Abstract. Time series models are developed for predicting future values of a variable which when cumulated is subject to an unknown saturation level. Such models are relevant for many disciplines, but here attention is focused on the spread of epidemics and the applications are for Coronavirus. The time series models are relatively simple but are such that their specification can be assessed by standard statistical test procedures. In the generalized logistic class of models, the logarithm of the growth rate of the cumulative series depends on a time trend. Allowing this trend to be time-varying introduces further flexibility and enables the effects of changes in policy to be tracked and evaluated.

Keywords: Generalized logistic, Gompertz curve, Kalman filter, Negative binomial distribution, Score-driven models, Stochastic trend.

Media summary

As an epidemic takes hold, projections of its trajectory and size enable health care providers and policy makers to plan and organize clinical and human resources to meet treatment requirements. The unknowns surrounding the COVID-19 virus make it particularly important and urgent to have reliable estimates of the course of the incidence of infection – of the daily numbers of new cases and of deaths to be expected.

Epidemiological models that seek to estimate epidemic trajectories fall into two broad classes. Compartmental models try to capture the structures and mechanisms of disease transmission in terms of how the numbers of the susceptible, exposed, infectious and recovered or deceased populations evolve over time. Inevitably when there is uncertainty about many clinical and transmission aspects of the disease, structural models are forced to rely on many assumptions and unknowns. Time series models on the other hand simply use historical data to make predictions.

*p.kattuman@jbs.cam.ac.uk

This article is © 2020 by author(s) as listed above. The article is licensed under a Creative Commons Attribution (CC BY 4.0) International license (https://creativecommons.org/licenses/by/4.0/legalcode), except where otherwise indicated with respect to particular material included in the article. The article should be attributed to the author(s) identified above.
This paper develops a new class of time series models that reflect epidemic trajectories and are able to make good forecasts even before new cases and/or deaths reach their peak. The models are relatively simple and transparent, and their specifications can be assessed by standard statistical test procedures. The nature of epidemic trajectories leads us to formulate a class of models in which the logarithm of the growth rate of the cumulative series follows a downward time trend. Allowing this trend to be time-varying introduces flexibility and enables the effects of changes in policy to be tracked and evaluated. The models are tested with applications to the United Kingdom and Germany and it is shown how they are able to adapt as the response of the population changes over time. Although the focus of the paper is on the spread of epidemics, and specifically COVID-19, the models are applicable in a wide range of disciplines.

1. Introduction

The progress of an epidemic typically starts off with the number of cases following an exponential growth path. Over time the growth rate falls and the total number of cases approaches a final level - the ‘leveling of the curve.’ Complex behavioural modeling of the progress of the disease, for example, by using the ‘semi-mechanistic Bayesian hierarchical model’ implemented by the team at Imperial College London, depends on many assumptions and unknowns; see Flaxman et al. (2020). Simple and transparent time series models may offer an alternative way of making predictions of the trajectory of the epidemic; see, for example, Chowell et al. (2016, section 2), where such approaches are called ‘phenomenological’. Similar issues arise in economics where there is a contrast between calibrated and Bayesian models based on economic theory and data-based time series models. Avery et al. (2020) give an excellent review of these issues.

The progress towards an upper bound or saturation level can be taken on board with a sigmoid curve, such as the logistic

\[ \mu(t) = \bar{\mu} / (1 + \gamma_0 e^{-\gamma t}), \quad \gamma_0, \gamma, \bar{\mu} > 0, \quad -\infty < t < \infty, \]

where \( \bar{\mu} \) is the final level, \( \gamma \) is a rate of progress parameter, \( \gamma_0 \) takes account of the initial conditions and \( e \) is Euler’s number. Logistic curves can be shown to arise from a model of a simple epidemic; see, for example, Daley and Gani (1999, ch. 2). More generally, sigmoid curves are used in many disciplines for a variety of applications, such as estimating the demand for new products and population growth of mammals subject to space and resource limitations. An early and influential discussion of growth curves was given by Gregg et al. (1964). More recently Panik (2014) describes growth curves and statistical methods for fitting them. Here we concentrate on an extension of the logistic curve called the generalized logistic (GL) or Richards curve; see Panik (2014, pp. 78–80). The Gompertz curve emerges as an important special case. Figure 1 shows two Gompertz growth curves together with the corresponding changes as given by the first derivatives. The effect of what turns out to be a key parameter, \( \gamma \), on the peak is evident.

By formulating a statistical model, parameters such as \( \gamma_0, \bar{\mu} \) and \( \gamma \) can be estimated from observations, denoted \( Y_t \), made at discrete times \( t = 1, \ldots, T \), on \( \mu(t) \). One option is to work directly with the level, by basing a time series model on a deterministic trend, as in (1.1); for example, Meade and Islam (1995) report fitting a variety of growth curves to telecommunications data, while Ciufolini and Paolozzi (2020) report fitting a deterministic trend to total Coronavirus cases in Italy. The methods discussed in Panik (2014, ch. 4) are restricted to deterministic curves, possibly with
a first-order autoregressive disturbance term. Our view is that, as with most economic and social time series, a deterministic trend fitted to the level is too inflexible. Instead we prefer to work with the change or the growth rate, with specification of the model informed by an assumption that the total follows a growth curve. The saturation level can be continually updated as new observations become available. A second advantage of working with the difference or the growth rate is that when logarithms are taken\footnote{One implication is that the forecasts of the change will tend asymptotically towards zero and never become negative. In some applications, $Y_t$ can go down as well as up and in these situations the methods proposed here do not apply. Instead flexibility can be brought into the deterministic growth curve by allowing parameters, such as $\mu$, to change over time. Young and Ord (1989) effectively do this by discounted least squares.}, estimation of the basic models derived from the GL class can be carried out by least squares regression, as proposed in Harvey (1984). Thus the method is viable even for a small number of observations. Finally the deterministic time trend in the estimating equations can be replaced by a stochastic one. This is particularly effective with a Gompertz function. The flexible trend, which can be estimated by the Kalman filter, as in the STAMP package of Koopman et al. (2020), allows parameters to evolve over time. As a result the model can adapt to significant events and changes in policy. Figure 2 shows predictions made for new cases of Coronavirus in Germany. The predictions, which include a day of the week effect, were made using data up to the end of March and show a downward trend even though the series was only just reaching what subsequently turned out to be its peak. Moving beyond the peak signals that $R_t$, the effective reproduction number at time $t$, has dipped below one; see Flaxman et al. (2020) and Aronson et al. (2020).

When numbers are small, as is the case with deaths at the beginning or end of an epidemic, there is a strong argument for adopting a negative binomial distribution. Models formulated under

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{Figure1.png}
\caption{Gompertz functions with final level of 100. The solid red curve (bold) has a point of inflexion equal to 30, with $\gamma_0 = 20$ and $\gamma = 0.1$. The solid blue curve, to the left, has a point of inflexion equal to 20, with $\gamma_0 = 20$ and $\gamma = 0.15$. The dotted lines show the corresponding change curves multiplied by 10.}
\end{figure}
Figure 2. Forecasts (dashed - blue) for new cases in Germany using data (solid red) up to March 31st. There was no data reported for May 1st. Data for May 2nd can be regarded as the sum of observations for May 1st and 2nd.

an assumption of Gaussianity need to be modified accordingly and we show how this may be done using the score-driven approach described in Harvey (2013) and implemented in the TSL computer package of Lit et al. (2020).

Section 2 describes the GL class of growth curves, while Section 3 discusses estimation. These methods are then applied in Section 4 to data on Coronavirus in UK hospitals and in Germany. Forecasts are set out and evaluated. The effects of policy, primarily the lockdown imposed on March 21st in the UK, are assessed in Section 5. There is concern about a potential second wave of infections as restrictions start to be eased and ways in which this might be monitored are discussed in Section 6. Section 7 concludes.

2. Growth curves

Let $\mu(t) \geq 0$ be a monotonically increasing function defined over the real line. The rate of change or ‘incidence curve’ is $d\mu(t)/dt \geq 0$. The generalized logistic is

$$
\mu(t) = \mu / (1 + (\gamma_0/\kappa)e^{-\gamma t})^\kappa, \quad \gamma_0, \gamma, \kappa > 0,
$$

where $\gamma$ is a growth rate parameter. The parameter $\kappa$ must be positive for there to be an upper asymptote; allowing $\kappa$ to be negative gives the class of general modified exponential (GME) growth curves. The logistic is obtained by setting $\kappa = 1$, while letting $\kappa \to \infty$ yields the Gompertz curve. When $\gamma_0$ is determined by the value of the curve at $t = 0$, it is

$$
\gamma_0 = \kappa \left[ (\mu / \mu(0))^{1/\kappa} - 1 \right].
$$

Differentiation yields
(2.3) \[ \ln d\mu(t)/dt = \rho \ln \mu(t) + \delta - \gamma t, \]

where \( \delta = \ln(\gamma_0/\bar{\mu}) \) and \( \rho = (\kappa + 1)/\kappa \), so \( 0 < \kappa < \infty \) implies \( 1 < \rho < \infty \). Alternatively, because \( d\mu(t)/dt = g(t)\mu(t) \),

(2.4) \[ \ln g(t) = (\rho - 1) \ln \mu(t) + \delta - \gamma t, \]

where \( g(t) \) is the growth rate of \( \mu(t) \). Note that \( \rho - 1 = 1/\kappa \).

The generalized logistic differential equation implied by (2.1) is

(2.5) \[ \frac{d\mu(t)}{dt} = \gamma \kappa \left[ 1 - \left( \frac{\mu(t)}{\bar{\mu}} \right)^{1/\kappa} \right] \mu(t). \]

The term in square brackets is less than one and tends to zero as \( \mu(t) \to \bar{\mu} \). The growth rate implied by (2.5) is

(2.6) \[ g(t) = \gamma \kappa \left[ 1 - \left( \frac{\mu(t)}{\bar{\mu}} \right)^{1/\kappa} \right]. \]

When \( \kappa = 1 \), (2.5) is a Riccati equation.

2.1. **Where is the peak?** The point of inflexion on the growth curve is the point at which the number of new cases, \( d\mu(t)/dt \), peaks. Differentiating the logarithm of \( d\mu(t)/dt = g(t)\mu(t) \) yields the condition

(2.7) \[ g(t) = -g_g(t), \]

where \( g_g(t) \) is the growth rate of the growth rate. It follows from (2.4) that the point of inflexion in the GL occurs when \( g(t) = \gamma/\rho \). The corresponding value of \( \mu(t) \) is

(2.8) \[ \mu^* = \bar{\mu} \rho^{-\kappa} = \bar{\mu}(\kappa/(\kappa + 1))^\kappa. \]

The change declines more slowly than it ascends when \( \kappa > 1 \); see, for example, Figure 1. When \( \gamma_0 \) is determined by \( \mu(0) \), as in (2.2), the peak is at

(2.9) \[ t^* = \ln \gamma_0/\gamma, \quad \gamma_0 > 1, \]

so it comes forward as \( \gamma \) increases.

2.2. **The Gompertz curve.** The Gompertz curve is

(2.10) \[ \mu(t) = \bar{\mu} \exp(-\gamma_0 e^{-\gamma t}), \quad \gamma_0, \gamma > 0, \quad -\infty < t < \infty. \]

When the GL is parametrized as in (2.1), letting \( \kappa \to \infty \) gives (2.10). Taking the logarithm of \( \mu(t) \) gives

(2.11) \[ \ln \mu(t) = \ln \bar{\mu} - \gamma_0 e^{-\gamma t} \]
leading to (2.4) with $\rho = 1$, that is,

\[(2.12) \quad \ln g(t) = \delta - \gamma t,\]

where $\delta = \ln(\gamma_0 \gamma)$. The point of inflexion is given in (2.9), which corresponds to $\mu^* = \overline{\mu}/e \approx 0.368\overline{\mu}$. The maximum value of the change is $0.368\gamma \mu$, while $g(t) = \gamma$. Letting $\kappa \to \infty$ in (2.8), so that $\rho \to 1$, also gives $\mu^* = \mu/e$.

2.3. Statistical distributions and epidemics. Writing a GL growth curve as $\mu(t) = \overline{\mu} F(t)$ allows $F(t)$ to be interpreted as the cumulative distribution function (CDF) of the log of a Dagum distribution; see Kleiber and Kotz (2003, pp. 212–213). Hence the corresponding probability density function (PDF), $f(t)$, is a special case of an Exponential Beta of the Second Kind (EGB2) distribution; see McDonald and Xu (1995).

The Gompertz distribution written using (2.10) has $f(t) = \gamma \gamma_0 F(t) \exp(-\gamma t)$. Figure 1 shows two Gompertz PDFs with the red bold dotted curve corresponding to $\gamma_0 = 20$ and $\gamma = 0.1$ and the blue dotted curve corresponding to $\gamma_0 = 20$ and $\gamma = 0.15$. The effect of increasing $\gamma$ is to raise the peak and bring it forward. Since $d\mu(t)/dt = \overline{\mu} f(t)$ the peak is brought down by a lower $\mu$.

When estimating a deterministic growth curve, some researchers, for example Ciufolini and Paolozzi (2020), prefer to use a sigmoid defined in terms of the Gaussian error function, which is the CDF of a Gaussian variate. Similarly, Murray (2020) use the Gaussian error function to model the logarithm of total Coronavirus cases in US states. It is difficult to see why the Gaussian error function, which has no closed form, might be preferred to the more flexible GL curve.

It follows from (2.5) that

\[(2.13) \quad d\mu(t)/dt = \overline{\mu} F(t)(1 - F(t)^{1/\kappa}),\]

where $\mu(t) = \overline{\mu} F(t)$. In a simple epidemic, $d\mu(t)/dt$ is proportional to a logistic growth curve, $F(t)(1 - F(t))$. Allowing $\kappa$ to be other than one gives more flexibility and is a useful generalization if the model provides a good fit to the data. Indeed complex mechanistic models of epidemics, with the population assigned to compartments labeled Susceptible, Infected, and Recovered (SIR), often produce incidence curves that are positively skewed. An example is the model of Giordano et al. (2020) which is based on a system of eight differential equations.

3. Statistical modeling

In the observational model, the cumulative total at time $t - 1$ replaces $\mu(t)$ and the (positive) change, $y_t = \Delta Y_t = Y_t - Y_{t-1}$, replaces $d\mu(t)/dt$. Formula (2.3) leads to

\[(3.1) \quad \ln y_t = \rho \ln Y_{t-1} + \delta - \gamma t + \varepsilon_t, \quad \rho \geq 1, \quad t = 2, \ldots, T,\]

where the disturbance terms $\varepsilon_t$ are assumed to be independently and identically distributed with mean zero and constant variance, $\sigma_\varepsilon^2$, that is, $\varepsilon_t \sim \text{IID}(0, \sigma_\varepsilon^2)$. Subtracting $\ln Y_{t-1}$ from both sides gives the form corresponding to (2.4), namely

\[(3.2) \quad \ln g_t = (\rho - 1) \ln Y_{t-1} + \delta - \gamma t + \varepsilon_t,\]
where \( g_t = y_t / Y_{t-1} \), although it may also be defined as \( \Delta \ln Y_t \). The parameters \( \rho, \delta \) and \( \gamma \) can therefore be estimated by regression. If \( \rho \) takes a specified value, \( \ln y_t - \rho \ln Y_{t-1} \) is simply regressed on a constant and time trend. The estimators of \( \delta \) and \( \gamma \) are then efficient when the disturbance is Gaussian, that is, \( \varepsilon_t \sim \text{NID}(0, \sigma^2) \).

The observational model for the Gompertz curve is

\[
\ln y_t = \ln Y_{t-1} + \delta - \gamma t + \varepsilon_t, \quad t = 2, ..., T,
\]

or the simple time trend regression

\[
\ln g_t = \delta - \gamma t + \varepsilon_t, \quad t = 2, ..., T.
\]

### 3.1. A time-varying trend: the dynamic Gompertz model

A stochastic trend may be introduced into (3.4) to give the dynamic Gompertz model

\[
\ln g_t = \delta_t + \varepsilon_t, \quad \varepsilon_t \sim \text{NID}(0, \sigma^2_e), \quad t = 2, ..., T,
\]

where

\[
\delta_t = \delta_{t-1} - \gamma_t - \eta_t, \quad \eta_t \sim \text{NID}(0, \sigma^2_\eta),
\]

\[
\gamma_t = \gamma_{t-1} + \zeta_t, \quad \zeta_t \sim \text{NID}(0, \sigma^2_\zeta),
\]

and the normally distributed irregular, level and slope disturbances, \( \varepsilon_t, \eta_t \) and \( \zeta_t \), respectively, are mutually independent. When \( \sigma^2_\eta = \sigma^2_\zeta = 0 \), the trend is deterministic, that is, \( \delta_t = \delta - \gamma t \) with \( \delta = \delta_0 \). On the one hand, when only \( \sigma^2_\zeta \) is zero, the slope is fixed and the trend reduces to a random walk with drift. On the other hand, allowing \( \sigma^2_\zeta \) to be positive, but setting \( \sigma^2_\eta \) to be zero gives an integrated random walk (IRW) trend, which when estimated tends to be relatively smooth. The degree of smoothness depends on the signal-noise ratio, \( q = \sigma^2_\zeta / \sigma^2_\varepsilon \).

The STAMP package of Koopman et al. (2020) can estimate a stochastic trend using techniques based on state space models and the Kalman filter. The Kalman filter outputs estimates of the state vector \((\delta_t, \gamma_t)'\). Estimates of the state at time \( t \) conditional on information up to and including time \( t \) are denoted \((\delta_{t|t}, \gamma_{t|t})'\) and given by the contemporaneous filter; the predictive filter, which outputs \((\delta_{t+1|t}, \gamma_{t+1|t})'\), estimates the state at time \( t + 1 \) from the same information set. It may sometimes be useful to review past movements by the smoother, which is the estimate of the state at time \( t \) based on all \( T \) observations in the series. Estimation of the unknown variance parameters is by maximum likelihood (ML). Tests for normality and residual serial correlation are based on the standardized innovations, that is, one-step ahead prediction errors, \( \hat{v}_t = y_t - \hat{\delta}_{t|-1} \), \( t = 3, ..., T \).

A stochastic trend can be introduced into the more general GL model. However, unless \( \rho \) is fixed, it may be hard to estimate in small samples.

The Kalman filter can be by-passed by adopting the reduced form, which comes from the innovations form of the Kalman filter; see Harvey (2013, pp. 3, 71). For the GL curve, the steady-state innovations form is

\[
\ln y_t = \rho \ln Y_{t-1} + \delta_{t|-1} + \hat{v}_t, \quad t = 3, ..., T,
\]
\[(3.7) \quad \ln g_t = (\rho - 1) \ln Y_{t-1} + \delta_{t+1} + \nu_t, \quad t = 3, \ldots, T, \]

where

\[
\begin{align*}
\delta_{t+1} &= \delta_{t+1} - \gamma_{t+1} + \alpha_1 \nu_t, \\
\gamma_{t+1} &= \gamma_{t+1} + \alpha_2 \nu_t. 
\end{align*}
\]

The filter for an IRW has \( \alpha_2 = \alpha_1^2/(2 - \alpha_1) \) with \( 0 \leq \alpha_1 \leq 1; \) see Harvey (2013, pp. 78–79). The implied value of \( q_\zeta \) is

\[(3.8) \quad q_\zeta = \frac{\alpha_1^4}{(2 - \alpha_1)^2(1 - \alpha_1)}. \]

The parameter \( \alpha_1 \) thus plays a role similar to that of the signal-noise ratio, \( q_\zeta \), and can be estimated by ML. When it is zero, the model reverts to (3.2).

### 3.2. Forecasts

Forecasts of future observations and an estimate of the final level can be obtained from the predictive recursions

\[(3.9) \quad \hat{g}_{T+\ell|T} = \tilde{\mu}_{T+\ell-1|T}^\rho \exp(\delta_{T}) \exp(-\gamma_{\ell}), \]
\[(3.10) \quad \tilde{\mu}_{T+\ell|T} = \tilde{\mu}_{T+\ell-1|T} + \hat{y}_{T+\ell|T}, \quad \ell = 1, 2, \ldots \]

where \( \delta_T \) is the level at time \( T \), that is, \( \delta_T = \delta - \gamma T \) in (3.1), and \( \hat{y}_{T|T} = Y_T \). As \( \ell \to \infty \), \( \tilde{\mu}_{T+\ell|T} \to \pi \).

Alternatively,

\[(3.11) \quad \hat{g}_{T+\ell|T} = \tilde{\mu}_{T+\ell-1|T}^{-1} \exp(\delta_T - \gamma_{\ell}), \quad \ell = 1, 2, \ldots, \]
\[(3.12) \quad \tilde{\mu}_{T+\ell|T} = \tilde{\mu}_{T+\ell-1|T}(1 + \hat{y}_{T+\ell|T}) \]

so that \( \hat{g}_{T+\ell|T} = \hat{y}_{T+\ell|T} \tilde{\mu}_{T+\ell-1|T} \) and \( \hat{y}_{T+\ell|T} = \tilde{\mu}_{T+\ell|T} \).

For the Gompertz model, where \( \rho = 1 \), the forecasts for the growth rate are simply

\[(3.13) \quad \hat{g}_{T+\ell|T} = \exp(\delta_T) \exp(-\gamma_{\ell}), \]

so (3.11) and (3.12) yield

\[(3.14) \quad \tilde{\mu}_{T+\ell|T} = \tilde{\mu}_{T|T} \prod_{j=1}^{\ell}(1 + \exp(\delta_T) \exp(-\gamma_j)). \]

A future point of inflexion is given at \( \ell^* = (\delta_T - \ln \gamma)/\gamma \). When \( 2 \ell \to \infty \), \( \tilde{\mu}_{T+\ell|T} \simeq Y_T \exp((\exp \delta_T)/(\exp \gamma - 1)) \).

An assumption of normality for the disturbances in (3.1) implies that, conditional on information at time \( t - 1 \), \( y_t \) is log-normal. Thus there may be a case for adding \( 0.5\sigma^2 \) to \( \delta_T \), where \( \sigma^2 \) is the

\[2 \ln \left[ \prod_{j=1}^{\ell} \left( 1 + g_{T+\ell} \right) \right] = \ln \sum_{j=1}^{\ell} \left( 1 + g_{T+\ell} \right) \simeq \sum_{j=1}^{\ell} g_{T+\ell}. \] Higher order terms can be neglected when \( g_{T+\ell} \) is small. Then \( \sum_{j=1}^{\ell} g_{T+\ell} = \delta_T \sum_{j=1}^{\ell} \exp(-\gamma_{\ell}) \to \delta_T/(\exp \gamma - 1) \) as \( \ell \to \infty \).
prediction error variance. (When there is no stochastic trend, $\sigma^2 = \sigma^2_\varepsilon$.) Predictive distributions of future observations may be obtained by simulation.

### 3.3. Models for the growth rate

A stochastic model for the logistic curve can be based on (2.6), as in Levenbach and Reuter (1976), by adding a serially independent Gaussian disturbance term so that

\[(3.15) \quad g_t = \gamma - (\gamma / \bar{\mu}) \hat{\mu}_{t|t-1} + \varepsilon_t, \quad t = 1, ..., T,\]

where $\hat{\mu}_{t|t-1}$ is an estimate of $\mu_t$, such as $Y_{t-1}$, based on information at time $t - 1$ and $\varepsilon_t$ is a serially independent Gaussian disturbance with mean zero and constant variance, $\sigma^2_\varepsilon$, that is $\varepsilon_t \sim \text{NID}(0, \sigma^2_\varepsilon)$. Regressing $g_t$ on $Y_{t-1}$ gives estimates of the key parameters $\gamma$ and $\bar{\mu}$. Generalization of this approach is based on (2.6) but leads to a nonlinear equation

\[(3.16) \quad g_t = \gamma \kappa - \left(\frac{Y_{t-1}}{\bar{\mu}}\right)^{1/\kappa} + \varepsilon_t, \quad t = 1, ..., T,\]

which requires a search over the range of $\kappa$ if it is unknown. In the Gompertz case, estimation is as straightforward as for the logistic model because

\[(3.17) \quad g_t = \gamma \ln \bar{\mu} - \gamma \ln Y_{t-1} + \varepsilon_t, \quad t = 1, ..., T.\]

Models based on the logarithm of the growth rate are preferred because they have better statistical properties. For example, the disturbance term is less likely to be heteroscedastic.

### 3.4. Small numbers: the negative binomial distribution

When $y_t$ is small, it may be better to specify its distribution, conditional on past values, as discrete. The usual choice is the negative binomial which, when parameterized in terms of a time-varying mean, $\xi_{t|t-1}$, and a fixed positive shape parameter, $\upsilon$, has probability mass function (PMF)

\[p(y_t|y_{t-1}, y_{t-2}, ...) = \frac{\Gamma(v + y_t)}{y_t!\Gamma(v)} \xi_{t|t-1}^{y_t} (v + \xi_{t|t-1})^{-y_t} (1 + \xi_{t|t-1}/v)^{-v}, \quad y_t = 0, 1, 2, ...,\]

with $\text{Var}_{t-1}(y_t) = \xi_{t|t-1} + \xi_{t|t-1}^2 / v$. An exponential link function ensures that $\xi_{t|t-1}$ remains positive and at the same time yields an equation similar to (3.1):

\[(3.18) \quad \ln \xi_{t|t-1} = \rho \ln Y_{t-1} + \delta - \gamma t, \quad \rho \geq 1, \quad t = 3, ..., T.\]

A stochastic trend may be introduced into the model, as in sub-section 3.2, by developing a filter for the time-varying trend similar in structure to that of (3.6). Because the observations are not Gaussian, the dynamic conditional score (DCS) framework described in Harvey (2013, pp. 77–79) is used to give

\[(3.19) \quad \ln \xi_{t|t-1} = \rho \ln Y_{t-1} + \delta_{t|t-1}, \quad t = 3, ..., T,\]

where $\delta_{t|t-1}$ is like the filter for the IRW in the Gaussian model, that is,

\[\text{This sub-section can be omitted without loss of continuity.}\]
\[
\begin{align*}
\delta_{t+1,t} &= \delta_{t,t-1} - \gamma_{t,t-1} + \alpha_1 u_t, \quad 0 \leq \alpha_1 \leq 1, \\
\gamma_{t+1,t} &= \gamma_{t,t-1} + \alpha_2 u_t, \quad \alpha_2 = \alpha_1^2 / (2 - \alpha_1),
\end{align*}
\]

but with \(u_t = y_t / \xi_{t,t-1} - 1\), which is the conditional score for \(\ln \xi_{t,t-1}\), that is \(u(y_t - \xi_{t,t-1}) / (v + \xi_{t,t-1})\), divided by the information quantity. The dynamic Gompertz model has \(\rho = 1\). Estimation is by maximum likelihood.

Predictions of future observations and the final level can be obtained from (3.9) and (3.10).

4. Forecasting Coronavirus in the UK and Germany

We began working on this project at the beginning of April. At that time Coronavirus was not as far advanced in the UK as in Italy, and our initial exploration was focussed on Italy. Figure 3 shows new cases, as measured by hospital admissions on the European Centre for Disease Prevention and Control (ECDC) website\(^4\), in the UK and Italy. Italy is seen to be clearly ahead and it is apparent that the decline is slower than the rise. This asymmetry is not consistent with a simple logistic model, (1.1), and later analysis confirmed this to be the case for most other countries, including the UK.

\[\text{Figure 3. New cases in the UK (red, bold) and Italy (blue) up to April 10th.}\]

Economic and social time series are typically subject to periodic variation, due to seasonal, day of the week and other effects. Preliminary analysis of data for hospital admissions and deaths in Italy indicate a day of the week pattern and this is confirmed for the UK; one reason is that laboratory confirmation for the virus tends to slow down during weekends. Day of the week effects are initially included in the equations by the introduction of a single harmonic cycle with period seven; this possesses only two parameters as opposed to the six needed for a full set of dummy variables.

\(^4\)The data on COVID-19 that we use to illustrate the time series models presented in this paper are discussed in Appendix A, which also discusses the limitations of COVID-19 data more generally.
Table 1. Estimates for UK Coronavirus cases (hospital admissions) as of April 13th and, in the last column*, as of April 20th. Fixed parameters are in italics. The maximized log-likelihood is ln \( L \); note that the last column is based on a different set of data and so its likelihood value should not be compared to those in the other columns.

Although we estimated many models in early April, the first results reported here for the UK are those obtained on April 13th, that is, with data up to the 12th, starting on March 5th. The models were updated on April 20th and 27th. Unfortunately, the data available to us after April 29th were not consistent with the earlier data as they were not confined to just tests carried out on hospital admissions.

4.1. Models fitted to new cases in the UK. Table 1 shows April 13th estimates for the GL, (3.2), the Gompertz, where \( \rho \) is set to one, and the dynamic Gompertz where the trend is an IRW. The results for the logistic model are not reported because it gives an inferior fit and is rejected by a ‘t-test’ on \( \hat{\rho} \); the standard error is in parentheses. The diagnostics presented are the Durbin-Watson (DW) statistic for residual serial correlation, a portmanteau \( Q \) statistic for serial correlation based on \( P \) autocorrelations, the Bowman-Shenton normality test statistic and a heteroscedasticity statistic consisting of the squares of the last third of the residuals over the first third.

There is a substantive reduction in the prediction error variance for all models when daily effects are included. The peak day is Friday, the same day as was found for Italy. The results for the GL are shown in the penultimate column and may be compared with those in the first column. The likelihood ratio statistic is 8.48 which is significant at the 5% level of a \( \chi^2 \) distribution. Figure 4 shows the fit and residuals from the GL. Figure 5 shows the histogram and correlogram for the GL residuals but using data up to April 20th.

4.2. Forecasts and forecast evaluation. The forecasts of new cases made on April 13th are shown in Figure 6, together with the actual values up to May 29th. The upper line is the dynamic Gompertz, while the lower lines are the GL with and without the daily effect. The actual outcome turns out to lie between the dynamic Gompertz and the GL. As with the German predictions shown in Figure 2, the important point is that for both models the forecasts are moving downwards even though the observations have barely peaked.
Predictions from the GL made on April 20th show little change from those made on April 13th. The final level under current policies is estimated to be 186,000. By contrast the dynamic Gompertz predictions change significantly. They are still higher than the GL predictions but the final level predicted (by the dynamic Gompertz model) is now 237,790, whereas on April 13th it was 308,960. The dynamic Gompertz predictions made on April 27th show little movement; the final predicted level is 253,800. Figure 7 shows the dynamic Gompertz predictions without the daily component, but the daily component was included when the models were estimated. The flexibility of the
Forecasts of new cases: GL model
New cases (data till April 13th)
Forecasts of new cases: Dynamic Gompertz model
New cases (data April 14th to 29th)

**Figure 6.** Forecasts (dashed) for new cases in the UK (using data up to April 13th).

A dynamic model resides in its ability to adapt to a situation in which the observations are falling less rapidly than indicated by the static GL.

**Figure 7.** Forecasts (dashed) for new cases in the UK made by Dynamic Gompertz model (using data up to April 20th, and up to April 27th).

4.3. Germany. The data for new cases in Germany are from the Robert Koch Institut (RKI) and constitute the confirmed cases of COVID-19 in all national hospitals and testing centres.\(^5\)

Fitting with data for all of March give the results in Table 2 for a dynamic Gompertz model with daily effects. Estimation of a GL give \(\hat{\rho} = 1.21(0.33)\), so the \(t\)-statistic is only 0.64. The

\(^5\)https://www.rki.de/DE/Home/homepage_node.htm
Table 2. Germany: Estimation results for dynamic Gompertz models with daily effects.

The likelihood is smaller at 16.43. There is no evidence of residual serial correlation in either model and although the heteroscedasticity statistic indicates a diminishing variance, its value seems to be heavily influenced by just one observation (March 9th) near the beginning of the series. The fit is shown in Figure 8.

Figure 8. New cases in Germany with smoothed estimates of level and slope components from dynamic Gompertz model.

The forecasts shown earlier in Figure 2 are quite remarkable in their accuracy over the next 36 days, that is, up to May 6th; see also Figure 9. The observations had not yet started to go down on April 1st, yet the sigmoid nature of the underlying growth curve means that the forecasts pick up the subsequent downward movement.

Fitting the dynamic Gompertz using data up to and including May 6th shows the model to be stable; see Table 2 and Figure 10. In particular the signal-noise ratio is almost unchanged. The observations in April are less variable than in March so the residuals are much smaller. The slope
appears to fall in early April but it then returns to a value similar to what it was at the end of March. In summary the dynamic Gompertz model works extremely well.\footnote{Even a deterministic trend gives a reasonably good fit for data from April 1st, although it is not as good as a model with a stochastic trend.}

4.4. Deaths. Deaths in Germany behave in a way similar to new cases: Figure 11 shows the log growth rates. There are some zero values in the first three weeks of March so the dynamic Gompertz is estimated using data from March 22nd. Because the numbers are smaller it is perhaps...
not surprising that the signal-noise ratio is estimated to be zero. In other words, we end up with the static Gompertz model.

The score-driven negative binomial model, introduced in sub-section 3.4, can be estimated with the TSL package of Lit et al. (2020). Observations from 11th March up to, and including, May 6th, are used. There are some zero values. Setting the \( \rho \) parameter to unity gives \( \bar{\alpha} = 0 \) - corresponding to a deterministic trend - and \( \bar{\gamma} = 0.071 \). The correlogram of the scores has a relatively high value at lag seven, indicating a daily effect. Modeling a fixed daily effect with dummy variables produces the fit in Figure 12, with the significantly increased log-likelihood. The parameter estimates are \( \bar{\gamma} = 0.070, \bar{\delta}_T = -4.14 \) and \( \bar{\upsilon} = 13.25 \). It is reassuring that the values of \( \bar{\gamma} \) and \( \bar{\delta}_T \) are consistent with those reported for new cases. The total number of deaths on May 6th was 6996 and the final total is predicted to be 8714. When we revised our paper on June 19th the total was 8946 and daily deaths had fallen to very low numbers.

Up to April 27th, the UK dataset on deaths from Coronavirus was restricted to deaths in hospitals. After April 28th, deaths in the community were included and the earlier figure revised. Although the relationship between deaths and new cases before April 28th is similar to that in Germany, there is a disconnect after that and the data on new cases are of very little help in predicting deaths.

Fitting a negative binomial model, with daily effects, to the series on UK deaths from March 17th to May 14th lead to an estimated \( \alpha \) of 0.34 (0.05). The changing trend is consistent with the Gaussian models fitted to new cases; the implied value of \( q_c \) is 0.007. The slope at the end of the series is 0.042 while \( \bar{\upsilon} = 29.91 \) (6.13). Figure 13 shows the forecasts of the underlying trend up to one month ahead. New observations after the forecasts were made are marked by dots.
Figure 12. Deaths in Germany and negative binomial filter with day of the week effect.

Figure 13. Forecasts (bold - blue) of daily deaths in the UK for one month ahead (using data up to and including May 14th).

5. THE EFFECT OF POLICY INTERVENTIONS

The effect of significant interventions, which may be a policy change, such as the introduction of a lockdown, or an external event, such as the arrival of a cruise ship in a small port, may be modeled by intervention (dummy) variables. The difficulty is that the pattern of the response is rarely known and so it becomes difficult to obtain a meaningful estimate of the final effect. Nevertheless some notion of the response can be obtained by making forecasts at the time the policy is thought to have become effective\(^7\) and comparing these forecasts with the actual outcome. In order to investigate

\(^7\)Flaxman et al. (2020) provide information on the incubation period.
this possibility, the dynamic Gompertz model was estimated on UK data up to and including March 31 (10 days after lockdown) using a fixed $q_\zeta$ of 0.001 (as estimated later with a larger sample and reported in Table 1). No day of the week effect was included, because the series is rather short. Figure 14 shows there is considerable overprediction. However, if the signal-noise ratio is increased to $q_\zeta = 0.01$, so the most recent observations receive more weight, the predictions are excellent. The final level is now approximately 290,000 as against 1.8 million. New cases are at their maximum, with a value of 5,400, on April 18th, whereas with $q_\zeta = 0.001$, they do not peak until May 20th, when they reach 21,500. Although the variation in predictions is huge, the same is true of many of the large models where the output can be very sensitive to the assumptions made.

![Figure 14. Forecasts of new cases in the UK, with $q = 0.01$ and $q = 0.001$ (dashed), using data up to April 1st.](image)

Information about the pattern of the response can also be obtained by fitting a dynamic Gompertz model and graphing estimates of the slope. Figure 15 shows filtered and smoothed estimates of slope in such a model based on data up to April 29th. The filtered estimates are most informative as they show the evolving changes in the slope after the lockdown of 21st March. As can be seen, the big falls occur at the beginning of April, with little or no change after mid-April. Figure 8 shows a very similar movement in Germany, but taking place a few weeks earlier.

Growth curves can be used to parameterize a gradual response to an intervention. A permanent change is captured with the CDF and a temporary one with the PDF. A logistic CDF gives a response curve $W(t) = 1/(1 + \gamma_0^l \exp(-\gamma_l^l(t - t_l^l)))$, where $t_l^l$ is the median. The $I$ superscript distinguishes the parameters $\gamma_0^l$ and $\gamma_l^l$ from the ones used to model the time series itself. With $t_l^L$ and $t_l^U$ denoting the beginning and the end of the time span during which gradual response to the intervention occurs, the intervention dummies are defined by $w_t = 0$ for $t < t_l^L$, $w_t = W(t)$ for $t = t_l^L$, $t_l^L + 1, ..., t_l^I, ..., t_l^U$ and $w_t = 1$ for $t = t_l^U + 1, ..., T$, or even just by $w_t = W(t)$ for $t = 1, ..., T$. With the Gompertz CDF the response is $W(t) = \exp(-\gamma_0^l \exp(-\gamma_l^l(t - t_l^l)))$; in this case the point of inflexion, corresponding to the maximum change, comes before the median of the time span between $t_l^L$ and $t_l^U$. Figure 16 shows the CDFs for logistic and Gompertz distributions with the median set to seven; for logistic $\gamma_l^l = 0.5$ and for Gompertz $\gamma_l^l = 0.2$. 
Figure 15. Filtered (solid blue) and smoothed (dashed - red) estimates of the slope in dynamic Gompertz model with daily effects (using data up to April 29th).

Figure 16. Cumulative distribution for logistic (green - solid) and Gompertz (red - dotted) functions with median set to 7. For logistic $\gamma = 0.5$ and for Gompertz $\gamma = 0.2$.

If the effect of a policy is to change $\gamma$ in the model, a slope intervention is needed in (3.2). Thus

$$\ln g_t = (\rho - 1) \ln Y_{t-1} + \delta - \gamma t - \beta t w_t + \varepsilon_t, \quad t = 1, \ldots, T,$$

but unless the sample is moderately large, $\rho$ will need to be fixed. When the full effect is realised, the slope on the time trend will have moved from $\gamma$ to $\gamma + \beta$. A positive $\beta$ will lower the growth rate, $g_t$, the peak of the incidence curve and the final level.

Fitting the static model in (5.1) to new cases in the UK with $\rho = 1$ and a logistic intervention, starting on March 26th and ending on April 12th, gives an estimate of $\beta$ equal to 0.020 (0.004) and an estimate of $\gamma$ also equal to 0.020. The picture in Figure 17 is not unreasonable but the estimate of the overall effect is 0.041 which may be a slight underestimate because minus one times the final
slope in Figure 15 is close to 0.05. When the slope is allowed to be stochastic, $\beta_T + \gamma$ is estimated at 0.054. However, a stochastic slope risks some confounding with the intervention variable and in fact the estimate of $\beta$ is reduced to 0.014 (0.006). Although neither model is completely satisfactory, both give a significant coefficient for the intervention variable, albeit after a degree of data mining.

![Figure 17. Estimates of logarithm of growth rate of total cases in UK with a logistic intervention and daily effect.](image)

6. A SECOND WAVE?

With the relaxation of a lockdown, the unwelcome prospect of a second wave of infections arises. Dynamic GL models can monitor this possibility by tracking changes in the growth rate of the incidence curve. This growth rate, $g_y(t)$, depends on a potentially time-varying $\gamma$ parameter, $\gamma(t)$, and the growth rate, $g(t)$. Differentiating the logarithm of $d\mu(t)/dt$ gives

\begin{equation}
(6.1) \quad g_y(t) = g(t) + g_y(t)
\end{equation}

and for GL curves, it follows from (2.4) that $g_y(t) = (\rho - 1)g(t) - \gamma(t)$. In the discrete time dynamic Gompertz model, (3.5), the negative of the growth rate of the growth rate, $g_y(t)$, is tracked by the filtered estimates of the slope, that is, $\gamma_{\ell|t}$, while the growth rate itself is tracked by the exponent of the filtered level, that is, $g_{\ell|t} = \exp(\delta_{\ell|t})$. Figure 18 shows $\gamma_{\ell|t}$ and $g_{\ell|t}$ for Germany from a dynamic Gompertz model, together with the daily number of new cases. The maximum is identified as April 3rd and after this date $g_{\ell|t} < \gamma_{\ell|t}$; see (2.7). The possible onset of a second wave is raised if at some point $\gamma_{\ell|t}$ starts to fall below $g_{\ell|t}$, or below $\rho g_{\ell|t}$ in the general case. This would signal that the reproduction number, $R_t$, has moved up above one.

The filtered estimates, $g_{\ell|t}$ and $\gamma_{\ell|t}$, are obtained by discounting of past observations, with the rate of discounting depending on the signal-noise ratio, $q_\kappa$. When a new policy is implemented, $q_\kappa$ may need to increase so that past observations are discounted at a faster rate, as illustrated in Figure 14. In these circumstances, the only viable tracking option may be to try a range of $q_\kappa$ values, bearing in mind the risk of triggering a false alarm. However, if the effects of a policy are
spread over a period of time, as with a gradual relaxation of a lockdown, a value of \( q_s \) estimated with the complete sample may be perfectly satisfactory.

If \( \gamma_{t|t-1} \) decreases, some idea of sampling variability is needed to assess the implications. This may be obtained from the standard deviation of \( \gamma_{t|t} \), as given by the Kalman filter. The variation in \( g_{t|t} \) is harder to determine, but since it is typically much smoother than \( \gamma_{t|t} \) it is likely to be small in comparison to the variation in \( \gamma_{t|t} \).

7. Conclusions

A new class of time series models is developed for predicting future values of a variable which when cumulated is subject to an unknown saturation level. Such situations arise in a wide range of disciplines. The models provide a simple and viable alternative, or complement, to the forecasts produced by large scale mechanistic models.

Generalized logistic growth curves provide the basis for our models. Estimating equations for the logarithm of the growth rate of the cumulative variable provide a good fit to the data, as assessed by standard statistical tests. Such models feature a time trend which can be made time-varying using the Kalman filter. The Gompertz model is a special case which works particularly well with a stochastic trend and, when this modification is made, the fit is often better than with an unrestricted generalized logistic model with a deterministic trend. The dynamic Gompertz can adapt to changing conditions and tracking the slope, especially the filtered slope, can be informative about the effect of interventions.

Additional components can be included in the models. These include seasonal or day of the week effects. The latter turned out to be relevant for new cases and deaths in our application to coronavirus in the UK and Germany. The dynamic Gompertz model worked well for both countries but was particularly impressive for German new cases. Deaths were successfully modelled by a Gompertz model with a negative binomial conditional distribution and a dynamic equation driven by the conditional score.
Estimating a model up to the point at which the effect of an intervention is likely to make itself felt and then making (unconditional) forecasts provides information about the effects of the intervention. The effectiveness of this approach is illustrated by fitting a dynamic Gompertz model to UK data and then investigating the effect of the Coronavirus lockdown. Further insight into the effect of the lockdown is given by tracking the filtered estimate of the growth rate of the growth rate. The impact of the lockdown can be estimated ex post by including a growth curve as an explanatory variable, thereby allowing the intervention to have a gradual response.

Finally, we suggest that the possibility of a second wave can be monitored by tracking the filtered estimates of new cases or deaths given by our model. Of course, if these methods are to be useful in practice they require reliable up-to-date observations on new cases, preferably at a disaggregated level. Implementing viable tracking procedures and relating them to current methods for tracking the reproduction number, $R$, is the next phase in our research program.

**Disclosure Statement.** The authors have no conflicts of interest to declare.

**Acknowledgments.** We would like to thank Dario Palumbo, Tom Pape, Mark Salmon, Stefan Scholtes, Jerome Simons, Robert Taylor, Steven Thiele, and Qingyuan Zhao for helpful comments and suggestions. We are also grateful to the Editor, Associate Editor and three referees for their constructive comments. Jonas Knecht supplied valuable research assistance.

**Appendix A. Data**

Surveillance data is the primary source of information for study of the dynamics of the pandemic. The data we use in this paper on time series of confirmed cases and deaths is sourced from the European Centre for Disease Prevention and Control (ECDC), which is a European Union agency with responsibility for EU level surveillance of infectious diseases. Since the beginning of the pandemic, through an Early Warning and Response System, ECDC has obtained daily numbers of confirmed cases of COVID-19 and deaths attributed to the virus from EU/EEA Member States and the UK.

ECDC’s data collection involves systematic screening of daily reports from health authorities by a team of epidemiologists, and an extensive validation process. The screened and validated daily reports are documented in a database, and the daily extract consisting of the total number of diagnosed COVID-19 infections and the number of deceased patients for each country is published on the ECDC website.

It is important to understand the possible sources of biases in the COVID-19 surveillance data. Angelopoulos et al. (2020) provide a systematic and insightful review of the many sources of bias in the time series of deaths, cases and recoveries relating to COVID-19. Their particular focus is on implications for the estimation of relative case fatality ratios between groups and countries. They show how some, though not all, biases can be addressed. Even when biases cannot be addressed, it is important to know their sources and implications.

In terms of the time series of cases, the definition of a “case” can vary from country to country, and may also change over time. Confirmation of cases is affected by the testing regime - test availability, and the criteria under which an individual is eligible for testing - which is liable to change over time. In the early days of the pandemic, severe cases will have been diagnosed more often than milder cases. True numbers of infections are likely to be much higher than reported in
countries where testing is limited. Biases also arise from imperfect or delayed reporting and errors in reporting.

Overall, deaths are more likely to be reported accurately by healthcare providers, but sources of bias remain. Time delays in reporting and the misattribution of deaths are not uncommon. In addition, national reporting guidelines during the early days are often focused on hospital deaths to the exclusion of deaths in community and care homes.

We focus on the UK and Germany due to our familiarity with the health data infrastructures and validation procedures in these countries. The data we use in this paper can be accessed from: https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide.

References

Angelopoulos, A., Pathak, R., Varma, R., & Jordan, M. (2020). On identifying and mitigating bias in the estimation of the COVID-19 case fatality rate. Harvard Data Science Review. 10.1162/99608f92.f01ee285

Aronson, J., Brasset, J., & Mahtani, K. (2020). When will it be over? An introduction to viral reproduction numbers. https://www.cebm.net/oxford-covid-19-evidence-service/

Avery, C., Bossert, W., Clark, A., Ellison, G., & Ellison, S. (2020). Policy implications of models of the spread of coronavirus: Perspectives and opportunities for economists. National Bureau of Economic Research Working Paper 27007. http://www.nber.org/papers/w27007

Chowell, G., Sattenspiel, L., Bausal, S., & Viboud, C. (2016). Mathematical models to characterize early epidemic growth. Physics of Life Reviews, 18, 66-97. https://doi.org/10.1016/j.phrev.2016.07.005

Ciufolini, I., & Paolozzi, A. (2020). Mathematical prediction of the time evolution of the COVID-19 pandemic in Italy by a Gauss error function and Monte Carlo simulations. The European Physical Journal Plus 135, 355. https://doi.org/10.1140/epjp/s13360-020-00383-y

Daley, D., & Gani, J. (1999). Epidemic modelling: An introduction. Cambridge: Cambridge University Press.

Flaxman, S., Mishra, S., Gandy, A., Unwin, H., Coupland, H., Mellan, T., Zhu, H., Berah, T., Eaton, J., Guzman, P., Schmit, N., Callizo, L., Team, I. C. C.-1. R., Whittaker, C., Winskill, P., Xi, X., Ghani, A., Donnelly, C., Riley, S., ... Bhatt, S. (2020). Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries. https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-13-europe-npi-impact/

Giordano, G., Blanchini, F., & Bruno, R. (2020). Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nature Medicine, 26, 855-860. https://doi.org/10.1038/s41591-020-0883-7

Gregg, J., Hossell, C., & Richardson, J. (1964). Mathematical trend curves: An aid to forecasting. ICI monograph no. 1. Edinburgh: Oliver and Boyd.

Harvey, A. (1984). Time series forecasting based on the logistic curve. Journal of the Operational Research Society, 35, 641-46.

Harvey, A. (2013). Dynamic models for volatility and heavy tails: With applications to financial and economic time series. Econometric Society Monograph, Cambridge University Press.

Kleiber, C., & Kotz, S. (2003). Statistical size distributions in economics and actuarial sciences. New York: Wiley.
Koopman, S., Lit, R., & Harvey, A. (2020). *STAMP 8.4: Structural time series analyser, modeller and predictor*. London: Timberlake Consultants Ltd.

Levenbach, H., & Reuter, B. (1976). Forecasting trending time series with relative growth rate models. *Technometrics, 18*, 26-72.

Lit, R., Koopman, S., & Harvey, A. (2020). *Time series lab - score edition*. https://timeserieslab.com

McDonald, J., & Xu, Y. (1995). A generalization of the beta distribution with applications. *Journal of Econometrics, 66*, 133-152.

Meade, N., & Islam, T. (1995). Forecasting with growth curves: An empirical comparison. *International Journal of Forecasting, 11*, 199-215.

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. *medRxiv*. https://doi.org/10.1101/2020.03.27.20043752

Panik, M. (2014). *Growth curve modeling: Theory and applications*. Hoboken: John Wiley & Sons.

Young, P., & Ord, J. (1989). Model selection and estimation for technological growth curves. *International Journal of Forecasting, 5*, 501-513.