer to the study by 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 for an infected individual, where $N = S + I + R$ is the total population. Then, $\beta I/N$ is the average number of contacts for a susceptible person with the infectious population in each time period, and β is the average number of potentially transmissible (“adequate”) contacts of one person with another person. The simple version leads to a set of ordinary differential equations:

$$\begin{aligned} dS(t)/dt &= -\beta I(t)S(t)/N, \\ dI(t)/dt &= \beta I(t)S(t)/N - \gamma I(t), \\ dR(t)/dt &= \gamma I(t), \end{aligned}$$

which sum to zero so N is constant. Given the initial conditions $R_1 = 0$, $I_1 = y_1$, and specific values of S_1 , β , γ , the ordinary differential equations can be integrated to find the time paths of S , I , R denoted by S_t , I_t , R_t . The model can be used to compute the evolution of an epidemic for the assumed parameters, where a richer model using Monte Carlo methods is generally employed for estimation.

Alternatively, we can estimate the parameters by nonlinear least squares from the observed data, as proposed

by Batista (2020). Given daily observations of $I_t + R_t$ in the form of confirmed cases y_t , we obtain the SIR residuals $\varepsilon_t(S_1, \beta, \gamma) = y_t - (I_t + R_t)$. We prefer to reparameterize the model in terms of $N_1/S_1, \beta, \gamma$ for a chosen N_1 and maximize $-\log \sum \varepsilon_t^2/T$. Thus, the assumption is that a susceptible person who becomes infectious (and then removed) is observed as infected among the confirmed cases.

Fig. 6 shows the forecast paths obtained by SIR compared with those from our average forecast described above for the recent UK history. The SIR model may be useful for describing the completed pandemic process, but these specific SIR forecasts offer very little guidance regarding the short-term movements in UK confirmed cases. Our procedure represented by average forecasts is unable to provide the sigmoid shape over long horizons but it provided better forecasts in the early stages and it will follow the slowdown toward and after the peak.

7. Conclusion

Accurate short-term forecasts of the COVID-19 pandemic are invaluable. The rapid increases in reported cases and deaths during the initial expansionary phase of the epidemic were often presented as surprising by the media, but they were partly predictable based on their trends. Forecasts also provide policy makers with advance warnings, thereby helping to allocate scarce public health resources and guide lockdown policy. We started producing real-time forecasts of COVID-19 from mid-March 2020 for many countries with the aim of addressing this need.

All forecasting models have different underlying assumptions and they use past data in different ways. Models based on well-established theoretical understanding and the available evidence are crucial for viable policy-making in disciplines based on observational data, but shifts in distributions can lead to systematic misforecasting. Consequently, there is an important role for short-term forecasts obtained using adaptive data-based models that are robust after distributional shifts. Our real-time forecasts of confirmed cases and deaths have filled this role, where they have proved timely, relevant,

and relatively accurate. Furthermore, an important aspect of COVID-19 forecasting is dating a peak in the epidemic, which is essential for indicating whether the disease is coming under control and if various policies are working. It is also important for monitoring to determine whether a second wave of the disease might occur. Flexible data-based short-term forecasts are essential for monitoring these events.

We found that our approach to modeling and forecasting, which was originally designed for application to economic data, yielded accurate forecasts for the COVID-19 data. The characteristics of economic data comprising non-stationarity in the form of stochastic trends and distributional shifts are also present in epidemiological data. Methods that adapt to this highly non-stationary, mis-measured data work well, and we showed that they can outperform several epidemiological models in the early stages, thereby providing an alternative complementary approach to forecasting.

We focused on short-term forecasts of seven days ahead. During the early stages of rapid growth, seven days seemed sufficiently long to be useful for policy makers, and many governments did not manage to implement much effective longer term planning. Forecasts of confirmed cases indicated the speed of spread, and deaths served as an indicator of the pressure on hospitals and care homes. It is unlikely that our forecasts would have much accuracy over longer horizons because of numerous policy reactions, both implemented and anticipated, as well as adaptive responses from individuals.

A further advantage of this data-based approach is that forecasting models are quicker to produce and run than structural models, which explains why the global forecasts obtained by [Petropoulos and Makridakis \(2020\)](#) and our forecasts were available from an early stage. As the pandemic slows down, the focus will shift to easing lockdowns and economic recovery, and thus more localized forecasts of confirmed cases will be helpful for preventing the resumption of growth, and our methods can be extended to this situation.

Our forecast method does not estimate structural parameters, and thus it provides no insights into the value of the reproduction number other than the growth of cases. Nonetheless, this nontheoretical approach to forecasting can obtain better forecasts than epidemiological models, as shown in the forecast comparisons with several such models. A combination of both approaches may be a fruitful avenue for future research.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

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 data collection. We also wish to thank Neil Ericsson and Andrew Martinez for helpful comments. A preliminary version of this paper was made available at: www.doornik.com/COVID-19 on 19 March 2020.

Appendix A. Countries included

Forecasts were made for the following 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, Luxembourg and Malta omitted.
- (3) UK, Norway, Switzerland (CH).

USA (54)

- (1) US,
- (13) CA, DC, NJ, NY, WA. More USA states from 2020-04-01 onward: CA, CO, CT, DC, FL, GA, IL, LA, MA, MI, NJ, NY, WA. No data available on 2020-04-02, after which we used the New York Times data.
- (53) Switched back to JH/CSSE on 2020-05-06 and included all states (and DC, Guam, and Puerto Rico).

Rest of the world (13)

- (7) Australia (AU), Brazil, Canada, Iran, Malaysia, Philippines, South Africa
- (4) India, Indonesia, Mexico, Turkey added from 2020-04-18.
- (1) Russia added from 2020-05-06.
- (1) Chile, Colombia, Mexico, Peru added from 2020-05-12.

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

Appendix B. Data revisions up to 2020-04-29

[Tables B.6](#) and [B.7](#) document the data revisions in the JH/CSSE data set when they exceeded 0.5%, but only for the countries for which we produced forecasts, and excluding USA states. It should be noted that there was usually a two-day gap between data downloads, so the actual revision date may have differed by one day.

For confirmed cases, we observed a large change for France when we ignored the revisions on March 23. We detect this first on April 17 (when comparing with the data from April 15) and noted that the revision applied to 10 observations (2020-04-04 to 2020-04-13, shortened to 04-04:13 in the table). The largest change of 28% was a downward revision by just over 30 000 cases. This is a

Table B.6

Revisions of cumulative confirmed count; JH/CSSE data. Each column is headed by the date when we detected the change. Each entry lists the date range affected and the largest absolute percentage change in the revision is shown in parentheses.

	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%)			

Table B.7

Revisions of cumulative deaths; JH/CSSE data. Each column is headed by the date when we detected the change. Each entry lists the date range affected and the largest absolute percentage change in the revision is shown in parentheses.

	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%)	

Table B.8

Revisions of cumulative deaths, ECDC data, and dates. Each column is headed by the ECDC date when we detected the change. Each entry lists the date range affected and the largest absolute percentage change in the revision is shown 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%)			

temporary change in the upward trend and the numbers were identical again on April 14. There was a correction to the dates for Mexico, where March 12 was deleted and the numbers shifted up, and the duplication of April 23 and 24 was also corrected.

The correction to the death count for Mexico was again an upward shift of the data, as shown in Table B.7. For the UK, there was a permanent redefinition from only counting deaths in hospitals to including non-hospital deaths (mainly care homes). For the USA, the revision applied to most of the history, but the observations for 2020-04-25 were again identical.

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

Table B.8 lists the revisions in the ECDC deaths data for the same countries up to 2004-04-27, which were all corrections for short periods. The forecast evaluation for Section 6.2 was based on the data from the first

available data set containing the outcomes. None of these corrections should have affected the MCIC forecasts.

Appendix C. Estimating the peak

Exponential growth is over when the daily cases have reached a peak and the reproduction of infections has slowed. The peak could be inferred from the actual daily increments but they tend to be erratic, so basing it on a smoothed trend produces more reliable estimates. We summarize the procedure used to determine the occurrence of a peak as follows (further details were provided by Castle, Doornik, and Hendry (2020)).

Averaged short-term forecasts were constructed as described in Section 4 from eight forecasts: DDL-TIS and DL-TIS both with damped $I(2)$, and each estimated up to T , $T-1$, $T-2$, $T-3$. Let τ_t denote the averaged smoothed trends from these eight models, and $C_t = \Delta\tau_t$ are the daily counts from the averaged smoothed trend. Next, let $\Delta_s y_t = y_t - y_{t-s}$ denote the s th backward difference and $\nabla_s y_t = y_t - y_{t+s}$ is the s th forward difference, where we drop the subscript s when $s = 1$.

A local peak requires that the observations before and after it are lower: $\Delta C_t > 0, \nabla C_t > 0$. For more certainty, we defined a local peak as $\Delta_i C_t > 0, \nabla_i C_t > 0, i = 1, \dots, p$. For confirmed cases, we set $p = 3$, and for deaths, which are more noisy, $p = 4$. A global peak was then defined as a local peak that was also the maximum of C_t . However, it was rejected if it was fewer than 5 days from the end of the sample and the last three were also sloping upward ($\Delta C_T > 0, \Delta C_{T-1} > 0$).

Finally, a peak was determined if there was either a global peak or if there was a global peak in $(C_{t-1} + 2C_t + C_{t+1})/4$ (setting $C_{T+1} = C_T$). After establishing the presence of a peak, exponential growth ended and we then switched the model from flexible trends to flexible steps, which provided more restrained trend forecasts. The actual peak was then derived from the averaged trends with the latter model. Clearly, a peak could disappear again after the resumption of growth. This was particularly noticeable for states in the USA.

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 Centre for Global Infectious Disease Analysis, mrc-ide.github.io/covid19-short-term-forecasts/index.html.
- Castle, J. L., Clements, M. P., & Hendry, D. F. (2015). Robust approaches to forecasting. *International Journal of Forecasting*, 31, 99–112.
- Castle, J. L., Doornik, J. A., & Hendry, D. F. (2019). Some forecasting principles from the M4 competition. *Economics Discussion*, Paper 2019-W01 Nuffield College, University of Oxford.
- Castle, J. L., Doornik, J. A., & Hendry, D. F. (2020). *Medium-term forecasting of the Coronavirus pandemic*. In progress Nuffield College, University of Oxford.
- Castle, J. L., Doornik, J. A., Hendry, D. F., & Pretis, F. (2015). Detecting location shifts during model selection by step-indicator saturation. *Econometrics*, 3(2), 240–264.
- Castle, J. L., Doornik, J. A., Hendry, D. F., & Pretis, F. (2020). *Trend-indicator saturation*. mimeo Department of Economics, Oxford University.
- Clements, M. P., & Hendry, D. F. (1999). *Forecasting non-stationary economic time series*. Cambridge, Mass: MIT Press.
- Doornik, J. A. (2009). Autometrics. In J. L. Castle, & 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*. Oxford: mimeo Nuffield College.
- Doornik, J. A., Castle, J. L., & Hendry, D. F. (2020). Card forecasts for M4. *International Journal of Forecasting*, 36, 129–134. [http://dx.doi.org/10.1016/j.ijforecast.2019.03.012](https://doi.org/10.1016/j.ijforecast.2019.03.012).
- Doornik, J. A., & Hendry, D. F. (2018). *OxMetrics: An interface to empirical modelling* (8th ed.). London: Timberlake Consultants Press.
- Doornik, J. A., & Juselius, K. (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 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., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, 36, 54–74. [http://dx.doi.org/10.1016/j.ijforecast.2019.04.014](https://doi.org/10.1016/j.ijforecast.2019.04.014).
- Petropoulos, F., & Makridakis, S. (2020). Forecasting the novel coronavirus COVID-19. *PLoS ONE*, 15, [http://dx.doi.org/10.1371/journal.pone.0231236](https://doi.org/10.1371/journal.pone.0231236).
- Walker, A., Pretis, F., Powell-Smith, A., & Goldacre, B. (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, Article 15205.