UNIVERSITY of York

This is a repository copy of *Modeling the Covid-19 Epidemic using Time Series Econometrics*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/177166/

Version: Accepted Version

Article:
Golinski, Adam orcid.org/0000-0001-8603-1171 and Spencer, Peter orcid.org/0000-0002-5595-5360 (Accepted: 2021) Modeling the Covid-19 Epidemic using Time Series Econometrics. Health Economics. ISSN: 1057-9230 (In Press)

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Modeling the Covid-19 Epidemic using Time Series Econometrics*

Adam Goliński† and Peter Spencer‡

Abstract

The classic 'logistic' model has provided a realistic model of the behaviour of Covid-19 in China and many East Asian countries. Once these countries passed the peak, the daily case count fell back, mirroring its initial climb in a symmetric way, just as the classic model predicts. However, in Italy and Spain and most other Western countries, the first wave of the epidemic was very different. The daily count fell back gradually from the peak but remained stubbornly high. The reason for the divergence from the classical model remain unclear. We take an empirical stance on this issue and develop a model framework based upon the statistical characteristics of the time series. With the possible exception of China, the workhorse logistic model is decisively rejected against more flexible alternatives.

1 Introduction

There are many different ways of analyzing and projecting the progress of an epidemic like Covid-19. Avery, Bossert, Clark, Ellison, and Ellison (2020) group these approaches into two broad types. On the one hand, there are the large computer models that analyze the spread of the disease and the effects of public health interventions in fine detail, like the one used by the UK government and its advisors (Ferguson et al., 2020). These 'mechanistic' models are largely theory-based and in that sense resemble the large theory-based models constructed by central banks to analyze the effect of their policy interventions on the economy. On the other hand, many epidemiologists fit curves suggested by the theoretical dynamics to the time series data and use these to make data-based predictions. An example is the model used by the Institute for Health Metrics and Evaluation ($IHME$) at the University of Washington to make the forecasts that underpin hospital resource planning in the US and White House briefings. This fits daily mortality data using the Gaussian bell curve (Murray, 2020). Avery, Bossert, Clark, Ellison, and Ellison (2020) classify these as 'phenomenological' models and note their resemblance to reduced-form econometric models. We also follow this data-based approach.

*We are grateful to Karl Claxton, Nigel Rice, Luigi Siciliani for insightful comments. We are also indebted to a referee for helpful suggestions.

†Department of Economics and Related Studies, University of York, YO10 5DD; adam.golinski@york.ac.uk.

‡Department of Economics and Related Studies, University of York, YO10 5DD; peter.spencer@york.ac.uk.

These two approaches complement each other nicely. Large-scale theoretical models are very useful for analyzing policy interventions and other structural changes but can miss important links, especially when confronted with 'black swan' events such as a financial breakdown or the emergence of a new variant of a virus like Covid-19. Moreover, it is hard to know what values to attribute to the parameters that govern the initial spread of a novel virus and the subsequent behavioural response. Small data-based models usually provide better forecasts. The Bank of England's Monetary Policy Committee uses both types of model for informing their decisions and forecasting (Burgess et al., 2013). Reduced form models can also be used to check the properties of the theoretical models and match them better to the data (Meenagh, Minford, and Wickens, 2009).

Although epidemiological models may differ in many respects, they are all based on the same underlying theory and invariably predict that the daily counts for infections and deaths follow a bell-shaped curve. In other words, once the peak is passed and the daily count begins to fall, it follows a path that mirrors the upward climb, before coming to an end. In the large mechanistic models, this classic pattern follows from the dynamics of the epidemic, which naturally slows as the disease runs through the population and immunity increases. Precautionary behaviour on the part of the public and policymakers are also likely to be important in slowing the spread of the disease. In the phenomenological models, the daily count follows a bell-shaped path by assumption and means that the cumulative count follows a sideways $S-$shaped logistic curve.

These symmetric dynamics have provided a reliable way of modelling outbreaks of influenza and other epidemics in the past. Indeed, simple regression models based on fitting the bell (or logistic) curve to the data, also proved accurate in predicting the path of the Covid-19 outbreak in China and many East Asian countries (Jia, Li, Jiang, Guo, and Zhao, 2020, Batista, 2020). However, the experience of Italy and Spain, which was followed by the US, the UK and many other countries, has been very different. The daily mortality figures have fallen back gradually from the peak in these countries, but have remained stubbornly high. This contrast is apparent in the daily infections series plotted for China and for Italy in Figure 1.

A positive skew in the national time series can appear because they aggregate data for areas that are hit by the virus at different times. However, data for hospital admissions and fatalities in hard-hit regions and cities like Lombardy in Italy and New York in the US also exhibit a pronounced skew.[1] A positive skew in the infections data may reflect measurement problems, such as improvements in the testing regime. It could also be due to the non-normality of community transmission. For example, the number of transmissions per person is known to have a long tail, due to the presence of 'superspreaders'. It has also been suggested that delays, such as the infection period, may have a gamma distribution with a long tail (Shen, Taleb, and Bar-Yam, 2020). Another plausible reason for the skew in mortality data is that the length of time from infection to death or recovery follows a gamma distribution (Hogg and Craig, 1978, Bird, 2013). Some people

[1]These data are available at: https://covidtracking.com/data.

recover or regrettably die very quickly, but others take much longer. This distribution is used by the MRC Biostatistics Unit to infer the true number of infections and the reproduction number (R) from mortality figures for the UK regions (Seaman and De Angelis, 2020). There is also a positive skew in the reporting lag (Bird and Neilsen, 2020, Birrell, Blake, van Leeuwen, and De Angelis, 2020).

Whatever the reason for this asymmetry, it is clear that the classic model has failed us badly this time. This has been documented by several recent studies. For example, Marchant, Samia, Tanner, and Cripps (2020) show that $IHME$ forecasts are usually overtaken by the data within a few days. We show that the classical model also fails in many other countries. Instead of trying to delve deeper into the data to try and find the reasons for this failure, we take an empirical stance and develop a model that is based upon the statistical characteristics of the national time-series. We model the daily mortality data for the first wave in fifteen countries published by the European Centre for Disease Control ($ECDC$). We use these data rather than infections because of the acute public and policy interest focus on these data and because they are less prone to measurement error.

We use the tools developed by econometricians to handle non-standard time-series, representing economic growth and speculative bubbles in financial markets for example, to analyze the dynamics of an epidemic. We compare the performance of the classical logistic model with flexible models based on the gamma and beta functions using an expanding data sample designed to mimic the data available to policymakers as an epidemic evolves in real time. At each stage, we identify the best fitting model using the Schwartz Information Criterion (SIC) and test forecasting performance using four-week projections.

This exercise identifies three distinct phases. The logistic model systematically under-forecasts over all three phases, but initially, in the upswing phase of the epidemic, its in-sample fit compares reasonably well with its rivals. But after a month or so, as the peak is passed and any asymmetries become apparent, the gamma and beta models fit much better than the logistic. Because they have greater flexibility to handle the initial stages of the epidemic, this is the case even in countries like Germany and Denmark that do not exhibit much of a skew. The gamma model is more robust than the beta in the face of data irregularities during this phase and generates more reliable forecasts. In the final phase, typically after a couple of months, the end-wave features become apparent. In most countries, mortality rates fall close to zero and the beta model with its distinct cut-off feature performs better than the gamma in terms of both in-sample fit and post-sample forecast accuracy. However, in the US, Brazil and Portugal, which move into the second wave without much of a hiatus, the gamma model still outperforms the beta over the full sample.

The next section of the paper sets out the three theoretical models and Section 3 explains how these are fitted to the data. The results are presented in Section 4. Section 5 offers some final observations and suggestions for future research.

2 Modelling an epidemic using time-series econometrics

Econometricians are used to dealing with difficult economic and financial times series. Their data often violate the classical assumptions adopted in the statistical texts and thus need to be handled using special techniques. For example, macroeconomic data like GDP exhibit exponential growth and financial prices can exhibit speculative bubbles that are explosive. They may respond with long and variable lags to policy interventions and exogenous shocks. These data may be measured with error and subject to structural shifts as behaviour or government policies change.

As Castle, Doornik, and Hendry (2020) argue, epidemiological data are fraught with similar problems. These econometricians have used sophisticated linear trend fitting techniques to decompose the cumulative death counts. They split each series into trend and remainder terms, then project them forward and recombine them to produce a forecast for the following week. As they note, a significant fall in outcomes relative to extrapolations from such models can be an indication that policy interventions are having the desired effect. Epidemiologists use similar models to separate the noise from the trend in the time-series and use the trend to estimate the reproduction number R, the number of people an infected person is likely to infect.

However, deviations from a linear trend can occur for many other reasons. For example, as the death toll mounts and people begin to worry about the virus and its consequences, they are likely to modify their behaviour in a way that reduces outcomes relative to a linear extrapolation. Longer term, the trend should bend as immunity builds up and the population becomes less susceptible to the disease. These endogenous feedback effects are built into the non-linear dynamics of the epidemiological models, which allow the trend to change as the epidemic progresses. This should in principle improve forecasts beyond the weekly horizon and make it more likely that systematic forecast errors are due to government interventions or other external influences.

2.1 The logistic process

These epidemiological models range from the large-scale computer models built by the Imperial College and other modelling groups to simple data-based curve-fitting techniques. For example, many epidemiologists fit a logistic curve to the cumulative number of infections $C(t)$:

$$\frac{C(t)}{K} = \frac{1}{1 + Ae^{-rt}}, \tag{1}$$

where: $A = K/C(0) - 1$, K is the final epidemic size and $r \geq 0$ the propagation or infection rate (see for example Batista (2020), equation (2)). However, in view of the well known issues around estimation with non-stationary data (Sims, 1980), we model the number of new cases. Differentiating (1) and substituting $Ae^{-rt} = K/C(t) - 1$ shows that the number of new cases at any time is a bell-shaped function of the

accumulated cases:

$$\frac{dC(t)}{dt} = rC(t)\left(1 - \frac{C(t)}{K}\right). \tag{2}$$

This model provides a simple way of allowing for the non-linear feedback mechanisms, loosely based on the SIR (susceptible, infectious, removed) model (Kermack, McKendrick, and Walker, 1927, Avery, Bossert, Clark, Ellison, and Ellison, 2020, Dimdore-Miles and Miles, 2020). Initially, with $C(0)$ cases observed when the outbreak is detected, all of them are 'infectious' and they will infect other 'susceptible' people at the rate r per unit of time (dt) causing $dC(0) = rC(0)dt$ new cases. Thus initially, the disease will spread exponentially, at the reproduction rate $\rho = dC(t)/(C(t)dt) = r$. However, various negative feedbacks then arise, which reduce the reproduction rate.

The classic feedback mechanism is provided by herd immunity. If people who have had the disease are less susceptible to catching it again, then they move into the 'removed' class. As they increase as a share of the population (N) the probability that an infectious person will meet a susceptible one falls from 1 to $(1 - C(t)/N))$. This results in $rC(t)(1 - C(t)/N)dt$ new cases per unit of time, resulting in (1) with $K = N$. However, there is a problem with this interpretation. If this were the only mechanism at work, we would expect K to be of a similar size to the population N. But it is much smaller than N empirically, suggesting that C is under-recorded. For example, Dimdore-Miles and Miles (2020) assume that the number of new cases that are symptomatic and recorded is a fraction π of the true number. If C represents the true number and C^o the recorded number, then substituting $C = C^o/\pi$ into (2) gives the model:

$$\frac{dC^o(t)}{dt} = rC^o(t)\left(1 - \frac{C^o(t)}{\pi K}\right).$$

Thus the estimator πK effectively replaces K. However, as they conclude the value of π would need to be extremely low to fully explain the low value of this estimate.

Precautionary feedbacks can also help to reduce the reproduction rate, as argued in the introduction. For example, as C grows, people are likely to modify their behaviour in a way that mimics the effect of immunity, reducing the reproduction rate $\rho = r(1 - C(t)/K)$ via the K parameter. This behaviour can be reinforced by government interventions like lockdown. On a more pessimistic view, if immunity from exposure to the disease is partial or tends to fall with the time since exposure, or indeed if the precautionary response depends upon the recent rather than the cumulative number of cases, then there may not be an upper limit to the cumulative number of cases.

The logistic model is designed to explain the transmission of a virus within a closed community. But apart from the country where the virus originates, all the initial cases will involve people that have recently entered the country and the number of new cases n will be related to the number of new arrivals rather than C. Thus in the initial stages, before community transmission begins: $dC(0) = ndt$ and not $dC(0) = nC(0)dt$

as implied by the logistic model.

However, the main problem with this model is that the bell and logistic curves are symmetric. The bell curve has a single peak at $C = K/2$. Once this is passed, the number of new cases begins to fall, following a path that mirrors the upward climb, before slowing to a stop as C approaches K. This was in fact the experience of China and many East Asian countries, which is why the logistic curve fits their data well. Unfortunately, the experience in many other countries has been very different. The number of deaths fell back from the peak, but then remained stubbornly high.

2.2 A long, thin-tailed epidemic

We need a more flexible model to allow for these possible effects. Mathematically, we can achieve this simply by raising the C and $(1 - C/K)$ terms in (2) by the powers α and β. This makes it a beta function, which is much more flexible:

$$\frac{dC(t)}{dt} = rC(t)^{\alpha}\left(1 - \frac{C(t)}{K}\right)^{\beta}. \tag{3}$$

Alternatively, suppose that the negative feedback effects mean that the reproduction rate ρ follows an exponential rather than a power law as the number of cases mounts: $\rho(t) = re^{-\gamma C(t)}$. Then:

$$\frac{dC(t)}{C(t)} = \rho(t)dt = re^{-\gamma C(t)}dt \Longleftrightarrow dC(t) = rC(t)e^{-\gamma C(t)}dt.$$

Importantly, in this model, there is no upper limit on the total number of cases as there is in the logistic and beta models. This makes it easier to fit to data sets in which the number of cases is falling but remains high, making it difficult to estimate the end-point parameter K. Theoretically, as noted, there could be situations in which the disease becomes endemic and there is no limit to the cumulative number of cases. The performance of this exponential feedback model can be improved by changing the power of the C term to α, thus giving the trend a form similar to that of the gamma density function:

$$dC(t) = rC(t)^{\alpha}e^{-\gamma C(t)}dt. \tag{4}$$

This function is used extensively in statistics to describe probability distributions, the well-known χ^2 distribution being a special case (Mood, Graybill, and Boes, 1973). Its mathematical properties are reviewed in Appendix 1.

These processes are non-stationary and should be handled using techniques developed for modelling non-stationary economic data, like growth and inflation. Their dynamics are dictated by stochastic differential equations with drift (i.e. trend) and volatility terms, like those used to model interest rates (Ait-Sahalia, 1996). We give the volatility term a form that is congruent with the drift.

3 Modelling the daily number of deaths

The logistic model outlined in the previous section was originally developed to explain the number of new infections. However, in the absence of mass testing, the *true* numbers of people who are infected and those that have recovered are likely to be much larger than those recorded, especially if there is a large proportion of asymptomatic cases. To avoid these measurement problems, we extend the reasoning of the logistic infections specification to track deaths instead, following Murray (2020) and many others.

Suppose for example that deaths $D(t)$ represent a constant lagged fraction of the *true* number of infections $C(t)$. Substituting this into (4) and suitably reinterpreting the parameters:

$$dD(t) = g(D(t)) = rD(t)^{\alpha} e^{-\gamma D(t)} dt. \tag{5}$$

We discretize (5) and use d_t to represent $dD(t)$ and the cumulative number of deaths as $D_t = \sum_{j=0}^{t-1} d_{t-j}$. Note that $d_t = D_t - D_{t-1}$.

In the empirical models we use ECDC data for daily mortality *rates*, which express the daily death count as a constant share of the population. This has the effect of normalizing the data to allow for country size, which is particularly important in cross-country comparisons such as these conducted in Section 4.5.[2] We use a rolling weekly average of the daily mortality rate instead of the daily series. This has the effect of smoothing out the erratic day-to-day movements often seen in the raw data as well as the weekend reporting lag seen in the US and several other countries. Specifically, denote the reported number of daily deaths in day t by d_t^o. We calculate its moving average $d_t = \frac{1}{7}\sum_{j=0}^{6} d_{t-j}^o$ and use it to find the cumulative value $D(t)$.

Finally, we add a congruent volatility specification:

$$d_t = rD_{t-1}^{\alpha} \exp\left(-\gamma D_{t-1}\right) + \sigma \left[rD_{t-1}^{\alpha} \exp\left(-\gamma D_{t-1}\right)\right]^{\delta} \epsilon_t, \tag{6}$$

where α, γ, r and σ are parameters to be estimated and $\epsilon_t \sim N(0,1)$ is a Gaussian error term. These variables are measured as a share of the population, per 10 million people. Similarly, the beta model corresponding to (3) is:

$$d_t = rD_{t-1}^{\alpha}(1 - D_{t-1}/K)^{\beta} + \sigma \left[rD_{t-1}^{\alpha}(1 - D_{t-1}/K)^{\beta}\right]^{\delta} \epsilon_t. \tag{7}$$

Setting $\alpha, \beta = 1$ simplifies this to the logistic specification:

$$d_t = rD_{t-1}(1 - D_{t-1}/K) + \sigma \left[rD_{t-1}(1 - D_{t-1}/K)\right]^{\delta} \epsilon_t \tag{8}$$

We found that δ was close to 0.75 for these models and countries and fixed this parameter at this value in

[2] We are particularly grateful to a referee for this suggestion. This is a scaling adjustment that does not affect the fit but removes the effect of the size of the population on the K and γ parameters.

the regressions reported here.[3]

4 Modeling the ECDC death data for the first wave

The previous section identifies three candidate models (logistic, gamma and beta). We estimate these models using data for the first wave provided by the $ECDC$. This source provides daily death (and infections) data from 1st January to 14th December 2020, when the $ECDC$ discontinued the daily series due to the effects of retrospective corrections, delays in reporting and similar problems.

To select the best model, we consider both the in-sample fit and (apart from China) the post-sample forecasting performance. To rank the models by fit we use the SIC, which adjusts the likelihood value appropriately for the number of parameters and observations to guard against over-fitting. To avoid the bias in estimates caused by an integer data count we start the estimation for each country from the date when the cumulative number of deaths exceeds 1.5 per 10 million people. The end of sample for each country is determined by the end of the first wave. Specifically, we end it at the beginning of the two-weekly period with the minimum death toll, counting from the beginning of the sample. (Please see the tables for more detail.) We then check that the best fitting model tends to provide the best forecasts. Finally, we report the full sample parameter estimates as well the estimates from an eight-week data sample, used to represent an on-going epidemic.

4.1 China

Table 1 shows the regression results obtained for the three rival models using the data for China. This was of course the first country to be hit by Covid-19 and managed to suppress it effectively by the end of March, when we end this sample. Figure 2 shows the in-sample fit of the logistic (red line), gamma (green line) and beta (blue line) regression models. As noted in the introduction, the bell-logistic model represents the behaviour of this outbreak nicely, although the beta drift model is better in terms of statistical criteria and with $\beta > \alpha$ indicates a small positive skew in the data.

4.2 Italy and Spain

We next analyze the daily mortality data for Italy and Spain, which were the first western countries to be overwhelmed by the virus and where the skew in the mortality figures first became apparent. The top panel of Figure 3 shows how the models fit the weekly average of the daily death data over the full sample. Table 2 reports the full sample parameter estimates. The beta model performs best for both countries over the full period. Arguably, a more relevant test is to ask how well these models fit and forecast as the epidemic evolves

[3]The Matlab code is available on the website: https://sites.google.com/york.ac.uk/adam-golinski/coronametrics.

in real time, helping to inform the tightness of government policy measures. To assess this, we estimate a weekly series of regressions with an expanding time window. In the second row panel of Figure 3, we plot the SIC value for each weekly regression for the three models.

This exercise identifies three distinct phases. Initially, the SIC criterion finds difficult to discriminate between these models. That is likely to be because they only differ in the way that they represent the negative feedback effects, which are not very powerful initially, making it very difficult to predict the final death toll.[4] However, after another a month, the gamma and beta models begin to outperform the logistic. This outperformance becomes more pronounced as the peak is passed and the data becomes more informative about the negative feedback effects. As noted in the introduction, this is because the daily mortality figures during the first wave were asymmetric: the downswing was more gradual than the initial upswing. Consequently, the logistic model, being symmetric, fits poorly during this phase and systematically under-predicts the subsequent number of deaths.

The poor forecasting performance of the logistic model is clear from the 'hedgehog' forecast charts shown in the third row panel. To construct this type of chart, we use parameter estimates from the expanding weekly sample models to make a succession of four-week ahead forecasts. The dots in this chart, which form the back of the hedgehog, show the cumulative number of deaths observed at the end of each week. The forecasts are shown as the spines of the hedgehog. After a couple of months, we identify a third and final phase, in which the beta models begin to outperform the gamma model. This is because the logistic and beta functions have a well-defined end value (K) for the final death toll, relative to population (which the gamma does not), terminating the wave decisively.

To illustrate the way that these models represent an on-going epidemic, Table 4 reports a set of parameter estimates for the eight-week data sample. The gamma and beta models both perform well at this point, significantly outperforming the logistic. Note in particular the very low K values for the final death toll implied by the logistic model. The K-values from the beta model provided a much more realistic projection of the final outcome. The bottom panel of Figure 3 shows how these models fit the four- and eight-week data samples.

4.3 The US and the UK

We next look at the US and the UK, two of the countries with the highest mortality rates during the first wave. The top panels of Figure 4 shows how the models track the full sample, the two central panels plot the SIC values and 4-week forecasts for the weekly regression models and the bottom panels shows the four- and eight-week fit. The parameter estimates are reported in Tables 2 and 4. Once again, we see an initial phase

[4]Reflecting this, our preliminary analysis of the data conducted in real time, suggested that the logistic model offered a useful tool for analyzing the initial spread of the virus, helping to identify the peak and to make tentative short-range forecasts. However, it proved difficult to predict the final death toll at this stage.

in which all three models have a similar fit, followed by a steady outperformance of the gamma and beta models in terms of both the fit and the forecasting performance. In the UK, we identify a third and final phase after around two months, in which the beta clearly outperforms the gamma model in these respects and the logistic does as well as the gamma model. The UK resembles the Italian experience in this sense.

However, the US experience was then very different because the weekly death toll never reached the lows seen in these countries over the late summer months. The beta and gamma models perform equally for the US until the mortality figures begin to move back up again in September, marking the beginning of the second wave. It is perhaps worth noting that the curves generated by these two US models overlap and smooth out the lumpy figures announced around the peak of the epidemic. The erratic nature of these data make forecasting very difficult and lead to large errors over the following month. However, the subsequent forecasting performance of the US gamma model is particularly impressive.

4.4 Other countries

We estimate these models for eight other West European countries[5]: France, Germany, Belgium, Denmark, Ireland, the Netherlands, Portugal and Sweden. We also estimate models for Canada and Brazil. Performance is shown graphically in Figures 6-9. The parameter estimates are reported in Tables 2 - 5.

The patterns observed in Italy, Spain, the US and the UK are also apparent in these figures. The SIC criterion reliably selects the model that provides the best forecasts. With the exception of Brazil and Portugal, which are similar to the US in progressing directly into a second wave, the beta drift model fits much better than the other models over the full period. However, over the first two or three months the gamma drift model invariably fits as well, if not better, and provides more robust forecasts than the beta. The gamma model fits and forecasts remarkably well for Brazil. In this case, the very low eight-week estimate of γ stands out, indicating that the negative feedback effects are still very weak at this point, with the epidemic still in the upswing phase. The logistic model systematically under-forecasts the spread of the virus in all these countries, unless and until the epidemic ends.

These patterns stand out despite the very different experiences of these countries during the first wave. Germany successfully combined a lockdown with mass population testing and had a much lower mortality rate than France (Figure 5) and other countries. Denmark was the second European country after Italy to go into lockdown, on 11 March, before any fatalities had occurred. Its peak mortality rate was similar to that seen in Germany and much lower than in its neighbour Sweden, which was exceptional in having relied upon individual responsibility rather than lockdown to contain the spread of the virus.

[5]East European countries remained largely unaffected by the first wave of the epidemic.

4.5 International comparisons

Time series models are designed to abstract from the noise in the data and provide estimates of the trend in the series. In the case of an on-going, non-linear process like an epidemic, they can also be used to indicate the rate at which the trend is increasing or decreasing and whether this rate is accelerating or decelerating. In this case, the trend in the cumulative death toll is the estimated number of daily deaths, described by the a drift function like equation (7). These properties can be seen from the shape of the curves in Figures 3 to 9. However, rather than 'eyeballing' charts it is often better to look at the results numerically, using well-known statistics, particularly when comparing different countries. Table 6 shows some of the basic numbers that describe the spread of the epidemic in the US, Canada and West European countries at the eight week stage. We use the gamma drift model to represent this. Its parameters arguably allow a broader assessment of the characteristics of the virus than comparisons of death tolls in different countries.

These statistics follow from the well-known mathematical properties of gamma-type functions, which are reviewed in the Appendix. The first three columns show the parameters estimated for each country, reproduced from Tables 4 and 5. These are first used to determine when the peak in the death toll is likely to have occurred. This can be difficult to gauge from the data visually, especially in countries like the US where the data is lumpy around the peak due to measurement problems. The table shows the date that the peak was reached in each country; the estimated number of daily deaths at the peak (corresponding to the height of the peak in each figure) and the cumulative number of deaths at that point. We then calculate the 'skewness' coefficient for each country, shown in column (vii). This indicates how different the decline from the peak was compared to the rise from the first few cases to the peak. The rows of this table are ranked in terms of skewness, starting with Belgium at the top.

One of the stand-out features of this table is the strong positive correlation between the skew and the daily ($R^2 = 0.94202$) and especially the cumulative ($R^2 = 0.99999$) mortality rates at the peak. Countries like Germany and Denmark that are judged to have dealt with the epidemic effectively compare very favourable in all of these respects with those like the UK and Belgium that were not. More interesting is the observation that all three variables are negatively correlated with the parameter γ, which of course acts as an indicator of the strength of the negative feedback effects. This suggests *prima facie* that the negative feedback effects are largely due to the effectiveness of government policy and precautionary behaviour by the public rather than herd immunity. Indeed, if we take the inverse of the gamma parameter (effectively expressing the exponent in (5) as D/K, as in the other two functions, rather than $\gamma \times D$), its correlation with the skew ($R^2 = 0.91409$), the daily ($R^2 = 0.99413$) and cumulative ($R^2 = 0.99371$) mortality rates is also very high. Importantly, this does not necessarily follow from the gamma function, since, for example, equation A1 in the Appendix shows that the skew depends upon both the α and the γ parameters. It only follows in Table 6 because the variation in α is small compared to the variation in γ.

Although this research project was initially aimed at modelling the features of the stubborn upper tail seen in Italy and elsewhere, these results also help us to a better understanding of the early, exponential, phase of this pandemic. In this early, pre-peak phase, attention is focussed on the time that it takes for the cumulative number of infections to double. One of the striking features of Tables 2 to 3 is that the estimates of the parameter $\rho(1) = r$ from the bell-logistic model are remarkably close, ranging from 0.07 for Sweden to 0.11 for Belgium . This parameter is important because it shows the daily growth rate during the initial phase of the epidemic, before the various negative feedback effects are significant. Thus the logistic model would suggest that country-specific factors are not significant in explaining the initial spread of the virus. However, the gamma drift model, which allows more flexibility in tracking the initial phase through the parameter α, suggests that the logistic model covers up important idiosyncratic effects. Table 6 shows the values of $\rho(0.5)$ and $\rho(1.5)$ and the respective doubling times during the initial phase. With $D = 0.5$ the doubling time ranges widely, from less than a day in Denmark to 3.14 for France, in strong contrast to the impression given by the logistic model.

5 Conclusion

This paper shows how the econometrician's toolkit can be used to develop a simple reduced form model of the time series generated by an epidemic. We illustrate this using daily mortality data generated by the first wave of Covid-19 for fourteen American and European countries. We use standard model selection techniques to find the model that best fits in-sample at any stage of the epidemic and show that this reliably generates the most accurate post-sample forecasts. With the exception of China, the logistic model frequently employed by epidemiologists to model time series data is decisively rejected against the more flexible gamma and beta models. These handle the very different wave characteristics seen in these countries remarkably well.

These time series models provide useful statistics that summarize the reproduction, morbidity and mortality rates in different countries. We could use these to look at the effects on these indicators of variations in containment and testing strategies across a cross-section of countries, while controlling for different demographic and other characteristics. One of the interesting findings that emerges from the present study is that there is a strong correlation between the parameters like γ and K that represent the strength of the negative feedback and the skew and peak mortality rates. The epidemic was less severe in countries like Germany, Denmark (and indeed China) that were generally regarded as being effective in dealing with the virus than they were in others like Belgium and the UK that were not and it seems likely that these feedback parameters reflect the efficacy of government policy.

However, such reduced form models have their limitations. Their dynamics are the result of a convolution of the long and possibly variable lag distributions involved in the data generation process and it is impossible

to unravel these without embedding them in a large scale structural model. In the present context, despite the *prima facie* argument that the negative feedback results from policy and precautionary behaviour rather than from the build up of immunity, we cannot be sure. Nor can we say whether the skew seen in the mortality data in many western countries during the first wave was due to the skew in the lag from infection to reinfection, the lag from infection to death or other effects. The difference between the experiences of these countries and East Asian countries remains to be explained.

Nevertheless, looking forward, we can potentially use reduced form models to identify structural breaks and perform similar tasks. Econometricians have a variety of handy tools for conducting this kind of work, including tests for discrete changes when the break-point is unknown a priori (due to the advent of a new variant for example) as well as tests for breaks at points when a change is likely to have occurred (due to a policy intervention for example). The explosive behaviour test proposed in a series of papers by Phillips, Shi and Yu (Phillips, Shi, and Yu, 2014, Phillips, Shi, and Yu, 2015) also looks potentially useful in this respect. This was developed for testing bubbles in financial time series, which are analogous to the early phase of a new virus or variant.

References

Ait-Sahalia, Y. (1996): "Testing continuous time models of the spot interest rate," *Review of Financial Studies*, 9, 385–42.

Avery, C., W. Bossert, A. Clark, E. Ellison, and S. Ellison (2020): "Policy implications of models of the spread of coronavirus: Perspectives and opportunities for economists," *Covid Economics, CEPR*, 12, 21–28.

Batista, M. (2020): "Estimation of the final size of the coronavirus epidemic by the logistic model," Discussion paper.

Bird, S. (2013): "Editorial: Counting the dead properly and promptly," *Journal of the Royal Statistical Society Series A*, 176, 815–817.

Bird, S., and B. Neilsen (2020): "Nowcasting of Covid-19 deaths in English hospitals," *mimeo, Nuffield College*.

Birrell, P., J. Blake, E. van Leeuwen, and D. De Angelis (2020): "Nowcasting," *Joint PHE Modelling Cell and MRC Biostatistics Unit COVID-19 Working Group*.

Burgess, S., et al. (2013): "The Bank of Englands forecasting platform: COMPASS, MAPS, EASE and the suite of models," *Working Paper No. 471*.

Castle, J. L., J. A. Doornik, and D. F. Hendry (2020): "Short-term forecasts of COVID-19 - preliminary version," *mimeo, Department of Economics, Nuffield College*.

Dimdore-Miles, O., and D. Miles (2020): "Assessing the spread of the novel coronavirus in the absence of mass Testing," *Covid Economics, CEPR*, 16.

Ferguson, N., et al. (2020): "Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand," *mimeo, Imperial College COVID-19 Response Team*.

Hogg, R. V., and A. T. Craig (1978): *Introduction to mathematical statistics (4th ed.)*. New York: Macmillan.

Jia, L., K. Li, Y. Jiang, X. Guo, and T. Zhao (2020): "Prediction and analysis of coronavirus disease 2019," *mimeo, Stanford University*.

Jones, A. M., J. Lomas, and N. Rice (2015): "Healthcare cost regressions: Going beyond the mean to estimate the full distribution," *Health Economics*, 24(9), 1192–1212.

Kermack, W. O., A. G. McKendrick, and G. T. Walker (1927): "A contribution to the mathematical theory of epidemics," *Proceedings of the Royal Society of London. Series A*, 115(772), 700–721.

Marchant, R., N. Samia, M. Tanner, and S. Cripps (2020): "Learning as we go: An examination of the statistical accuracy of COVID19 daily death count predictions," *mimeo, University of Sidney*.

Meenagh, D., P. Minford, and M. Wickens (2009): "Testing a DSGE model of the EU using indirect inference," *Open Economies Review*, 20, 435–471.

Mood, A. M., F. A. Graybill, and D. C. Boes (1973): *Introduction to the Theory of Statistics, 3rd Edition*. McGraw Hill.

Murray, C. (2020): "Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months IHME COVID-19 health service utilization forecasting team," *IHME COVID-19 health service utilization forecasting team*.

Phillips, P. C., S.-P. Shi, and J. Yu (2014): "Specification sensitivity in right-tailed unit root testing for explosive behaviour," *Oxford Bulletin of Economics and Statistics*, 76, 315–333.

——— (2015): "Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500," *International Economic Review*, 56(4), 1043–1078.

Seaman, S., and D. De Angelis (2020): "Challenges in estimating the distribution of delay from COVID-19 death to report of death," *Joint PHE Modelling Cell, and MRC Biostatistics Unit*.

Shen, C., N. N. Taleb, and Y. Bar-Yam (2020): "Review of Ferguson et al "Impact of non-pharmaceutical interventions"," *New England Complex Systems Institute*.

Sims, C. A. (1980): "Macroeconomics and Reality," *Econometrica*, 48(1), 1–48.

Appendix: Properties of the gamma drift function

The mathematical properties of gamma-type functions are well known. This function is used to define the gamma probability distribution Mood, Graybill, and Boes (1973). Health economists are also used to dealing with skewness and kurtosis. For example Jones, Lomas, and Rice (2015) use generalized gamma and beta functions to handle the tails in health cost distributions.

The peak in mortality

The value of the drift in (6) at any time gives the expected number of deaths and is obtained by substituting the cumulative number of deaths at that time. However, analytically, it is often easier to work in continuous rather than discrete time, using (5) instead of (6). Gamma-type functions have a single peak. This is found by taking the first derivative of the drift:

$$\frac{\partial}{\partial D} r D^{\alpha} e^{-\gamma D} = r D^{\alpha-1} e^{-D\gamma} \left(\alpha - D\gamma\right)$$

and equating this to zero by setting $D = \alpha/\gamma$. These values are shown in column of the first panel of Table 6 and substituting them back into (5) gives the estimate of the daily death toll at the peak. The peak corresponds to the mode in the gamma distribution. Similarly, the skewness coefficient shown in column (vii) follows from that of the gamma distribution (Mood, Graybill, and Boes (1973)):

$$Skew = \frac{(\alpha+1)(\alpha+2)(\alpha+3)}{\gamma^3}. \tag{A1}$$

Doubling time

Although this paper is focussed on the features of the stubborn upper tail, our results also help us to a better understanding of the early, exponential, phase of this pandemic. In that phase, attention is focussed on the time that it takes for the cumulative number of infections to double. This can also be used to assess the initial behaviour of the mortality rate. But to gauge that we need to have an estimate of the reproduction rate $\rho(D(t))$. Table 6 shows three estimates implied by the gamma drift function (7), expressed as the expected daily change as a decimal fraction of the cumulative. The first is evaluated at the start of the sample with $D = 1.5$, suggesting that is in the range of $0.2 - 0.6$ for most countries. The second is a pre-sample extrapolation value for $D = 0.5$. The third is the value at the peak.

We can then calculate the doubling time by dividing (5) by D to get the percentage or logarithmic change:

$$d \ln D(t) = dD(t)/D(t) = \rho(D(t))dt = rD(t)^{\alpha-1} e^{-\gamma D(t)} dt,$$

which we can integrate to get the approximation: $\ln D(t+\tau)/D(t) = \rho(D(t))\tau$, where τ is measured in days. The doubling time is found by setting $D(t+\tau)/D(t) = 2$ and solving for τ :

$$\tau = \ln 2/\rho(D(t)) = 0.693/\rho(D(t)).$$

As noted in Section 4.5, the values in the table range widely across different countries.

Tables and figures

Model	SIC	r	K	α	β	$\gamma \times 10,000$	σ
			China				
Bell	7.9117	0.1578 0.0080	3,305 3	-	-	-	1.0310 0.0961
Gamma	8.5713	0.1648 0.0561	-	1.1527 0.0684	-	13.650 1.036	1.1969 0.1166
Beta	5.4226	0.7353 0.0635	3,347 7	0.7904 0.0148	1.1470 0.0323	-	0.2619 0.0226

Table 1: Estimates of the logistic, gamma and beta models. Standard errors are reported in small font. The estimation period is from 23 January to 31 March 2020.

Model	SIC	r	K	α	β	$\gamma \times 10,000$	σ
			Italy (01/03/2020 - 02/08/2020)				
Logistic	7.9062	0.0826 0.0046	5,526 1.27	-	-	-	1.5935 0.1128
Gamma	7.8276	0.1234 0.0296	-	1.1771 0.0440	-	9.7435 0.3990	1.1917 0.0806
Beta	4.7824	1.1271 0.1048	5,830 6.57	0.6834 0.0142	1.2163 0.0243	-	0.3114 0.0181
			Spain (12/03/2020 - 16/07/2020)				
Logistic	8.4716	0.1017 0.0068	6,461 1.15	-	-	-	1.8673 0.1515
Gamma	8.4090	0.1803 0.0479	-	1.1517 0.0471	-	8.7862 0.3752	1.2378 0.0912
Beta	6.8855	2.8428 0.6945	6,464 3.52	0.5626 0.0352	0.9622 0.0377	-	0.7590 0.0523
			US (18/03/2020 - 08/07/2020)				
Logistic	8.6075	0.0766 0.0052	4,007 19.39	-	-	-	1.6842 0.1395
Gamma	4.9271	0.4083 0.0302	-	0.8774 0.0134	-	8.8797 0.1404	0.2355 0.0159
Beta	4.9720	0.4101 0.0303	466,913 91,619	0.8761 0.0134	412.3503 81.1246	-	0.2359 0.0157
			UK (17/03/2020 - 01/09/2020)				
Logistic	7.9793	0.0790 0.0046	6,226 1.33	-	-	-	1.6887 0.1175
Gamma	7.9272	0.1062 0.0267	-	1.2231 0.0458	-	9.8539 0.3996	1.3202 0.0891
Beta	4.5045	1.0576 0.0947	6,280 5.53	0.7060 0.0134	1.3073 1.0214	-	0.3037 0.0169
			Canada (22/03/2020 - 06/09/2020)				
Bell	6.1274	0.0695 0.0033	2,445 1.27	-	-	-	1.1160 0.0725
Gamma	6.1712	0.0780 0.0170	-	1.2314 0.0470	-	22.4705 0.9927	1.0066 0.0659
Beta	3.7953	0.3642 0.0369	2,500 6.64	0.8075 0.0183	1.4263 0.0405	-	0.3447 0.0194
			Brazil (28/03/2020 - 07/11/2020)				
Bell	8.2387	0.0370 0.015	7,820 33.84	-	-	-	1.4078 0.0782
Gamma	4.9886	0.3222 0.0188	-	0.7513 0.0095	-	3.1001 0.0532	0.2402 0.0115
Beta	4.7114	0.4958 0.0300	10,456 435.49	0.6565 0.0103	1.6636 0.1506	-	0.2054 0.0098

Table 2: Full sample estimates of the logistic, gamma and beta models. Standard errors are reported in small font.

Model	SIC	r	K	α	β	$\gamma \times 10,000$	σ
Belgium (15/03/2020 - 05/07/2020)							
Logistic	9.2934	0.1106 0.0068	8,4219 3.13	-	-	-	1.7737 0.1436
Gamma	9.2518	0.1644 0.0442	-	1.1559 0.0468	-	6.5104 0.3128	1.4130 0.1111
Beta	5.0843	1.1039 0.0623	8,539 9.17	0.7435 0.0083	1.3049 0.0179	-	0.2001 0.0135
The Netherlands (14/03/2020 - 18/07/2020)							
Logistic	6.9724	0.1041 0.0063	3,550 1.58	-	-	-	1.3729 0.1277
Gamma	7.5372	0.1334 0.0371	-	1.1999 0.0563	-	15.8447 0.8330	1.2587 0.0980
Beta	4.3135	1.0951 0.1092	3,569 3.11	0.6867 0.0163	1.1468 0.0249	-	0.3154 0.0203
France (11/03/2020 - 29/07/2020)							
Logistic	7.5175	0.0959 0.0055	4,507 1.18	-	-	-	1.4726 0.1084
Gamma	7.4870	0.0667 0.0158	-	1.3608 0.0473	-	14.6807 0.6061	1.2611 0.0932
Beta	4.4665	0.4076 0.0380	4,630 11.91	0.8893 0.0155	1.6833 0.0401	-	0.3230 0.0197
Germany (21/03/2020 - 05/09/2020)							
Logistic	5.1518	0.0734 0.0047	1,122 0.4203	-	-	-	1.2248 0.0887
Gamma	4.6092	0.0885 0.0213	-	1.4104 0.0629	-	63.5154 2.8542	0.9395 0.0668
Beta	1.5931	0.4767 0.0399	1,157 2.91	0.8043 0.0173	1.7157 0.0400	-	0.2367 0.0132
Denmark (20/03/2020 - 21/08/2020)							
Logistic	4.6832	0.0813 0.0058	1,070 0.02	-	-	-	1.3750 0.1012
Gamma	4.7736	0.1234 0.0412	-	1.2720 0.0680	-	60.8744 2.9943	0.9256 0.0672
Beta	2.8761	0.8059 0.1260	1,101 4.67	0.6937 0.0320	1.5538 0.0662	-	0.4131 0.0250
Sweden (13/03/2020 - 28/08/2020)							
Logistic	8.0976	0.0675 0.0037	5,680 2.14	-	-	-	1.6309 0.1116
Gamma	7.5395	0.2110 0.0423	-	1.0138 0.0353	-	8.2822 0.3063	0.9736 0.0593
Beta	5.7591	1.2272 0.1653	5,721 9.32	0.6216 0.0204	1.0739 0.0322	-	0.4605 0.0260
Ireland (16/03/2020 - 31/08/2020)							
Logistic	6.2453	0.0854 0.0045	3,623 0.07	-	-	-	1.4050 0.0854
Gamma	6.9755	0.0466 0.0127	-	1.4575 0.0590	-	20.1799 1.9434	1.4507 0.1116
Beta	4.7667	0.4531 0.0705	3,6919 11.72	0.8477 0.0266	1.5749 0.0610	-	0.5644 0.0333
Portugal (22/03/2020 - 23/08/2020)							
Logistic	7.5537	0.0591 0.0053	1,743 2.39	-	-	-	2.0289 0.1787
Gamma	4.7037	0.5786 0.0911	-	0.8522 0.0328	-	26.1216 0.7786	0.4910 0.0293
Beta	4.3883	1.1209 0.1935	2,371 160.87	0.6395 0.0381	2.9911 0.4672	-	0.4167 0.0246

Table 3: Full sample estimates of the logistic, gamma and beta models. Standard errors are reported in small font.

Model	SIC	r	K	α	β	$\gamma \times 10,000$	σ
Italy (01/03/2020 - 16/04/2020)							
Logistic	8.4873	0.1520 0.0089	4,3830 57.63	-	-	-	1.1210 0.1140
Gamma	4.8970	0.4998 0.0250	-	0.8707 0.0092	-	5.5416 0.1004	0.1513 0.0143
Beta	5.1327	0.5371 0.0336	19,127 7,434.60	0.8506 0.0130	9.1594 4.1333	-	0.1642 0.0158
Spain (12/03/2020 - 7/05/2020)							
Logistic	9.7410	0.1596 0.0111	5,528 37.93	-	-	-	1.5203 0.1610
Gamma	6.4511	0.5670 0.0440	-	0.9063 0.0136	-	5.7420 0.1137	0.2440 0.0231
Beta	6.0448	0.7220 0.0660	11,707 2,309.90	0.8444 0.0194	4.7457 1.3590	-	0.1908 0.0195
US (18/03/2020 - 13/05/2020)							
Logistic	7.1653	0.1482 0.0079	2,550 37.79	-	-	-	0.8704 0.0872
Gamma	4.6268	0.3964 0.0285	-	0.8858 0.0147	-	9.0802 0.2766	0.2034 0.0193
Beta	4.7256	0.4216 0.0292	229,161 88,439.62	0.8703 0.0136	199.1241 73.6595	-	0.2054 0.0194
UK (17/03/2020 - 12/05/2020)							
Logistic	8.7651	0.1537 0.0086	4,794 50.04	-	-	-	1.1210 0.1135
Gamma	4.9953	0.4682 0.0213	-	0.8929 0.0082	-	5.4596 0.0815	0.1402 0.0133
Beta	5.1558	0.5722 0.0304	9,446 867.75	0.8397 0.0103	3.5651 0.4871	-	0.1461 0.0147
Canada (22/03/2020 - 17/05/2020)							
Bell	5.5679	0.1376 0.0057	1,612 32.32	-	-	-	0.5583 0.0543
Gamma	3.6031	0.3184 0.0219	-	0.8642 0.0156	-	10.5394 0.4900	0.1793 0.0170
Beta	3.9660	0.3643 0.0301	2,456 380.46	0.8182 0.0191	1.6071 0.04235	-	0.2071 0.0208
Brazil (28/03/2020 - 23/05/2020)							
Bell	4.7548	0.1136 0.0053	1,248 73.79	-	-	-	0.5176 0.0507
Gamma	2.2857	0.3844 0.0211	-	0.7027 0.0139	-	0.1000 0.7289	0.1303 0.0123
Beta	4.2741	0.7484 0.0665	20,531 2,078.31	0.5798 0.0162	0.0050 0.0748	-	0.3675 0.0165

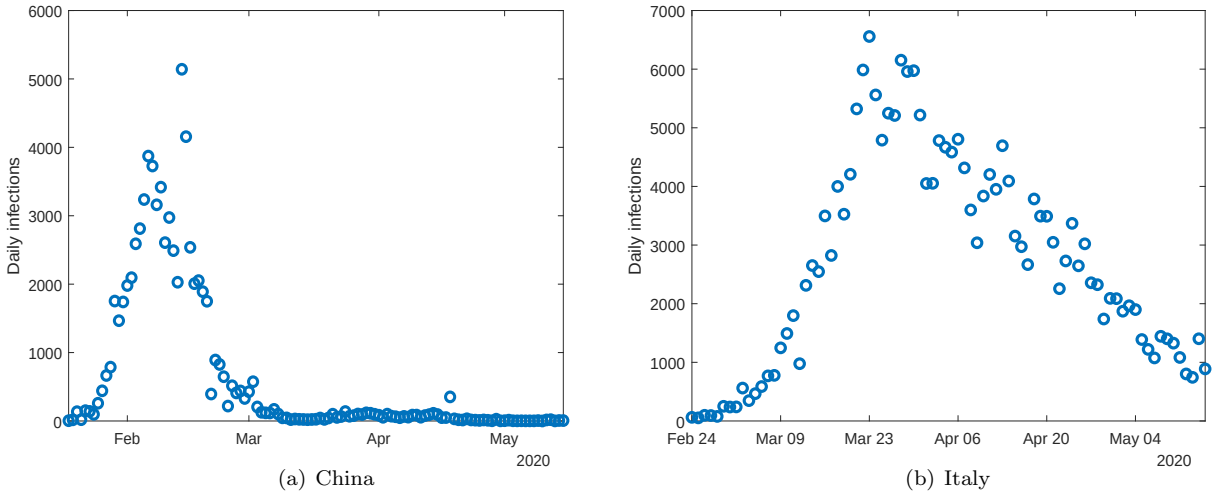
Table 4: Estimates of the logistic, gamma and beta models at the 8-week stage. Standard errors are reported in small font.

Model	SIC	r	K	α	β	$\gamma \times 10,000$	σ
Belgium (15/03/2020 - 10/05/2020)							
Logistic	9.9357	0.1635 0.0100	7,466 49.18	-	-	-	1.4080 0.1450
Gamma	8.0598	0.5382 0.0702	-	0.9029 0.0220	-	3.9301 0.1469	0.4594 0.0439
Beta	6.9573	0.7156 0.0480	12,638 652.71	0.8356 0.0100	3.1893 0.2633	-	0.2524 0.0155
The Netherlands (14/03/2020 - 9/05/2020)							
Logistic	8.3878	0.1510 0.0100	3,094 32.17	-	-	-	1.2187 0.1272
Gamma	5.1897	0.5964 0.0432	-	0.8461 0.0140	-	8.3912 0.2060	0.2022 0.0192
Beta	5.2603	0.6035 0.0435	72,270 23,219	0.8425 0.0139	59.1401 19.4683	-	0.2022 0.0190
France (11/03/2020 - 6/05/2020)							
Logistic	7.5114	0.1720 0.0072	3,806 23.03	-	-	-	0.7774 0.0761
Gamma	6.4715	0.2281 0.0268	-	1.0469 0.0226	-	8.5744 0.3037	0.4054 0.0388
Beta	5.5733	0.3423 0.0286	4,920 256.19	0.9337 0.0167	2.1520 0.2508	-	0.2541 0.0233
Germany (21/03/2020 - 16/05/2020)							
Logistic	6.0344	0.1442 0.0081	947 7.57	-	-	-	0.7651 0.0775
Gamma	3.6587	0.4031 0.0337	-	0.8859 0.0206	-	28.8674 0.9494	0.2016 0.0191
Beta	2.3962	0.5726 0.0286	1,200 246.36	0.7573 0.0130	1.6599 0.1525	-	0.1021 0.0097
Denmark (20/03/2020 - 15/05/2020)							
Logistic	6.8360	0.1399 0.0105	923 7.17	-	-	-	1.0642 0.1147
Gamma	3.8710	0.6464 0.0612	-	0.7997 0.0228	-	29.8669 1.0056	0.2109 0.0200
Beta	3.8216	0.7180 0.0571	3,062 911.35	0.7589 0.0150	7.3850 2.7276	-	0.1985 0.0180
Sweden (13/03/2020 - 8/05/2020)							
Logistic	8.0904	0.1434 0.0085	3,399 63.97	-	-	-	1.0399 0.1058
Gamma	6.4488	0.4625 0.0549	-	0.8439 0.0228	-	5.8766 0.3452	0.3869 0.0369
Beta	6.5188	0.4635 0.0550	252,661 223,953	0.8432 0.0228	147.3981 131.2944	-	0.3869 0.0369
Ireland (16/03/2020 - 11/05/2020)							
Logistic	7.3545	0.1635 0.0088	2,986 34.26	-	-	-	0.9201 0.0924
Gamma	6.5280	0.3314 0.0568	-	0.9429 0.0347	-	8.6326 0.5825	0.5168 0.0506
Beta	6.4445	0.4701 0.0847	3,920 643.46	0.8459 0.0386	1.7610 0.0684	-	0.4677 0.0458
Portugal (22/03/2020 - 17/05/2020)							
Logistic	7.4439	0.1451 0.0120	1,161 8.72	-	-	-	1.2484 0.1382
Gamma	5.2371	0.8438 0.1213	-	0.7410 0.0335	-	21.2559 1.2594	0.3571 0.0341
Beta	4.4049	1.3262 0.1318	1,271 38.21	0.5900 0.0223	1.0063 0.1155	-	0.2262 0.0214

Table 5: Estimates of the logistic, gamma and beta models at the 8-week stage. Standard errors are reported in small font.

Parameters				Peak mortality rate			Skew	Reproduction rates			Doubling time	
	r	α	$\gamma \times 10000$	Date	Daily	Cumulative		ρ_0 $D = 0.5$	ρ_1 $D = 1.5$	$\rho(peak)$	$D = 0.5$	$D = 1.5$
	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	(ix)	(x)	(xi)	(xii)
Belgium	0.5382	0.9029	3.9311	7 April	236	2,296	5.484	0.575	0.517	0.102	1.204	1.340
UK	0.4682	0.8929	5.4960	8 April	141	1,624	3.878	0.504	0.447	0.087	1.374	1.547
Spain	0.5669	0.9062	5.7420	30 March	181	1,578	3.769	0.605	0.545	0.114	1.145	1.270
Italy	0.4998	0.8707	5.5415	30 March	126	1,571	3.751	0.546	0.473	0.08	1.268	1.462
Sweden	0.4625	0.8438	5.8765	14 April	91	1,436	3.430	0.515	0.433	0.064	1.345	1.597
France	0.2281	1.0468	8.5744	5 April	136	1,220	2.943	0.221	0.232	0.111	3.139	2.984
Ireland	0.3313	0.9428	8.6326	14 April	94	1,092	2.611	0.344	0.323	0.086	2.010	2.142
Netherlands	0.5964	0.8460	8.3911	7 April	89	1,008	2.408	0.663	0.559	0.088	1.044	1.238
US	0.3963	0.8858	9.0801	14 April	72	975	2.329	0.429	0.377	0.074	1.615	1.833
Canada	0.3184	0.8641	10.5390	26 April	44	819	1.957	0.349	0.301	0.053	1.982	2.303
Portugal	0.5786	0.8522	26.1210	21 April	34	326	0.779	0.640	0.543	0.105	1.083	1.277
Germany	0.4031	0.8859	28.8674	7 April	26	306	0.732	0.435	0.383	0.086	1.590	1.808
Denmark	0.6463	0.7996	29.8669	1 April	25	267	0.641	0.741	0.593	0.095	0.934	1.168

Table 6: Summary statistics for gamma drift models.



(a) China (b) Italy

Figure 1: Covid-19 infections in China and Italy.

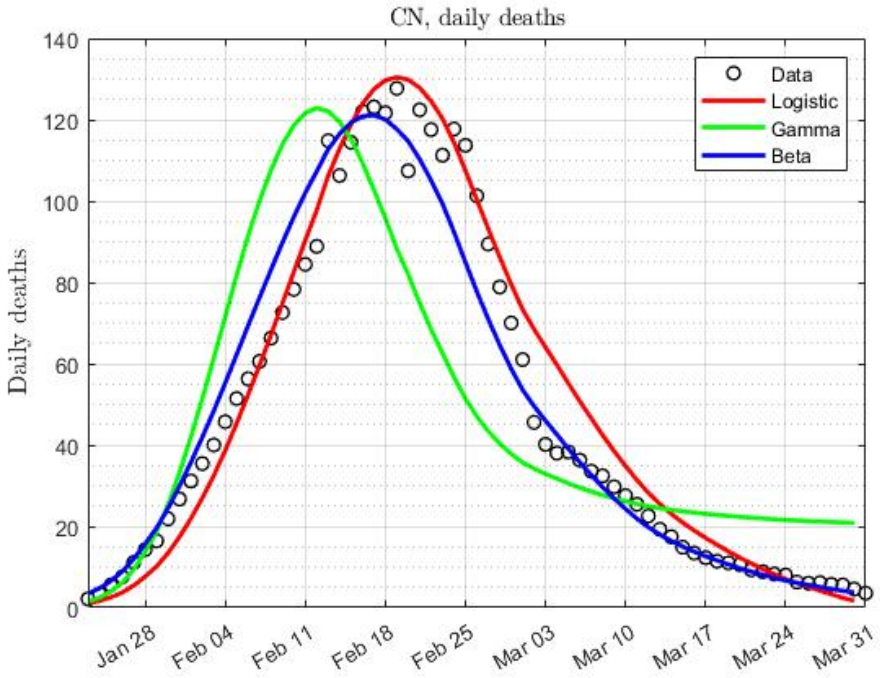
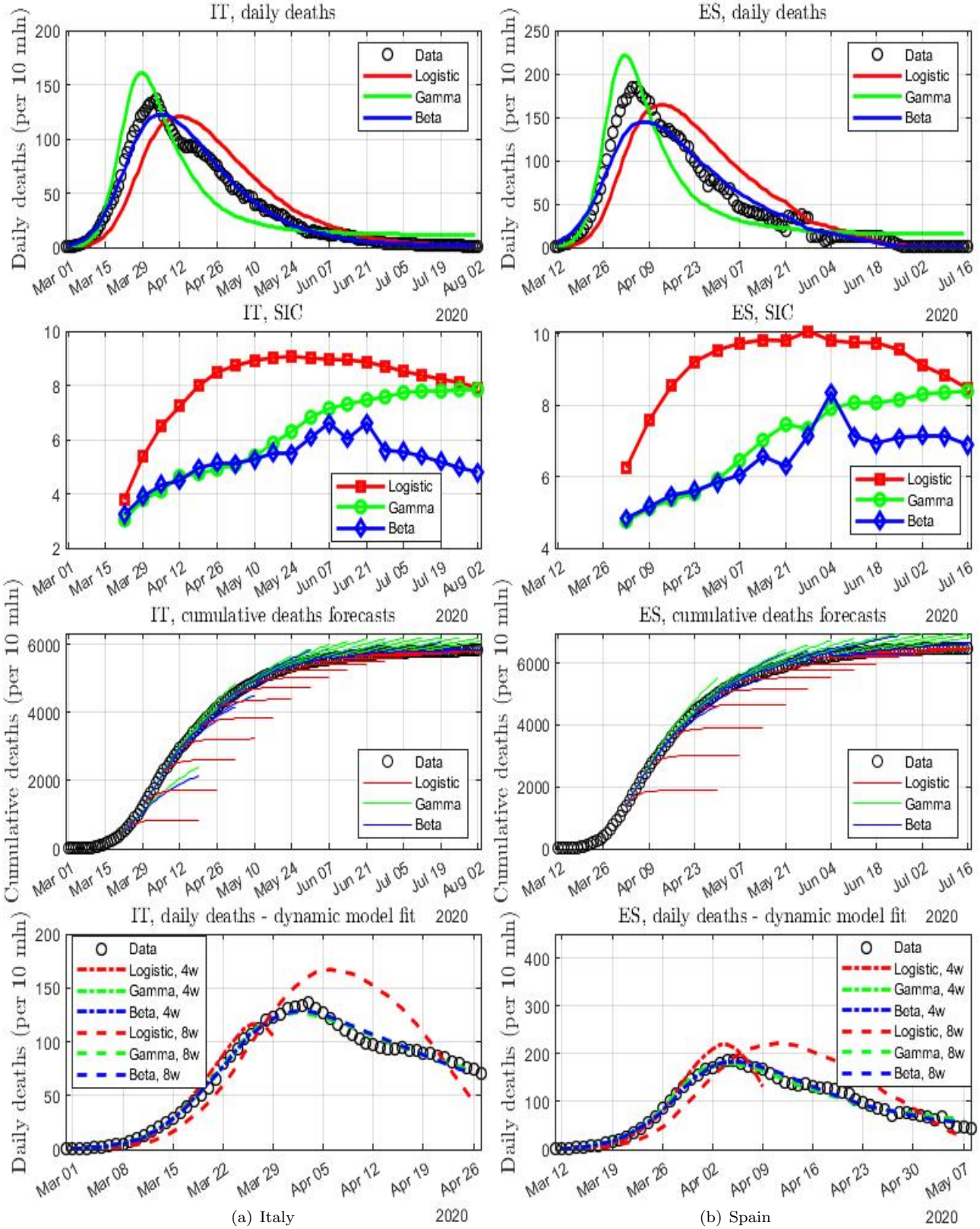
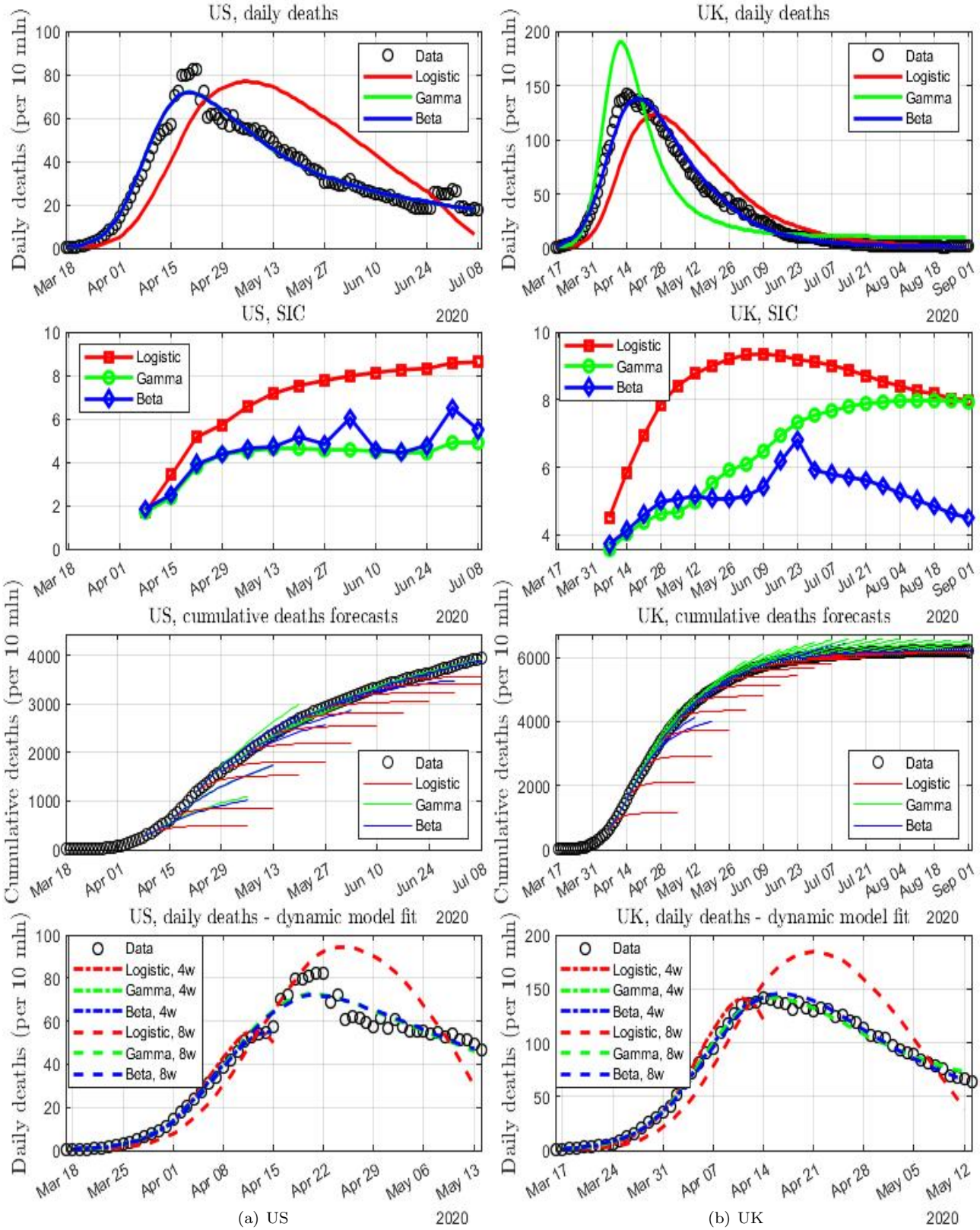


Figure 2: In sample fit of daily deaths for China. The dots show data for the daily death tolls. The red line shows the regression fits for the logistic model, the green line for the gamma model and the blue line for the beta model.



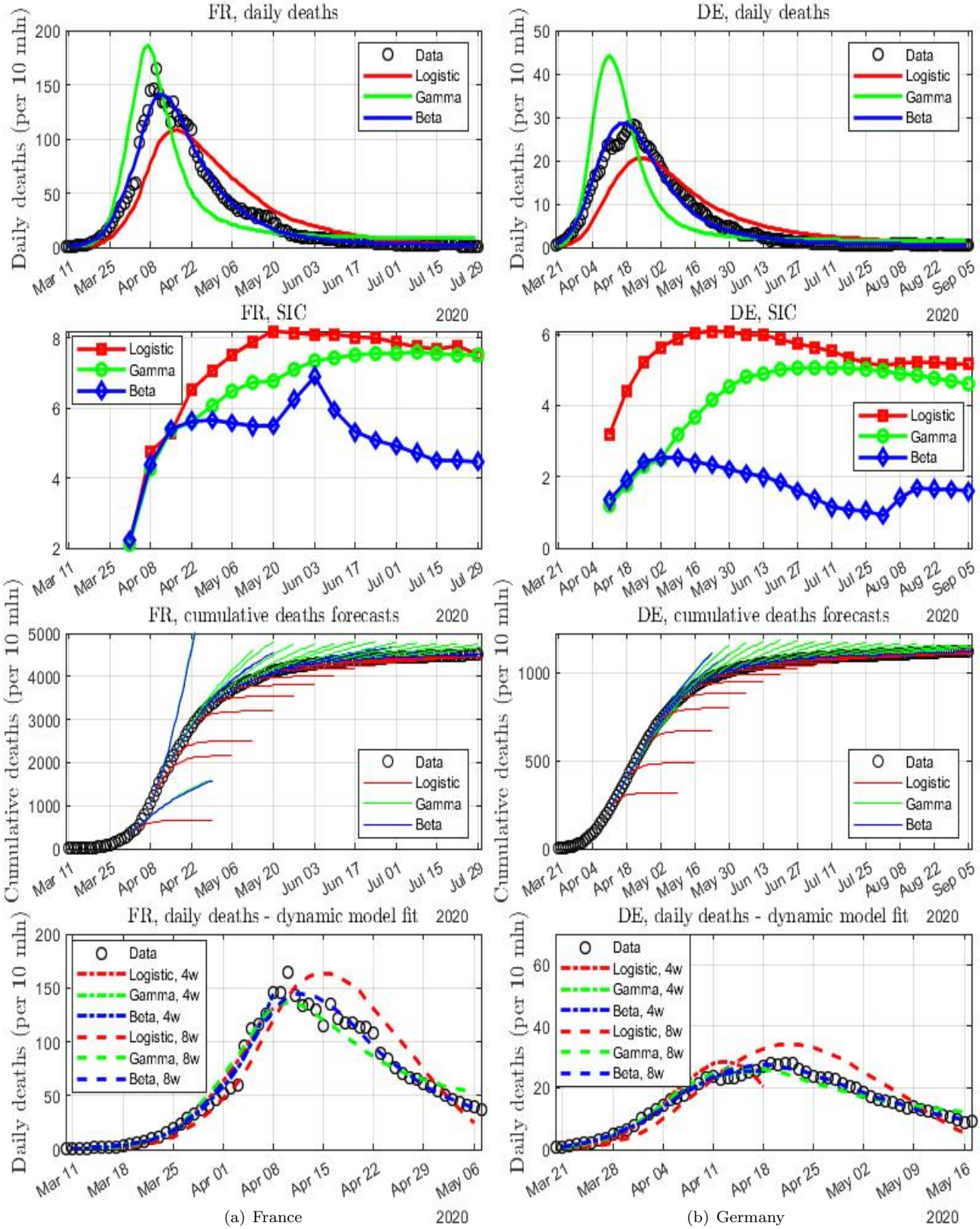
(a) Italy (b) Spain

Figure 3: In-sample fit (top row), SIC and rolling out-of-sample forecasts of cumulative deaths calculated on an expanding window (second and third central rows, respectively) and dynamic fit (bottom row) for Italy (left) and Spain (right).



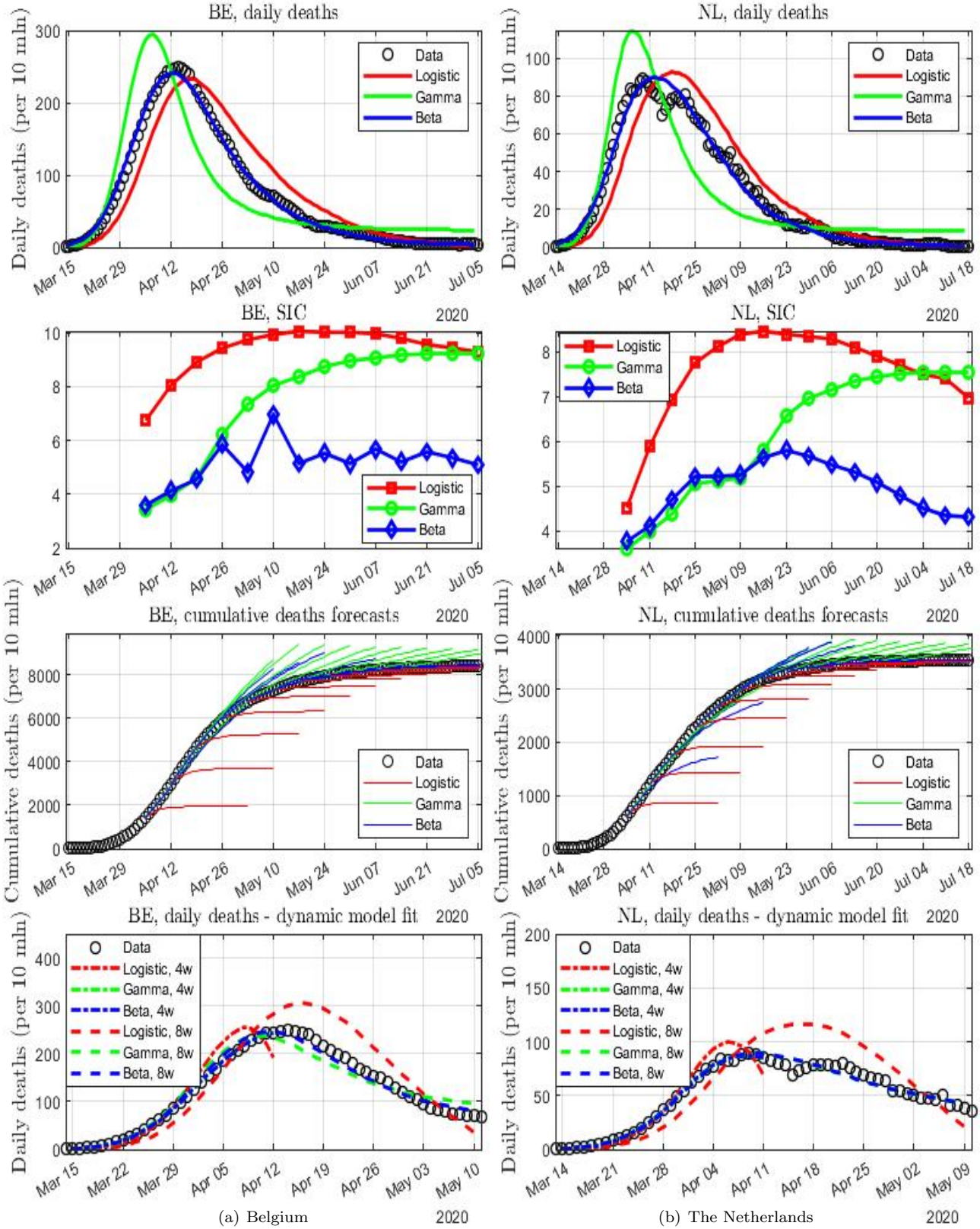
(a) US (b) UK

Figure 4: In-sample fit (top row), SIC and rolling out-of-sample forecasts of cumulative deaths calculated on an expanding window (second and third central rows, respectively) and dynamic fit (bottom row) for the US (left) and the UK (right).



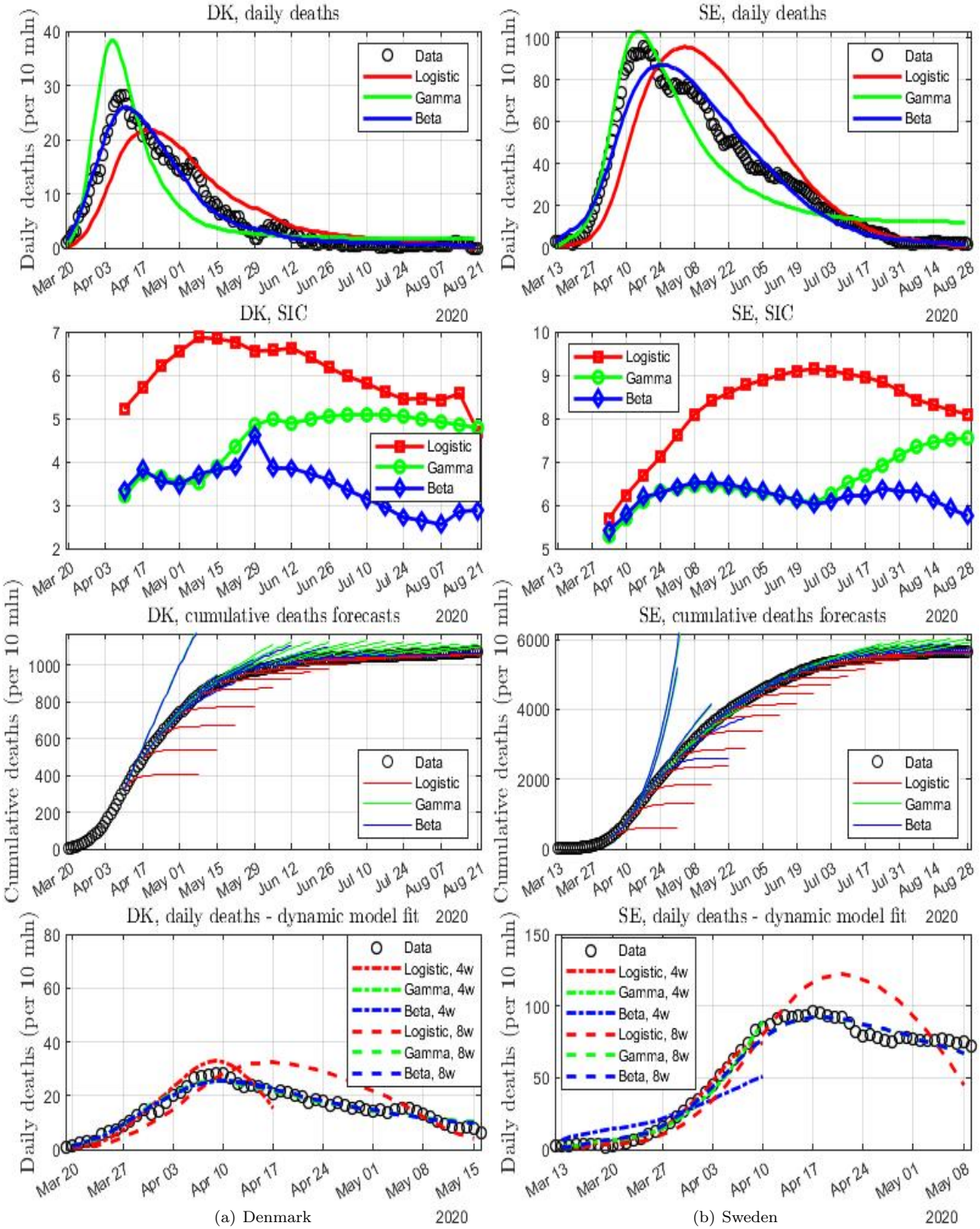
(a) France 2020 (b) Germany 2020

Figure 5: In-sample fit (top row), SIC and rolling out-of-sample forecasts of cumulative deaths calculated on an expanding window (second and third central rows, respectively) and dynamic fit (bottom row) for France (left) and Germany (right).



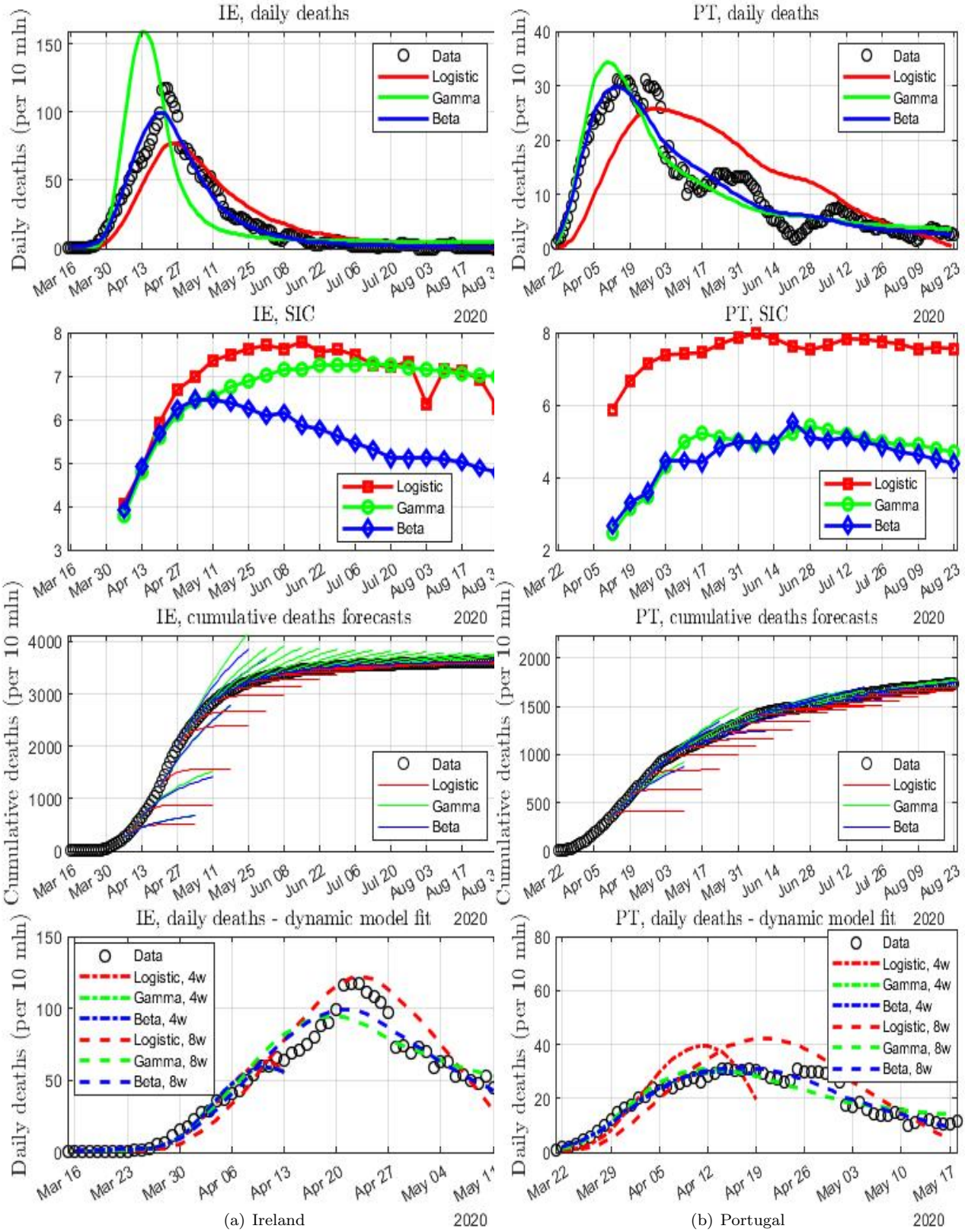
(a) Belgium (b) The Netherlands

Figure 6: In-sample fit (top row), SIC and rolling out-of-sample forecasts of cumulative deaths calculated on an expanding window (second and third central rows, respectively) and dynamic fit (bottom row) for Belgium (left) and the Netherlands (right).



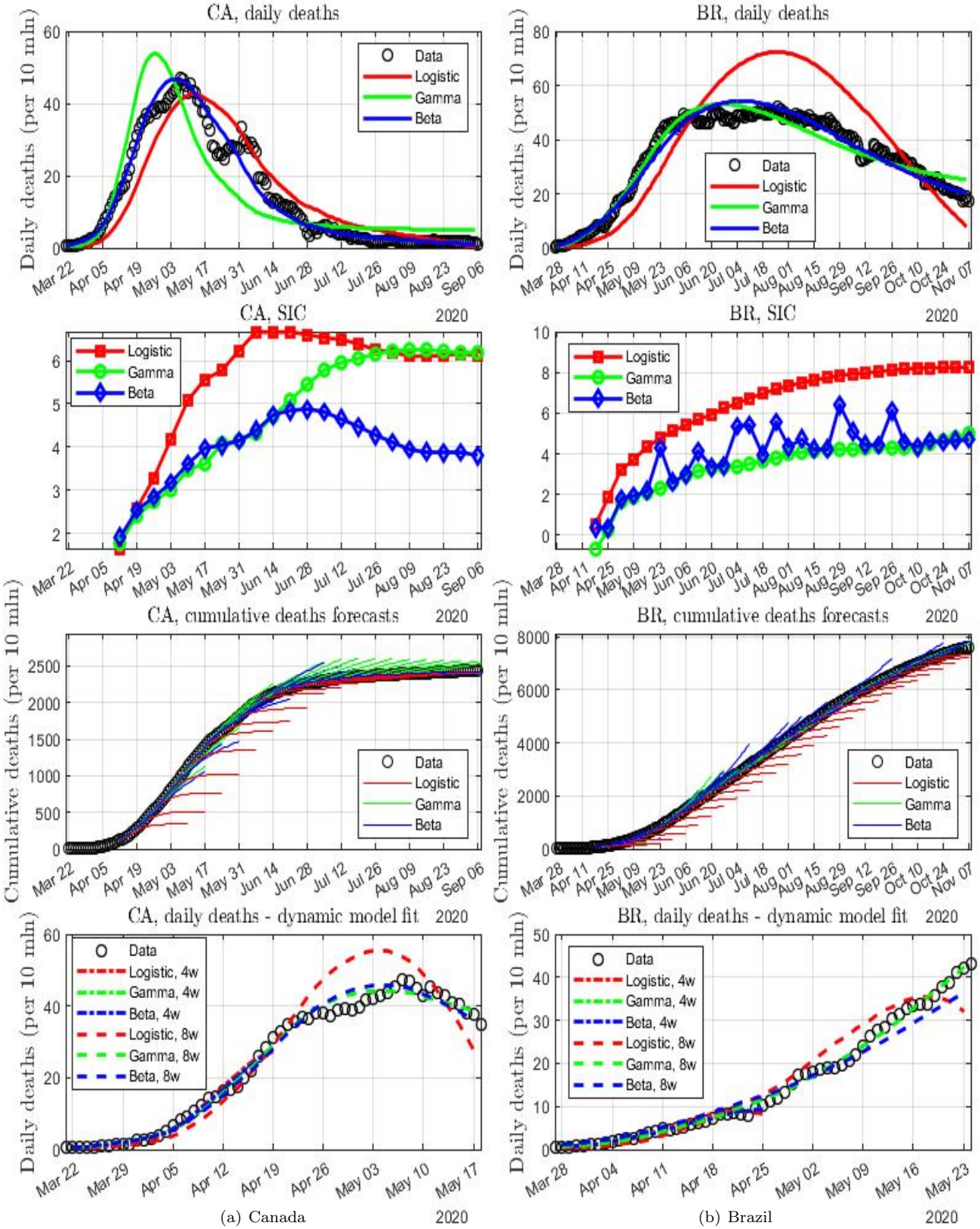
(a) Denmark (b) Sweden

Figure 7: In-sample fit (top row), SIC and rolling out-of-sample forecasts of cumulative deaths calculated on an expanding window (second and third central rows, respectively) and dynamic fit (bottom row) for Denmark (left) and Sweden (right).



(a) Ireland (b) Portugal

Figure 8: In-sample fit (top row), SIC and rolling out-of-sample forecasts of cumulative deaths calculated on an expanding window (second and third central rows, respectively) and dynamic fit (bottom row) for Ireland (left) and Portugal (right).



(a) Canada (b) Brazil

Figure 9: In-sample fit (top row), SIC and rolling out-of-sample forecasts of cumulative deaths calculated on an expanding window (second and third central rows, respectively) and dynamic fit (bottom row) for Canada (left) and Brazil (right).