Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Generalized trapezoidal ogive curves for fatality rate modeling

Johan René van Dorp¹,², Ekundayo Shittu³, Thomas A. Mazzuchi¹

School of Engineering and Applied Science, The George Washington University, 800 22nd Street, Washington DC, 20052, United States

ARTICLE INFO
Article history:
Received 17 July 2020
Accepted 29 July 2020
Available online 4 August 2020

Keywords:
Least squares curve fitting
Distribution theory
Forecasting

ABSTRACT
The construction of a continuous family of distributions on a compact set is demonstrated by concatenating, in a continuous manner, three probability density functions with bounded support using a modified mixture technique. The construction technique is similar to that of generalized trapezoidal (GT) distributions, but contrary to CT distributions, the resulting density function is smooth within its bounded domain. The construction of Generalized Trapezoidal Ogive (GTO) distributions was motivated by the COVID-19 epidemic, where smoothness of an infection rate curve may be a desirable property combined with the ability to separately model three stages and their durations as the epidemic progresses, being: (1) an increasing infection rate stage, (2) an infection rate stage of some stability and (3) a decreasing infection rate stage. The resulting model allows for asymmetry of the infection rate curve opposite to, for example, the Gaussian Error Infection (GEI) rate curve utilized early on for COVID-19 epidemic projections by the Institute for Health Metrics and Evaluation (IHME). While other asymmetric distributions too allow for the modeling of asymmetry, the ability to separately model the above three stages of an epidemic’s progression is a distinct feature of the model proposed. The latter avoids unrealistic projections of an epidemic’s right-tail in the absence of right tail data, which is an artifact of any fatality rate model where a left-tail fit determines its right-tail behavior.

© 2020 The Authors. Published by Elsevier Ltd.
This is an open access article under the CC BY-NC-ND license.
(http://creativecommons.org/licenses/by-nc-nd/4.0/)

1. Introduction
The recent outbreak of the severe acute respiratory syndrome coronavirus (SARS-CoV)-2, nicknamed COVID-19, in Wuhan in 2019, has spread and cascaded all over the world. Accompanying this global pandemic are numerous studies aimed at estimating and forecasting the scale of the fatalities. Unfortunately, very few, if any, of these studies have been sacrosanct in both their analysis and the prescriptions to policy. The plurality of fatality rate estimates has led to policy ambivalence with significant implications on the macro-economic indicators, the induced economic uncertainties, and effects on financial markets. At least in part, drivers and accelerators of these indicators are the global responses to the pandemic that includes, e.g., the population-wide social distancing risk mitigation measure. The presence of wide ranging fatality rate estimates, pitting expert guesses against each other, unfortunately results in an erosion of public trust in these fatality rate estimates and the models that produce them. The latter in turn complicates the adoption of proposed risk mitigation measures by a mistrusting public with potentially dire consequences. While the objective of this paper is also to contribute to the modeling effort in refinement of the early studies that have provided less than optimal and often knee-jerk reactions or forecasts, the value of this contribution is in offering a prudent approach toward fatality rate estimation as an epidemic progresses.

The plurality of fatality estimates results in part from a plurality of models that generate these estimates. A study using January and February data on COVID-19 cases and deaths in China applied a linear regression model to estimate the fatality rate for some of the clusters of the incidences of the disease in China. The study finds that the case fatality rate ranged from 0.15% to 5.25% with a concluding remark that the fatality rate is lower than the previous SARS epidemics. The lower end of this fatality rate was echoed in an analysis that points out the uncertainty of available official data in China that might lead to ambiguous results or inaccurate forecasts by non-trivial orders of magnitude. Other studies show that the impacts have been downplayed. SIR models are based on susceptibility (no immunity) S, count of infected I,

https://doi.org/10.1016/j.csfx.2020.100043
Z590-0544/© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license.
(http://creativecommons.org/licenses/by-nc-nd/4.0/)
and resistance (recovered or dead) \( R[3] \) illustrates how tenuous it is to estimate the fatality rate from the disease with emphasis on the uncertainty over the joint distribution of the fatality rate and the initial number of active cases. It is evident that model parameterization as the disease evolves becomes more tenuous when mitigation methods are simultaneously considered. Forecasting models have been cautionary in their predictions. For example, while the use of time series forecasting with models from the exponential smoothing family [8,13] have been proven to capture a variety of trend and seasonal forecasting patterns, non-seasonality assumes a prominent role [19].

The modeling approaches applied differ in their complexity and the assumptions they make. The latter emphasizes that uncertainties are not only rooted in the many unknowns about the contagious virus and the complicated mixture of human and government intervention policies [11], but also in the multitude of modelling approaches with varying assumptions that produce the fatality estimates. The latter has resulted in scrutiny of these approaches [5,22]. In Section 2, we present a preamble of modeling ensembles of the curve-fitting kind including bi-weekly cumulative fatality estimates resulting from three models using Italy COVID-19 fatality data. The data is compiled by the Johns Hopkins University Center for Systems Science and Engineering (JHU CCSE) from various sources [9]. Aside from assuming the functional form of a curve to be fitted, these models are purely data driven. Even in that case, the example demonstrates model uncertainty effects across these models resulting from sequentially fitting fatality rate curves as data accumulates over the progression of an epidemic. It is not unthinkably that with the inclusion of additional fatality rate curve models, with the same mathematical artifact where a left tail estimation cannot be decoupled from a right tail estimation, even larger model uncertainties may be observed.

Section 3 introduces the functional form of the proposed fatality rate model herein, termed Generalized Trapezoidal Ogive (GTO). It consists of three stages, being: (1) an increasing infection rate stage, (2) an infection rate stage of some stability and (3) a decreasing infection rate stage. The manner of its construction, which is somewhat technical, is described in the appendix. In Section 4, a least-squares fitting procedure is developed for the GTO model that allows for decoupling of the estimation of an epidemic’s three stages as it progresses. It is the latter fitting procedure that allows for prudence in providing third stages estimates in the absence of right-tail data. Indeed, not forecasting too far in the future is a time-honored prudent principle in the forecasting literature derived from the experience that past “historical” data may not be indicative of the future. In addition forecasting in general suffers from increasing uncertainties as the length of time for which one forecasts increases [7]. There is no empirical evidence, to the best of our knowledge, to suggest this does not apply to forecasting an epidemic’s progression. Section 5 presents several GTO fatality rate curve estimates using COVID-19 fatality data [9] from a number of different countries in Europe obtained utilizing the curve fitting procedure described in Section 4. Some closing remarks are provided in Section 6.

2. Model preamble

The functional form \( f(t|\alpha, \beta, \gamma) \) given by (1) was used for fatalities curve and infection rate curve modeling for the COVID-19 epidemic by the Institute for Health Metrics and Evaluation (IHME) [15,16],

\[
f(t|\alpha, \beta, \gamma) = \frac{\gamma}{2} \left[ 1 + \text{erf}(\alpha(t - \beta)) \right]
= \frac{\gamma}{2} \left( 1 + \frac{2}{\sqrt{\pi}} \int_{0}^{\alpha(t - \beta)} \exp(-x^2) dx \right),
\]

(1)

That functional form utilizes the Gauss error function \( \text{erf}(\cdot) \) which is equivalent to

\[
f(t|\alpha, \beta, \gamma) = \gamma \Phi(\sqrt{\pi} \times \alpha(t - \beta)),
\]

(2)

where \( \Phi(\cdot) \) is the cumulative distribution function of a standard normal distribution function. In (1) and (2), the \( \gamma \) parameter controls the total number of fatalities, the \( \beta \) parameter is the point in time at which the infection rate is at its maximum and the \( \alpha \) parameter controls the speed of infection or the infection rate

\[
\frac{df(t|\alpha, \beta, \gamma)}{dt} = \gamma \alpha \sqrt{\pi} \Phi(\sqrt{\pi} \times \alpha(t - \beta)),
\]

(3)

In (3), \( \Phi(\cdot) \) is the classical density function of a standard normal distribution. From the functional form of (2) and (3), it follows that the infection rate curve (3) has a symmetry axis at \( t = \beta \) which implies that the manner in which the infection rate increases up to \( \beta \) is the same as the manner in which the infection rate decreases after \( \beta \). Furthermore, by implication, the time to the maximum infection rate \( \beta \) from a negligible one then also equals to the amount of time from maximum infection rate \( \beta \) to a near zero infection rate.

During the COVID-19 epidemic the parameters \( \alpha, \beta, \gamma \) in (2) and (3) were estimated as the epidemic progressed, say on a weekly-basis or even more frequent. As a result, in the early stages of the epidemic, the estimates of \( \alpha, \beta, \gamma \) in (2) and (3) were based on fatality numbers in the left-tail of the fatality rate curve, in the absence of right-tail fatality numbers. As per the previous paragraph, these parameters determine the “estimated/forecasted” duration and severity of the epidemic that, in turn, served as a driver of risk mitigation measure implemented or recommended, such as the duration of “lock-down” periods. In the absence of that right-tail data of the fatality rate curve, however, such early estimates carry a great deal of uncertainty and inaccuracy with them. Truth be told, however, this is not a result of the symmetry of the functional form of (2) and (3), but an artifact of any fatality curve model where its parameters determine both its left-tail and right-tail behaviors.

Alternative infection rate models \( f(t|\gamma, \Theta) \) follow from (2) by substituting a cumulative distribution function (cdf) \( G(\cdot|\Theta) \) for the standard normal cdf \( \Phi(\cdot) \) in (2) as follows;

\[
f(t|\gamma, \Theta) = \gamma G(t|\Theta),
\]

(4)

where \( \Theta \) are its parameters and \( \gamma > 0 \) as in (2) and (3) controls the total number of fatalities. One obtains for the infection rate curve in that case;

\[
\frac{df(t|\gamma, \Theta)}{dt} = \gamma g(t|\Theta),
\]

(5)

where \( g(t|\Theta) \) is the probability density function (pdf) associated with the cdf \( G(\cdot|\Theta) \). To illustrate the large potential inaccuracies/fluctuations of early “forecasts”, Fig. 1 demonstrates the cumulative fatality (CF) curves and fatality rate (FR) estimated using the standard normal cdf \( \Phi(\cdot) \) in (2), the gamma\((\alpha, \beta)\) cdf and the log-normal\((\mu, \sigma)\) cdf with pdfs

\[
g(t|\alpha, \beta) = \frac{t^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha} \exp(-\frac{t}{\beta})
\]

(6)

and

\[
g(t|\mu, \sigma) = \phi(\frac{\ln(x) - \mu}{\sigma})
\]

(7)

respectively. Contrary to (2) and (3), (6) and (7) both allow for the modeling of asymmetry in the fatality rate curve (5). The curves in Fig. 1 were fitted to fatality data \( y = (y_1, \ldots, y_k) \) for Italy on a rolling bi-weekly basis using the least squares estimation procedure that minimizes

\[
SS(\Theta) = \sum_{i=1}^{k} [y_i - \gamma(G(t_i|\Theta))]^2
\]

(8)
Fig. 1. Italy COVID-19 fatality data [9] with seven biweekly least-squares fitted fatality curves. A: Gaussian CF Curves (2); B: Gaussian FR Curves (3); C: Gamma CF Curves with cdf of (6); D: Gamma FR Curves (6); E: log-normal CF Curves with cdf of (7); F: log-normal FR Curves (7).
where $G_i(\cdot; \Theta)$ is given by (4), $t_i = (t_1, \ldots, t_k)$, and $y_i$ are the cumulative number of fatalities reported [9] on days $t_i$, $i = 1, \ldots, k$. The parameters in Fig. 1 were estimated using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) optimization algorithm available in the open-source platform R [17]. At 14 days, the distribution parameter starting values in (2), (6) and (7) were set equal to 1 in that optimization procedure, whereas the starting value for the parameter $\gamma$ in (5) was set equal to the cumulative number of fatalities observed up to that point in time. For subsequent biweekly least-squares, fitting the starting values for the distribution parameters were set to those of the prior biweekly optimization results, and the starting value of the parameter $\gamma$ was updated to the cumulative number of fatalities observed.

From Figs. 1C and 1D, one observes that the biweekly estimated gamma curves both underestimate and overestimate the number of fatalities with a max fatality of 42,319 estimated at 42 days into the epidemic, whereas after 85 days a total number of 31,368 fatalities were observed. The next biweekly fatality estimate at 56 days reduced to 25,978, about a 40% reduction. From Fig. 1E and 1F, one observes that for the log-normal fatality curves this estimation behavior is even more pronounced with a max fatality estimate of 49,343 at 42 days into the epidemic and a reduced fatality estimate at 56 days equal to 28,400, a 42% reduction. From Fig. 1B, in the cases of the Gaussian fatality rate curves, one observes that, consistently, a duration below 100 days until negligible infection rates is estimated. On the other hand, in the cases of the gamma and the log-normal curves in Fig. 1D and 1F, a duration of about up to 150 days or more than 150 days is estimated (at 42 days into the epidemic) until negligible infection rates, respectively. Summarizing, Fig. 1 demonstrates model uncertainty effects across the three different fatality rate models depicted when sequentially fitting fatality rate curves as such data accumulates over the progression of an epidemic.

Given the potential for unintended consequences resulting from fatality and epidemic duration estimates that “miss the mark” as a result of those model uncertainties in the absence of right-tail data, a fatality rate curve model is developed herein that allows for a separate estimation of the left and right tail behavior of an epidemic’s progression. It utilizes a bounded density function $g(\cdot | \Theta)$, termed the GTO distribution, reminiscent of generalized trapezoidal distribution functions with a uniform central stage, also known as Generalized Trapezoidal Uniform (GTU) distributions [24]. Similar to the GTU density function, it has bounded support to account for the duration of an epidemic being bounded, and it consists of three stages: (1) a growth stage, (2) a stability stage, and (3) a decay stage. A growth parameter $m$ and decay parameter $n$ model the growth and decay in Stage 1 and Stage 3, respectively. The length of the three stages are modeled with parameters

$$0 \leq a < b \leq c < d. \quad (9)$$

Note that in Eq. (9) one can set $b = c$, which eliminates the second stability stage from a GTO fatality curve model should this be apparent from the data collected. An epidemic’s total duration modeled by the parameter $d$ can be set arbitrarily large. An advantage of the GTO distribution over GTU distributions for fatality rate curve modeling is that its density is smooth over its bounded support, which may be a desirable property in such a context.

### 3. GTO distributions

The approach towards constructing the GTO pdf $g(\cdot | \Theta)$ requires one to specify: (i) the beginning and end of each of the three stages ($a, b, c, d$), (ii) the growth behavior of the first stage (parameter $m$), (iii) the decay behavior of the third stage (parameter $n$), and (iv) the requirement that the relative likelihood at the end of the growth stage $[a, b]$ and at the beginning of the decay stage $[c, d]$ equals one, i.e.,

$$g(b | \Theta)/g(c | \Theta) = 1. \quad (10)$$

To allow for nonlinear growth and decay the probability density functions

$$g_1(t | \Theta) = \frac{m}{b-a} \left( \frac{x-a}{b-a} \right)^{m-1} \left( \frac{4m-2}{3m-1} - \frac{2m-2}{3m-1} \left( \frac{t-a}{b-a} \right)^m \right), \quad a < t \leq b \quad (11)$$

and

$$g_3(t | \Theta) = \frac{n}{d-c} \left( \frac{d-t}{d-c} \right)^{n-1} \left( \frac{4n-2}{3n-1} - \frac{2n-2}{3n-1} \left( \frac{d-t}{d-c} \right)^n \right), \quad c \leq t < d. \quad (12)$$

are chosen for the first and the third stage, respectively. The density function $g_2(t | \Theta)$ during the second stage will be assumed to be of a constant value, i.e.,

$$g_2(t | \Theta) = \frac{1}{c-b}, \quad b \leq t \leq c. \quad (13)$$

The functional form of the GTO density function is next

$$g(t | \Theta) = \begin{cases} 
\pi_1(\Theta) \times g_1(t | \Theta), & a < t \leq b, \\
\pi_2(\Theta) \times g_2(t | \Theta), & b \leq t \leq c, \\
\pi_3(\Theta) \times g_3(t | \Theta), & c \leq t < d,
\end{cases} \quad (14)$$

where $\Theta = \{a, b, c, d, m, n\}, a < b \leq c < d, m \geq \frac{1}{2}, n \geq \frac{1}{2}$, and $\pi_i(\Theta), i = 1, \ldots, 3$, are the stage probabilities of the fatality rate curve (5) or the proportion of fatalities in each stage of an epidemic. The main challenge in the construction of the pdf (14) is to select the probabilities $\pi_1(\Theta)$, $\pi_2(\Theta)$, $\pi_3(\Theta)$ so that the overall density function $g(t | \Theta)$ in (14) be continuous. This turns out to be a nontrivial problem, but was solved in the derivation of Generalized Trapezoidal (GT) distributions in [23]. That same approach is utilized herein to prove the following proposition.

**Proposition.** The probability density function given by (14) with parameters

$$\Theta = \{a, b, c, d, m, n\}, a < b \leq c < d, m \geq \frac{1}{2}, n \geq \frac{1}{2}$$

generates a density function

$$p(y | m) = my^{m-1} \left[ \frac{4m-2}{3m-1} - \frac{2m-2}{3m-1} y^m \right], \quad 0 \leq y \leq 1, m \geq 1/2, \quad (16)$$

with cdf

$$P(y | m) = y^m \left[ \frac{4m-2}{3m-1} - \frac{m-1}{3m-1} Y^m \right], \quad 0 \leq y \leq 1, m \geq 1/2, \quad (17)$$

and quantile function

$$P^{-1}(q | m) = \frac{2m-1}{m-1} \left[ 1 - \sqrt{1 + q \left( \frac{m-1}{2m-1} \right) \left( \frac{3m-1}{m-1} \right)^2} \right], \quad 0 \leq q \leq 1, m \geq 1/2. \quad (18)$$
introduced in [10]. The density function (12) is a linear transformation of a reflected ogive density function. An important property that was used to construct the ogive pdf (16) in [10] is

$$\lim_{y \to 1} \frac{dp(y|m)}{dy} = 0. \quad (19)$$

The density (16) was termed an ogive density since the pdf allows, with property (19), for strictly increasing S-shaped density curves. It is that latter property (19) of ogive pdfs (16) that ensures that the GTO density function (14) is smooth and differentiable at its stage transitions, i.e., at b and c. The parameters m and n are the growth and decay parameters of the stages provided $m > 1$, $n > 1$. Finally, since the cdf $P(y|m)$ (17) and quantile function $P^{-1}(q|m)$ (18) are available in a closed form, it follows that the cdf and the quantile function of GTO distributions are also available in a closed form. The GTO cdf $G(t|\Theta)$ follows as,

$$G(t|\Theta) = \begin{cases} 
\pi_1(\Theta)P\left(\frac{t-a}{b-a} \right|m), & a < t \leq b, \\
\pi_1(\Theta) + \pi_2(\Theta) - \frac{b-t}{b-c}, & b \leq t \leq c, \\
1 - \pi_3(\Theta)P\left(\frac{d-t}{n}\right), & c \leq t \leq d,
\end{cases} \quad (20)$$

where $\Theta = \{a, b, c, d, m, n\}, a < b \leq c \leq d$, $m \geq \frac{1}{2}, n \geq \frac{1}{2}$, and $\pi_i(\Theta), i = 1, \ldots, 3$ are given by (15). The quantile function $G^{-1}(p|\Theta)$, follows as

$$G^{-1}(p|\Theta) = \begin{cases} 
p^{-1}(\pi_1(\Theta)(m(b-a) + a), & 0 < p \leq \pi_1(\Theta), \\
p - \pi_1(\Theta) + \frac{\pi_2(\Theta)}{\pi_3(\Theta)}(c-b) + b, & \pi_1(\Theta) \leq p \leq 1 - \pi_3(\Theta), \\
d - p^{-1}\left(1 - \frac{p}{\pi_3(\Theta)}\right)n(d-c), & 1 - \pi_3(\Theta) \leq p < 1.
\end{cases}$$

In other words, contrary to, e.g., the infection rate curves used in Fig. 1, no numerical approximation procedures are required to evaluate the CF curve (4) that utilize the GTO cdf. In the next section, a sequential fitting procedure of the stage parameters using the least squares method is developed.

4. Sequential fitting of a GTO fatality curve

An objective of fitting fatality curves in an epidemic is the application of a decision rule as to the timing of the progression of the epidemic. Such a progression typically consists of an increasing fatality and infection rate stage, potentially followed by a stage of relative stability and finally a stage where the epidemics fatality and infection rates decreases. While the determination of these stages after an epidemic has run its course is perhaps straightforward, the determination of these stages as an epidemic progresses is complicated due to reporting behavior of daily infection and fa-
tality counts. For example, in the case of the COVID-19 epidemic fatalities reported on Tuesdays were routinely higher than those on other days in the week, whereas fatalities reported on Sundays and Mondays were lower. This cyclical behavior can be observed for the fatalities reported in Italy in Figs. 1B, 1D and 1E. Similar over-reporting and under-reporting patterns were observed in daily infection and fatality counts for other countries as well.

Cyclical reporting of daily infection and fatality counts not only complicates the fitting of an infection and fatality rate curve as an epidemic progresses and the determination of its stages, but it also questions the rationale of daily reporting of infection rate and fatality counts, without reporting them in the context of that cyclical nature. Providing a methodology under such reporting conditions may help inform policy decision making in terms of extending or relaxing risk mitigation strategies as the epidemic progresses. It is worthwhile to note here that the fluctuations observed in the fitted curves depicted in Fig. 1 are not a result of that cyclical reporting behavior since the curves were fitted on a bi-weekly basis. That being said, one can expect even larger fluctuations and model uncertainties should one choose to fit infection and fatality rate curves on a daily basis, under such cyclical reporting conditions.

4.1. Fitting the first stage

Consider the least-squares objective function \( SS(\hat{\Theta}|\mathbf{y}, t, k) \) (8). In (8) the vector \( \mathbf{y} \) contains the cumulative fatality and infection counts at the times specified by \( t = (1, \ldots, k) \). The first days of the epidemic was selected in Fig. 1, such that the first fatality and infection were counted on the second day. As the epidemic progresses the function \( SS(\hat{\Theta}|\mathbf{y}, t, k) \) increases as a function of \( k \) since additional squared residuals are added to (8) on a day-to-day basis. To assess if a transition from the first stage has occurred to the second stage of the epidemic, the following setting of the parameters is utilized:

\[
b = c = d, n = 1.
\]

The effect of the parameter setting (21) on the GTO cdf (20) is that it only consists of a single stage. Next, as more data becomes available the total number of fatalities (in the first stage) \( \gamma \) and the growth parameter \( m \) are fitted on a daily basis by minimizing the least squares function \( SS(\hat{\Theta}|\mathbf{y}, t, k) \) (8), while setting \( b = k \). To evaluate if the first stage has ended, the progression of

\[
SS(\hat{\Theta}|\mathbf{y}, t, k)/\gamma
\]

is tracked, where \( \gamma \) is the estimated cumulative number of fatalities and infections on day \( k \). Fig. 2A, plots the progression of optimal values of \( SS(\hat{\Theta}|\mathbf{y}, t, k) \) (8) as \( k \) increases and Fig. 2B plots the progression of the normalized optimal values \( SS(\hat{\Theta}|\mathbf{y}, t, k)/\gamma \) (22). From Fig. 2B, one observes that \( SS(\hat{\Theta}|\mathbf{y}, t, k)/\gamma \) shows a fluctuating behavior up to day 38, followed by a strictly increasing behavior thereafter. From that latter increasing behavior, it may be concluded that the daily fatality counts started to level off at day 38.
The first stage fit of the cumulative GTO fatality rate curve (4) up to day 38 is depicted in Fig. 2C with optimal parameter values $\tilde{\tau} = 11.101$ and power parameter $\tilde{m} = 4.411$. Hence, it is estimated that the first stage of the epidemic accounted for 11,101 fatalities. The associated first stage GTO fatality counts curve (5) is depicted in Fig. 2D. Note the significant drop-off in observed fatality counts following day 38 in Fig. 2D.

### 4.2. Fitting the second stage

The procedure to fitting the second stage is similar to that of fitting the first stage except that the following setting of the parameters is utilized:

$$\hat{b} = \hat{b}, \; \hat{b} \leq \hat{c} = \hat{d}, \; n = 1.$$  

Thus in the COVID-19 Italy case $b$ is kept fixed at the value $\hat{b} = 38$ and $SS(\tilde{\Theta}, |x, c, k|$ (8) is minimized over the parameters $\gamma$ and the first stage power parameter $m$ as the number of days $k$ in the epidemic progresses onward from $\hat{b} = 38$, while setting $c = k$.

The COVID-19 Italy fatality data and the results are depicted in Fig. 3. As in Fig. 2, Fig. 3A tracks the progression of $SS(\tilde{\Theta}, |x, c, k|$ (8) and Fig. 3B tracks the progression of $SS(\tilde{\Theta}, |x, c, k|$ (22). Similar to Fig. 2B, one observes a strictly increasing behavior of $SS(\tilde{\Theta}, |x, c, k|$ (22) following day 43, which can be interpreted as

![GTO Model for Number of Covid Deaths - Italy](image)
having entered the decreasing stage of the epidemic. Truth be told, the increase in $SS(\hat{\Theta} | y, \tilde{t}, k) / \hat{\gamma}$ from day 43 to day 44 is only minor, and a modeling decision could have been made to set $c = 44$ as opposed to $c = 43$. The optimal left branch power parameter of the GTO fatality curve $\gamma \mathcal{C}(t | \Theta)$ (4) is now $\hat{m} = 4.224$ (updated from the value $\hat{m} = 4.411$ when only the first stage was fitted in Fig. 2), whereas the estimated total number of fatalities during the first and second stage is now estimated at $\hat{\gamma} = 15,030$ thousand. In other words, it is estimated that during the second stage the number of fatalities increased by $15,030 - 11,101 = 3,929$ thousand. The first and second stage fit of the cumulative GTO fatality rate curve (4) up to day 43 is depicted in Fig. 3C. The associated first stage and second stage GTO fatality counts curve (5) up to day 43 is depicted in Fig. 3D. Note the significant reduction in observed fatality counts following day 43 in Fig. 3D.

4.3. Fitting the final stage

To fit the complete GTO fatality curve $\gamma \mathcal{C}(t | \Theta)$ (4) with $\mathcal{C}(t | \Theta)$ given by (20) the following setting of the parameters is utilized $b = \hat{b}$, $c = \hat{c}$, $c < d$, where $\hat{b}$ and $\hat{c}$ follow from the first stage and second stage fitting procedure. Hence in the COVID-19 Italy case, we preset $b = 38$ and $c = 43$. Next, $SS(\Theta | y, \tilde{t}, k)$ (8) is minimized over the parameters $\gamma$ and the first and second stage power parameters $m$ and $n$ for different values of $d$, where $d > k$ can be chosen arbitrarily large. The results for the COVID-19 Italy case depicted in Fig. 1 (i.e. at $k = 85$ days into the epidemic) are depicted in Fig. 4 starting with a value $d = 150$ days up to $d = 800$ days. From Fig. 4A, one observes a decreasing behavior of the optimal value of $SS(\Theta | y, \tilde{t}, k)$ (8) and an increasing behavior of the maximum estimated number of fatalities $\hat{\gamma}$ as one increases $d$. In addition, one observes from Fig. 4B that the optimal value for $\gamma$ starts to level at $\hat{d} = 800$. The optimal values of the other parameters, while setting $b = 38$, $\hat{c} = 43$ and $\hat{d} = 800$, are: $\hat{\gamma} = 33.377$, $\hat{m} = 3.705$, $\hat{n} = 42.480$.

In Fig. 4C the observed number of fatalities up to 85 days is indicated at 31,368. Hence an additional 33,377 – 31,368 = 2,009 fatalities are estimated 85 days into the epidemic by the GTO fatality curve model. From Fig. 4D, one observes a daily fatality count close to zero at about 150 days into the epidemic.

5. Fit comparison of some European countries

In this section, an overall fit comparison is presented between the GTO fatality curve model (20) and the Gaussian (2), gamma (6) and log-normal (7) ones that all follow from (4) and (5). Fig. 5 provides a visual fit comparison utilizing 138 days of Italy COVID-19 data [9]. That data set was also utilized to demonstrate the sequential GTO fitting procedure in Section 4.

As in Sections 2 and 4, the curves in Fig. 5 were estimated utilizing the Broyden–Fletcher–Goldfarb–Shanno optimization algorithm available in R [17] but, this time, using 138 days of fatality rate data. The fitted parameter values are provided in Table 1 together with the optimal value of the objective function $SS(\hat{\Theta} | y, \tilde{t}, k)$ (8) indicated by $SS$ in Table 1. Fig. 5A plots the cumulative fatality rate curves (4) and Fig. 5B plots the fatality counts per day (5). The differences between the fitted models are more apparent in Fig. 5B than Fig. 5A. One observes through careful visual inspection from Fig. 5B that the Gaussian model defined by (1) and (2) demonstrates the worst performance. That Gaussian model was used early on in the COVID-19 pandemic by the IHME institute [15,16]. The Gaussian model is followed in performance by the fitted gamma model (6). The log-normal model (7) outperforms the GTO model (14) in Fig. 5 in terms overall fit.
Fig. 6. Fatality curves comparison fitted to COVID-19 fatality data from the United Kingdom (UK), Spain (SP), and The Netherlands (NED) [9]. The fatality curves fitted are the GTO Model (20), the Gaussian model (2), the gamma model (6) and finally the log-normal model (7). The parameter values of these fitted models are provided in Table 1. A: UK cumulative fatality curves; B: UK daily fatality count curves; C: SP cumulative fatality curves; D: SP daily fatality count curves; E: NED cumulative fatality curves; F: NED daily fatality count curves.
This ordering of fit performance also follows from the SS-values in Table 1 for Italy. That being said, the cumulative number of fatalities estimated by the GTO model, log-normal and gamma models are similar in value, indicated by the \( \tilde{y} \)-values in Table 1. None of the fitted models could capture the fluctuations in the reported daily death numbers as is demonstrated in Fig. 5B. While the log-normal model performs best in the least-squares sense, it suffers from overall fatality estimates fluctuations in a bi-weekly fitting regimen, as demonstrated in Section 2. This is the result of the mathematical artifact in the log-normal model that prevents de-coupling of the fitting of the first stage and the last stage of an epidemic. The same applies to the Gaussian and gamma models. The latter was the main objective for the introduction of the GTO model as explained in Sections 3 and 4. One visually observes from Fig. 5A and 5B that both the GTO model (20) and the log-normal model (7) are close in comparison, with the biggest difference observed in the left tails of their fatality curves in Fig. 5B.

Fig. 6 provides a comparison of the fitted GTO Model (20), the Gaussian model (2), the gamma model (6) and finally the log-normal model (7) to COVID-19 fatality data [9] from the United Kingdom (UK), Spain (SP), and The Netherlands (NED). The parameters of these fitted models are also provided in Table 1. From Fig. 6, we observe a close fit between the fitted GTO and log-normal fatality rate curves as in Fig. 5. The gamma model again comes third and the Gaussian model performs last. The same observations can be made from the SS column in Table 1. The log-normal model also consistently provides the highest number of estimated cumulative fatalities, with the GTO model coming second in this regard as well.

6. Closing remarks

The data reporting problems of COVID-19 fatality data identified in Fig. 5 are once more apparent from Fig. 6B, 6D and 6F. In fact, the cyclical behavior of reporting fatalities are more pronounced in Fig. 6B and 6F than in Fig. 5B. Other reporting issues can be observed in Fig. 6C. For example, the cumulative number of fatalities decreases at one point in Fig. 6C (resulting in a negative daily fatality count in Fig. 6D). This is followed in Fig. 6C by a period where the reported cumulative number of fatalities remained constant resulting in turn in a period of zero fatality counts in Fig. 6D. After that period a large spike in fatality counts is displayed in Fig. 6D after 100 days. These data reporting issues further complicate the fitting of fatality curves and, quite frankly, the interpretation of the “closest” fit based on the data provided. No doubt, it would appear that an improvement in the reporting process of fatality data during an epidemic is called for. In particular, it would be desirable that such a reporting process does not result in a cyclical reporting behavior, but also that this process is consistent with the policies and procedures under which fatalities attributable to an epidemic are reported over its course.

In the absence of such data collection and reporting practices, the utilization of the GTO model compares well in the overall fit to the “best” fitting log-normal model, in the least squares sense, while providing a fatality curve fitting approach that allows for prudence in providing cumulative fatality estimates as an epidemic progresses. Providing large fluctuating cumulative fatality estimates, as e.g., demonstrated for the log-normal model in Fig. 1, has the potential of eroding public trust in models that provide these estimates/forecasts. In fact, [21] points out that a platform of trust is more fundamental to risk management than risk communication. Hence, the erosion of public-trust complicates the adoption of proposed risk mitigation measures by the public to combat an epidemic, with potentially dire consequences. Perhaps the prudence that the GTO fatality curve model offers in providing fatality estimates can contribute to reducing such an erosion of public trust.

Declaration of Competing Interest

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

Appendix A. Construction of the GTO distribution

Proposition. The probability density function given by (14) with parameters

\[ \Theta = \{a, b, c, d, m, n\}, a < b < c < d, m \geq \frac{1}{2}, n \geq \frac{1}{2} \]

that follows from expressions (10), (11), (12) and (13) that utilize stage probabilities

\[
\begin{align*}
\pi_1(\Theta) &= \frac{n^2(b-a)(3m-1)}{m(d-c)(3n-1) + 2dn^2(c-b) + n^2(b-a)(3m-1)}, \\
\pi_2(\Theta) &= \frac{2n^2(c-b)}{m(d-c)(3n-1) + 2dn^2(c-b) + n^2(b-a)(3m-1)}, \\
\pi_3(\Theta) &= \frac{n^2(b-a)(3m-1)}{m(d-c)(3n-1) + 2dn^2(c-b) + n^2(b-a)(3m-1)}
\end{align*}
\]

is continuous and smooth over its support \([a, d]\).

Proof. Expression (14) is obtained using a mixture of three pdfs \(g_i(t|\Theta)\) given by (11), (12) and (13), i.e.

\[
g(t|\Theta) = \sum_{i=1}^{3} \pi_i(\Theta) g_i(t|\Theta), \quad a \leq t < d, \quad \text{elsewhere,} \quad (A.24)
\]

where \(\sum_{i=1}^{3} \pi_i(\Theta) = 1, \pi_i(\Theta) > 0\). Utilizing (11), (12) and (13) and (A.24) the density function of the proposed generalized trapezoidal ogive pdf given by (A.24) can be rewritten as in (14). It will be convenient to write the mixture weights \(\pi_i(\Theta)\), \(i = 1, 2, 3\), in the form

\[
\pi_1(\Theta) = \beta p, \quad \pi_2(\Theta) = (1 - \beta), \quad \pi_3(\Theta) = \beta(1 - p) \quad (A.25)
\]

where \(0 < \beta < 1, 0 < p < 1\). This implies that \(\pi_i(\Theta) > 0, i = 1, 2, 3\) and

\[
\sum_{i=1}^{3} \pi_i(\Theta) = \beta p + (1 - \beta) + \beta(1 - p) = 1. \quad (A.26)
\]

From (A.25), utilizing (10)–(14), we have

\[
1 = \int_{a}^{b} g_i(b|\Theta) \frac{\beta p g_1(b|\Theta)}{\beta(1-p)g_3(b|\Theta)} = \frac{p \frac{2}{\beta-a} \frac{3}{3m-1}}{\left(1-p\right) \frac{2}{\beta-c} \frac{3}{3n-1}},
\]

where \(g_i(b|\Theta) = \lim_{t \to b} g(t|\Theta)\) and \(g_i(c|\Theta) = \lim_{t \to c} g(t|\Theta)\), yielding

\[
p = \frac{n^2(b-a)(3m-1)}{m(d-c)(3n-1) + 2dn^2(c-b) + n^2(b-a)(3m-1)}. \quad (A.27)
\]

Observe that \(p\) does not depend on \(\beta\). Also the stipulations \(a < b < c < d, m > \frac{1}{2}, n > \frac{1}{2}\) imply \(0 < p < 1\).

Continuity of (14) at \(b\) will follow from the requirement that

\[
g_1(b|\Theta) = g_2(b|\Theta), \quad (A.28)
\]

implying with (14) and (A.26) that

\[
\beta pg_1(b|\Theta) = (1 - \beta)g_2(b|\Theta), \quad (A.29)
\]

yielding

\[
\beta = \frac{g_2(b|\Theta)}{pg_1(b|\Theta) + g_2(b|\Theta)}. \quad (A.30)
\]
The choice of $\beta$ in (A.30) assures $0 < \beta < 1$ and continuity of $g_x(\cdot|\Theta)$ (14) at $b$. Utilizing (10), (11), (A.27) and (A.30), it follows after some algebraic manipulations that

$$\beta = \frac{m^2(d-c)(3n-1) + n^2(b-a)(3m-1)}{m^2(d-c)(3n-1) + 2n^2m^2(c-b) + n^2(b-a)(3m-1)}. \quad (A.31)$$

Substituting (A.31) and (A.27) into (A.25) we arrive at (23) given by

$$\begin{align*}
&\pi_1(\Theta) = \frac{n(b-a)(3m-1)}{m^2(d-c)(3m-1) + 2n^2m^2(c-b) + n^2(b-a)(3m-1)}, \\
&\pi_2(\Theta) = \frac{2mn^2(c-b)}{m^2(d-c)(3n-1) + 2n^2m^2(c-b) + n^2(b-a)(3m-1)}, \\
&\pi_3(\Theta) = \frac{n^2(b-a)(3m-1)}{m^2(d-c)(3n-1) + 2n^2m^2(c-b) + n^2(b-a)(3m-1)}. \quad (A.32)
\end{align*}$$

Finally, substitution of (11), (12), (13) and (A.32) into (A.24), yields (14). Smoothness of the density function (14) over its support $(a, d)$ follows directly from the smoothness of the pdfs (11), (12), (13), the property (19) of ogive pdfs and the density of the central stage of (13) being constant. □

**CRediT authorship contribution statement**

**Johan René van Dorp**: Conceptualization, Methodology, Software, Writing - original draft. **Ekundayo Shittu**: Writing - original draft. **Thomas A. Mazzuchi**: Supervision.

**References**

[1] Anastassopoulou C, Russo L, Tsakris A, Siettos C. Data-based analysis, modelling and forecasting of the COVID-19 outbreak. PLoS ONE 2020;15(3):e0230405.
[2] Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? The Lancet 2020;395(10228):931–4.
[3] Atkeson A. How deadly is COVID-19? Understanding the difficulties with estimation of its fatality rate. Tech. Rep., National Bureau of Economic Research; 2020.
[4] Baker SR, Bloom N, Davis SJ, Terry SJ. Covid-induced economic uncertainty. Tech. Rep., National Bureau of Economic Research; 2020.
[5] Baud D, Oji X, Nielsen-Saines K, Musso D, Pomar L, Favre G. Real estimates of mortality following COVID-19 infection. Lancet Infect Dis 2020.
[6] Goerrieri V, Lorenzoni G, Straub L, Werning I. Macroeconomic implications of COVID-19: Can negative supply shocks cause demand shortages? Tech. Rep., National Bureau of Economic Research; 2020.
[7] Hyndman RJ, Athanasopoulos G. Forecasting: principles and practice. OTexts: 2018.
[8] Hyndman RJ, Koehler AB, Snyder RD, Grose S. A state space framework for automatic forecasting using exponential smoothing methods. Int J Forecast 2002;18(3):439–54.
[9] JHU CSSE. Novel coronavirus (COVID-19) cases data. 2020. https://data.humdat.org/dataset/novel-coronavirus-2019-ncov-cases: Accessed July 11, 2020.
[10] Krotz S, Van Dorp JRF. Beyond beta: other continuous families of distributions with bounded support and applications. World Scientific; 2004.
[11] Luo J. Predictive monitoring of COVID-19. SUTD Data-Driven Innovation Lab; 2020.
[12] Mahase E. Coronavirus: COVID-19 has killed more people than SARS and MERS combined, despite lower case fatality rate. 2020.
[13] Makridakis S, Wakerfield A, Kirkham R, et al. Predicting medical risks and appreciating uncertainty. Foresight: The International Journal of Applied Forecasting 2019;52:28–35.
[14] McKibbon WA, Fernando R. The global macroeconomic impacts of COVID-19: Seven scenarios 2020; https://www.brookings.edu/research/ the-global-macroeconomic-impacts-of-covid-19-seven-scenarios. Accessed August 4, 2020.
[15] Murray C, et al. Forecasting covid-19 impact on hospital bed-days, icu-days, ventilator-days and deaths by us state in the next 4 months. medRxiv 2020. doi:10.1101/2020.03.27.20043752.
[16] Murray C, et al. Forecasting the impact of the first wave of the covid-19 pandemic on hospital demand and deaths for the us and european economic area countries. medRxiv 2020. doi:10.1101/2020.04.21.20074732.
[17] Nash JC, Varadhan R, et al. Unifying optimization algorithms to aid software system users: optimum for R. J Stat Softw 2011;43(1):1–14.
[18] Nicola M, Alsafi Z, Sohrabi S, Kerwan A, Al-Jabir A, Iosifidis C, et al. The socio-economic implications of the coronavirus and COVID-19 pandemic: a review. Int J Surg 2020.
[19] Petropoulos F, Makridakis S. Forecasting the novel coronavirus COVID-19. PLoS ONE 2020;15(3):e0231236.
[20] Rajgor DD, Lee MH, Archuleta S, Bagdasarian N, Quek SC. The many estimates of the COVID-19 case fatality rate. Lancet Infect Dis 2020.
[21] Slovic P. Perceived risk, trust, and democracy. Risk Anal 1993;13(6):675–82.
[22] Spaanschalski P, Blažynska-Spajchaslna A, Kobiela J. Estimating case fatality rates of COVID-19. Lancet Infect Dis 2020.
[23] Van Dorp JRF, Krotz S. Generalized trapezoidal distributions. Metrika 2003;58(1):85–97.
[24] Van Dorp JRF, Ramdou SC, Pérez JC, Pleguezuelo RH. An elicitation procedure for the generalized trapezoidal distribution with a uniform central stage. Deci Analy 2007;4(3):156–66.
[25] Yang S, Cao P, Du P, Wu Z, Zhuang Z, Yang L, et al. Early estimation of the case fatality rate of COVID-19 in mainland china: a data-driven analysis. Ann Transl Med 2020;8(4).
[26] Zhang D, Hu M, Ji Q. Financial markets under the global pandemic of COVID-19. Finance Res Lett 2020;101528.