



Faculty of Sciences
Department of Mathematics & Statistical Sciences

Modelling COVID-19 Pandemic in Southern African Region

by

Violet Zivai Nyamajiwa

Reg. No: 21100064

MSc(Statistics), BIUST, Botswana

A Dissertation Submitted to the Faculty of Sciences in Partial Fulfilment of the
Requirements for the Award of the Degree of Master of Science in Statistics of BIUST

Supervisor(s):

Dr. Lesego Gabaitiri

Department of Mathematics & Statistical Sciences,

Faculty of Sciences, BIUST

E-mail: gabaitiril@biust.ac.bw

Prof. Broderick Oluyede

Department of Mathematics & Statistical Sciences,

Faculty of Sciences, BIUST

E-mail: oluyedeo@biust.ac.bw

Dr. Onkabetse Vincent Mabikwa

Department of Mathematics & Statistical Sciences,

Faculty of Sciences, BIUST

E-mail: mabikwa@biust.ac.bw

October 21, 2024

DECLARATION REGARDING THE WORK AND COPYRIGHT

Candidate Violet Z. Nyamajiwa

Student ID: 21100064

Thesis Titled: Modelling COVID-19 Pandemic in Southern African Region

I **Violet Z. Nyamajiwa**, certify that the Thesis is all my own original work and that I have not obtained a degree in this University or elsewhere on the basis of any of this work.

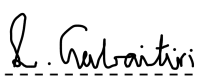
This dissertation/thesis is copyright material protected under the Berne Convention, the Copyright and Neighboring Rights Act, Act. No. 8 of 2000 and other international and national enactments, in that behalf, on intellectual property. It must not be reproduced by any means, in full or in part, except for short extracts in fair dealing; for researcher private study, critical scholarly review or discourse with an acknowledgement, without the written permission of the office of the Postgraduate School, on behalf of both the author and the BIUST.

Signed: 

Date: 21/10/2024

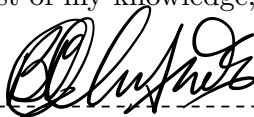
Primary Supervisors Dr. Lesego Gabaitiri, Prof. Broderick Oluyede and Dr. Onkabetse V. Mabikwa

I **Dr. Lesego Gabaitiri**, hereby confirm that I have inspected the above titled thesis and, to the best of my knowledge, it is based on the original work of the candidate.

Signed: 

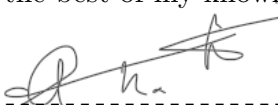
Date: October 23, 2024

I **Prof. Broderick Oluyede**, hereby confirm that I have inspected the above titled thesis and, to the best of my knowledge, it is based on the original work of the candidate.

Signed: 

Date: OCTOBER 23, 2024

I **Dr. Onkabetse V. Mabikwa**, hereby confirm that I have inspected the above titled thesis and, to the best of my knowledge, it is based on the original work of the candidate.

Signed: 

Date: October 23, 2024

Abstract

The global impact of COVID-19 incidence and mortality propelled world leaders under the guidance of the World Health Organization to agree on strategies and goals to reduce infections and mortality. The Southern African region was more vulnerable to the disease due to the high prevalence of comorbidities. Hence, it is crucial to estimate incidence, mortality, and administered vaccinations to continuously monitor this disease. This estimation generally requires a suitable distribution. Failure to find a suitable distribution can result in wrong inferences. Thus, this research aimed to identify generalized distributions that can best fit and model the COVID-19 data in the Southern African region. Graphical and numerical summaries were used to determine the most suitable probability distributions for modelling COVID-19 data in Southern Africa. Two generalized distributions and their sub-models, as well as the two classical distributions, were evaluated to estimate parameters and identify the most suitable distribution for describing the characteristics of COVID-19 data in the region. The maximum likelihood estimation method was used to estimate the unknown parameters of the models. The results indicated that the Type I Heavy-Tailed Log-Logistic distribution provided a better fit for COVID-19 weekly incidence, mortality, and administered vaccinations in Southern Africa and at the seasonal level compared to its competitors. However, the Topp-Leone Half Logistic-Odd Burr X-Log-Logistic distribution offered a better fit for incidence in the spring period. This study provides valuable insights in modelling heavy-tailed/non-monotonic distributions such as the COVID-19 datasets used in this research.

keywords: COVID-19, Goodness-of-Fit, Probability Distributions, Maximum Likelihood Estimation and Southern Africa.

Dedications

To my family, my late mother and brother. I love you deeply.



Acknowledgements

Firstly, I would like to thank the Almighty God for his guidance and protection during my study at the Botswana International University of Science and Technology. It was a very tough road for me, but God made it possible, and I am forever grateful to him. I would also like to extend my profound gratitude to my academic supervisors, Dr. L. Gabaitiri, Prof. B. Oluyede, and Dr. O. V. Mabikwa, for their efforts in overseeing my research, their guidance, their unwavering support, and for giving me valuable feedback, which made it possible for this research to be completed. I am deeply indebted to my kids and my grandmother (Theresa) for this great achievement, which is due to their love, support, and encouragement throughout this journey.

—God Bless you all



Contents

Declaration Regarding the Work and Copyright	i
Contents	vi
List of Figures	vii
List of Tables	viii
List of Abbreviations	ix
1 Introduction and Literature Review	1
1.1 Introduction	1
1.2 Problem Statement	4
1.3 Aim of the Research	5
1.4 Objectives	5
1.5 Justification of the Study	5
2 Graphical and Numerical Summary	6
2.1 Introduction	6
2.2 Data Description	6
2.3 Graphical Presentation of the Data	6
2.3.1 Time Series Plots	7
2.3.2 Box and Q-Q Plots	9
2.4 Summary Statistics	12
2.4.1 Summary Statistics for COVID-19 Data in Southern Africa	13
2.5 Chapter Summary	15
3 Generalized Families of Distributions	16
3.1 Introduction	16
3.2 Generalized Families of Distributions	16
3.3 Fitting Weekly COVID-19 Incidence, Mortality and Administered Vaccinations Data	17
3.3.1 Type-I Heavy-Tailed Exponentiated Generalized-Log-Logistic (TIHTEG-LL) Distribution	18
3.3.2 The Topp-Leone-Exponentiated Half Logistic-Odd Burr X-Log-Logistic (TLEHLOBX-LL) Distribution	19
3.3.3 Log-Logistic	21
3.3.4 Weibull	21
3.4 Parameter Estimation	21

3.4.1	Maximum Likelihood Estimation (MLE)	21
3.5	Statistical Analysis	22
3.5.1	Goodness-of-fit Statistics	22
3.5.2	Graphical Assessments	23
3.6	Chapter Summary	24
4	Results and Discussions	25
4.1	Introduction	25
4.2	Results for Incidence in Southern African Region	25
4.2.1	Estimates and Model Selection Criteria for Incidence in Autumn	28
4.2.2	Estimates and Model Selection Criteria for Incidence in Winter	30
4.2.3	Estimates and Model Selection Criteria for Incidence in Spring	32
4.2.4	Estimates and Model Selection Criteria for Incidence in Summer	34
4.3	Results for Mortality in Southern African Region	35
4.3.1	Estimates and Model Selection Criteria for Mortality in Autumn	38
4.3.2	Estimates and Model Selection Criteria for Mortality in Winter	40
4.3.3	Estimates and Model Selection Criteria for Mortality in Spring	42
4.3.4	Estimates and Model Selection Criteria for Mortality in Summer	44
4.4	Results for Administered Vaccinations in Southern African Region	45
4.5	Discussions	47
5	Conclusions	49
5.1	Introduction	49
5.2	Conclusions	49
5.3	Limitations	50
5.4	Areas for Further Research	50
	Appendix	51
A	R Code for Chapter Two on Southern Africa Data	51
B	R Code for Chapter Two on Southern Africa Data	55

List of Figures

2.1	Time series plot of the weekly incidence and mortality for COVID-19 in Southern Africa.	7
2.2	Time series plot of the administered vaccinations for COVID-19 in Southern Africa.	7
2.3	Monthly seasonal index plot for weekly COVID-19 incidence and mortality in southern Africa.	8
2.4	Box-plots for weekly COVID-19 reported incidence, mortality and administered vaccinations in Southern Africa.	9
2.5	Q-Q plots for weekly COVID-19 reported incidence, mortality and administered vaccinations in Southern Africa.	10
2.6	Box-plots for seasonal COVID-19 incidence in southern Africa.	10
2.7	Q-Q plots for seasonal COVID-19 reported incidence in southern Africa.	11
2.8	Box-plots for seasonal COVID-19 mortality in Southern Africa.	11
2.9	Q-Q plots for seasonal COVID-19 mortality in Southern Africa.	12
3.1	Density Plots for the TLEHLOBX-LL and TIHTEG-LL Distributions.	17
4.1	Fitted densities, Ecdf, Kaplan-Meier Survival Curve, and Hazard Rate Function Plot for Incidence in Southern Africa.	27
4.2	Fitted densities, Ecdf, Kaplan-Meier Survival Curve, and Hazard Rate Function Plot for Incidence in Autumn.	29
4.3	Fitted densities, Ecdf, Kaplan-Meier Survival Curve, and Hazard Rate Function Plot for Incidence in Winter.	31
4.4	Fitted densities, Ecdf, Kaplan-Meier Survival Curve, and Hazard Rate Function Plot for Incidence in Spring.	33
4.5	Fitted densities, Ecdf, Kaplan-Meier Survival Curve, and Hazard Rate Function Plot for Incidence in Summer.	35
4.6	Fitted densities, Ecdf, Kaplan-Meier Survival Curve, and Hazard Rate Function Plot for Mortality in Southern Africa.	37
4.7	Fitted densities, Ecdf, Kaplan-Meier Survival Curve, and Hazard Rate Function Plot for Mortality in Autumn.	39
4.8	Fitted densities, Ecdf, Kaplan-Meier Survival Curve, and Hazard Rate Function Plot for Mortality in Winter.	41
4.9	Fitted densities, Ecdf, Kaplan-Meier Survival Curve, and Hazard Rate Function Plot for Mortality in Spring.	43
4.10	Fitted densities, Ecdf, Kaplan-Meier Survival Curve, and hazard rate function plot for mortality in summer.	45
4.11	Fitted densities, Ecdf, Kaplan-Meier Survival Curve, and Hazard Rate Function Plot for Administered Vaccinations in Southern Africa.	47

List of Tables

2.1	Summary statistics for COVID-19 incidence, mortality and administered vaccinations in Southern Africa.	13
2.2	Summary statistics for seasonal COVID-19 incidence in Southern Africa.	14
2.3	Summary statistics for seasonal COVID-19 mortality in Southern Africa.	14
4.1	Estimated parameters for the fitted distributions.	25
4.2	GoF statistics for COVID-19 incidence in Southern Africa.	26
4.3	Estimated parameters for the fitted distributions.	28
4.4	GoF statistics for COVID-19 incidence in autumn.	28
4.5	Estimated parameters for the fitted distributions.	30
4.6	GoF statistics for COVID-19 incidence in winter.	30
4.7	Estimated parameters for the fitted distributions.	32
4.8	GoF statistics for COVID-19 incidence in spring.	32
4.9	Estimated parameters for the fitted distributions.	34
4.10	GoF statistics for COVID-19 incidence in summer.	34
4.11	Estimated parameters for the fitted distributions.	36
4.12	GoF statistics for COVID-19 mortality in Southern Africa.	36
4.13	Estimated parameters for the fitted distributions.	38
4.14	GoF statistics for COVID-19 mortality in autumn.	38
4.15	Estimated parameters for the fitted distributions.	40
4.16	GoF statistics for COVID-19 mortality in winter.	40
4.17	Estimated parameters for the fitted distributions.	42
4.18	GoF statistics for COVID-19 mortality in spring.	42
4.19	Estimated parameters for the fitted distributions.	44
4.20	Distribution fitting to COVID-19 mortality in summer.	44
4.21	Estimated parameters for the fitted distributions.	46
4.22	GoF statistics for COVID-19 administered vaccinations in Southern Africa.	46

List of Abbreviations

AIDS	Acquired Immune Deficiency Syndrome
A*	Anderson-Darling
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
CAIC	Corrected Akaike Information Criterion
cdf	Cumulative Distribution Function
COVID-19	Corona Virus Disease-2019
GoF	Goodness-of-fit
HIV	Human Immunodeficiency Virus
K-S	Kolmogorov-Smirnov
K-M	Kaplan-Meier
LL	log-likelihood
pdf	Probability Density Function
Q-Q	Quantile-Quantile
SI	Seasonal Index
SARS-CoV-2	Severe acute respiratory syndrome corona virus 2
SACU	Southern African Customs Union
WHO	World Health Organisation

1. Introduction and Literature Review

1.1 Introduction

The unprecedented emergence of the coronavirus disease (COVID-19) plagued humanity worldwide. The severe acute respiratory syndrome coronavirus (SARS-CoV-2) emerged in December 2019 from Wuhan, China (WHO, 2020), and swiftly propagated across all continents. The emergence of COVID-19 in Africa exacerbated an already precarious situation in a region battling with poor health systems, poverty, and infectious illnesses such as HIV/AIDS, malaria and cholera. In light of these contributing factors, the continent was more vulnerable to the pandemic than other continents (Mthethwa et al., 2022; Lone and Ahmad, 2020). The initial occurrence of COVID-19 in the Southern Africa region was documented in South Africa on the 5th of March 2020 (Lone and Ahmad, 2020). Subsequently, the first fatality related to the virus was recorded in Zimbabwe on the 23rd of March 2020 (Nyabunze and Siavhundu, 2020). The virus spread around the Southern African countries in less than three months, with Lesotho being the last to report its first COVID-19 case and death on the 13th of May 2020 and the 9th of July 2020, respectively (De Alencastro et al., 2020). It is noteworthy that the emergence of such a pandemic requires researchers to study the trends and behavior of the disease with the objective of acquiring a comprehensive understanding of the prevailing circumstances. As such, researchers have been applying statistical and mathematical modelling techniques to provide policy-makers and medical scientists with solid data-driven insights and non-pharmaceutical interventions to better manage the pandemic. Many countries around the world implemented non-pharmaceutical interventions, such as social isolation, mandatory mask use, and movement restrictions (e.g., lockdowns), to effectively mitigate the virus transmission. However, the public measures in the Southern African region were coupled with a lot of challenges, including failure to comply with lockdowns, due to the diverse and contrasting living conditions. Many countries in the region have a high portion of the informal sector, where individuals depend on daily wages to sustain themselves and their families (Stiegler and Bouchard, 2020). The easing of lockdowns, as reported in Kuwonu (2020), coupled with the emergence of new variants, including the beta variant initially detected in South Africa (Tegally et al., 2021) and the more contagious delta variant, worsened the severity of the disease in the region (WHO, 2022).

To reduce the burden of the disease and related deaths, pharmaceutical interventions, including vaccines, were developed and distributed a few months after the outbreak. These served to manage the illness and slow down the rate of infection, and they are still being distributed. Following guidance from the World Health Organization, South Africa and Zimbabwe took the lead as the initial countries in the Southern African region to initiate the vaccination program on the 17th and 18th of February 2021, respectively (Dzinamarira et al., 2021). Since then, there has been an observed decline in the overall number of reported incidence and mortality. Currently, life is almost back to normal; countries have lifted their COVID-19 restrictions and reopened their

economies. Even though COVID-19 is no longer considered a global public health emergency, people are still contracting the illness and continue to lose their lives due to the disease. One of the reasons some countries are still experiencing these COVID-19 challenges is, among others, lack of proper vaccination programs (Ayenigbara et al., 2021). This slow progress in vaccination has prolonged the pandemic and led to the emergence of new variants that evade the immunity conferred by vaccines.

COVID-19 necessitates a thorough assessment of the extent and severity of its spread in the region. Thus, given the inherent uncertainty associated with such a complex and rapidly evolving disease, it is imperative to employ probabilistic methods that are both timely and efficient in order to model and analyse the patterns exhibited by the available data on COVID-19. As highlighted by Daniel et al. (2022), this approach enables a comprehensive understanding of the characteristics of the data. Furthermore, Olumide et al. (2013) noted that probability distributions provide appropriate methodologies for describing a wide range of natural, health and societal issues, making them particularly relevant for analysing COVID-19 data. Thus, by utilizing probability distributions, researchers and policy-makers can acquire valuable knowledge about the dynamics of the disease, assisting in the development of effective mitigation and control strategies.

To date, several techniques have been developed or applied to examine the distribution and trends of COVID-19 in various regions across the globe. For instance, Gholami et al. (2020) discovered that the Weibull distributions outperformed the Log-normal and Normal distributions in fitting daily COVID-19 incidence and mortality in Iran. Ahsan-ul Haq et al. (2022) conducted a study to compare different statistical distributions for modelling COVID-19 cases in Pakistan. The distributions examined were the Weibull, Power Function, Log-Logistic, Log-Normal, Inverse Weibull, Gumbel, Burr III, Burr XII, and Birnbaum-Saunders distributions. The findings indicated that the Weibull distribution was the most accurate in fitting the observed data. Furthermore, Ramadhan et al. (2023) used the Weibull, Gamma, Log Normal, Amarendra, Rani, and Akash statistical distributions to examine the frequency of COVID-19 deaths in Indonesia. According to the results of their study, the Weibull distribution best fits the observed data. Although the Weibull distribution was found to be the best fit for these studies, it is frequently criticized in the literature for its inability to accurately capture the unique characteristics of non-monotone hazard functions. As a result, this limitation can lead to poor model fit and inaccurate inference in such situations. Daniel et al. (2022) evaluated the effectiveness of the exponential gamma distribution, along with the pre-existing gamma and exponential distributions. The primary objective was to identify the distribution that provided the most accurate fit for COVID-19 fatalities in Nigeria during the period from March 2020 to April 2021. The findings revealed that the exponential gamma distribution exhibited superior performance compared to the existing distributions. Chikobvu and Sigauke (2020) conducted an analysis on COVID-19 deaths in South Africa using four classical statistical distributions, namely the generalized Gamma distribution, which includes the Gamma, Weibull, Exponential, and Lognormal distributions. The empirical findings indicate that the Gamma distribution

demonstrated the highest level of accuracy in fitting the observed data.

Other notable research in other disciplines of statistics in modelling and analysing data on COVID-19 includes the susceptible, infectious, and recovered (SIR) model, which is widely regarded as a gold standard for understanding communicable disease transmission and has been extensively applied in COVID-19 pandemic modelling (Dolton, 2021). Many studies have utilized the SIR model to investigate COVID-19 dynamics, including the work by Sinkala et al. (2020), who predicted the temporal progression of COVID-19 transmission in Africa and individual countries using the susceptible-infectious-recovered mathematical model. The findings indicated that there is a probable increase in monthly COVID-19 cases expected for June 2020 compared to the preceding months. Variations of the SIR model, such as the susceptible, exposed, infectious, and recovered (SEIR) model was used by Zhao et al. (2020b) to predict the outcomes of the three distinct intervention strategies, namely suppression, mitigation, and mildness, in South Africa, Egypt, and Algeria. Ndlovu et al. (2022) also modified the susceptible vaccinated, infected, and recovered (SVIR) model and integrated the moving displacements and signal approaches to account for the impact of seasonal patterns. As a result of the adjustments, the model successfully predicted the timing and magnitude of the fourth wave. Another extension was pursued by Amouzouvi et al. (2021) and Giordano et al. (2020) who applied the susceptible, infected, diagnosed, ailing, recognized, threatened, healed, and extinct (SIDARTHE) modelling technique to study COVID-19 in selected African Countries and Italy, respectively. However, these dynamical system models rely heavily on adequate data for model validation; thus, due to limited data availability during the initial phases of disease progression, particularly in developing countries, models may produce estimates that are not reliable. Hence, parameters could have not been accurately determined which leads to questionable prediction results.

A study by Alsulami et al. (2021) found higher COVID cases in high-population-density nations (France, Germany, and Italy) using decision trees statistical models. Zhang (2021) used an ARIMA model to predict COVID-19 incidences and fatalities in the US and Europe but found that the actual data after January 2021 was below the forecast values, possibly due to a failure to use the time series stationary transformations to make the data trend-free before modelling. Subsequently, Maleki et al. (2020) developed a two-piece scale mixture of normal (TP-SMN) to predict COVID-19 transmission and death rates. The proposed TP-SMN model outperformed the conventional time-series model. However, the model only focused on total cases and deaths, neglecting other measures like total recoveries, etc., which could make the model more practical. Achoki et al. (2020) utilized a covariate-based instrumental variable regression model to estimate COVID-19 biomarkers in Africa, including prevalence, incidence, and mortality rate. However, the lack of key covariates and the fact that some countries had not yet reached a stable infection rate were major drawbacks of the model, making future trend modelling an open-ended exercise.

In addition, El-Morshedy et al. (2022) developed a counting regression model called Discrete Generalized Lindley using daily coronavirus infections and fatalities. The model accommodates

under-dispersion, equi-dispersion, and over-dispersion. The study used three COVID-19 data sets with excessive over-dispersion to evaluate the model's performance. Interestingly, the study found that countries with high alcohol consumption have lower death rates. Akhtar (2020) used statistical methods such as the probability density function, changing point detection, and polynomial fitting to analyze the COVID-19 pandemic curve and successfully estimated the infected population over 30 days. Machado and Lopes (2020) used heuristic model fitting to assess the transmission of COVID-19 and compare countries through hierarchical clustering and multidimensional scaling. The study revealed patterns that highlighted similarities and differences between countries, identifying key features of the complex dynamics. Tawiah et al. (2021) employed zero-dominated models to examine the progression of COVID-19 fatalities in Ghana. Their analysis revealed that the negative binomial autoregressive technique provided the most accurate fit and projected a significant increase in deaths shortly after November 30, 2020, followed by a rapid decline. Mthethwa et al. (2022) developed a Markov Switching (MS)-volatility model incorporating heavy-tailed distributions to evaluate the risk of SARS-CoV-2 mortality in South Africa. They discovered that combining the MS (3)-GARCH (1,1) model with long-tailed distributions was an efficient method for estimating the likelihood of COVID-19 deaths.

From the literature reviewed thus far in the framework of distribution theory, it is crucial to acknowledge that despite the inability of some classical distributions and dynamical models to capture different characteristics of data, researchers continue to rely on classical distributions. Therefore, in this study, we will explore some classical and generalized distributions to find a model that provides a better fit for the data under consideration.

1.2 Problem Statement

Application of probability distributions in health care studies could be useful to other researchers and critical for decision-making purposes. Hence, it is important to correctly apply these models. The challenge is always identifying the distribution that can suitably model skewed or non-monotonic datasets, such as COVID-19. Thus, studies in the field of distribution theory are few and have not been widely used to analyse COVID-19 data (Ramadhan et al., 2023). This raises the need to focus on data-driven distributions when modelling COVID-19 disease. Therefore, this study seeks to identify generalized distributions that can effectively capture and adequately describe the unique characteristics of COVID-19 datasets in Southern Africa. The incidence, mortality and vaccination datasets will be examined. This involves identifying and examining contemporary distributions that offer improved modelling capabilities in the analysis of COVID-19 data. This research will enhance and help with understanding the complex dynamics surrounding the pandemic and aid in developing statistical models that are more precise and reliable for similar datasets.

1.3 Aim of the Research

This research aims to identify generalized distributions that can best fit and model the COVID-19 data in Southern Africa.

1.4 Objectives

The specific objectives of this study are to:

- identify the types of distributions required to model COVID-19 datasets in Southern Africa.
- estimate parameters and fit the identified models.
- evaluate the selected distributions, and select the model that provides the best fit.

1.5 Justification of the Study

The results of the COVID-19 analysis have been widely presented and reported in the literature (Gholami et al., 2020; Daniel et al., 2022; Ramadhan et al., 2023), highlighting the importance of modelling this disease. Despite the numerous excellent research publications in the framework of distribution theory most researchers rely on classical distributions, and there are few studies in the literature on the application of generalized classes and families of distributions. Hence, generalized distributions were developed to address the shortcomings of these standard models. Generalized distributions are ideal because of their ability to simultaneously account for tail variations and model extreme values, such as those found in COVID-19 datasets. As a result, this study seeks to fill a gap in the field of data-driven distributions, and the findings are expected to be useful in addressing the challenges, making more accurate decisions about COVID-19 disease in the Southern African region, and will serve as a reference point for future studies.

2. Graphical and Numerical Summary

2.1 Introduction

One of the major difficulties faced by researchers is selecting a family of distributions that can fit the data well (Bryson, 1974; Daniel et al., 2022). This stems from the need to accurately represent the underlying patterns and characteristics of the data. Failure to find an appropriate distribution can result in wrong inferences. According to Daniel et al. (2022), the distribution pattern of most health data guides its analysis. Thus, the focus of this chapter is to determine the types of generalized distributions required to model the COVID-19 in the Southern Africa through a comprehensive summary of the data, both graphically and numerically. The summary was divided into two sections. Section 2.3 presents the graphical approach in two parts. In the first sub-section (2.3.1), time-series line plots were used to visually represent the changes in COVID-19 data over time. In the second sub-section (2.3.2), Q-Q plots and bar plots were employed to show concentration, the extreme values, and to determine data distributions. Section 2.4 provides key summary statistics from the COVID-19 data. The R^{C} statistical software, was used for graphical and descriptive analysis.

2.2 Data Description

This study considers the use of online data acquired on the open-source portal Our World in Data (<https://ourworldindata.org/coronavirus>) in collaboration with the World Health Organisation to model COVID-19 in Southern Africa over time. The database includes detailed country-level information such as the date COVID-19 was reported, the nation, the region, new infections, accumulated infections, new mortalities, and accumulated mortalities, to name a few. The Southern African region, as defined by the African Union(<https://au.int/>), comprises Angola, Botswana, Eswatini, Lesotho, Malawi, Mozambique, Namibia, South Africa, Zambia, and Zimbabwe. To have a better understanding of the COVID-19 situation in the region, we aggregated data from these southern African countries into an entity. Weekly incidence and mortality data were extracted from March 2020 to December 2022, while vaccination data were collected from February 2021 to December 2022, with predefined timeframes as shown in the Appendix A. Importantly, this study addresses a significant research gap, as there have been no previous studies specifically focused on modelling COVID-19 in the Southern African region.

2.3 Graphical Presentation of the Data

The section presents the graphical representation of the weekly Southern African COVID-19 data in the form of time series plots, boxplots and Q-Q plots.

2.3.1 Time Series Plots

Figures 2.1 and 2.2 shows an illustration of the reported COVID-19 incidence, mortality and administered vaccinations in Southern Africa. The plots provide a visual representation of how the data changed over time. The X-axis represents the date when the cases, deaths and vaccinations were reported, and the Y-axis represents the number of cases, deaths and administered vaccinations reported, respectively.

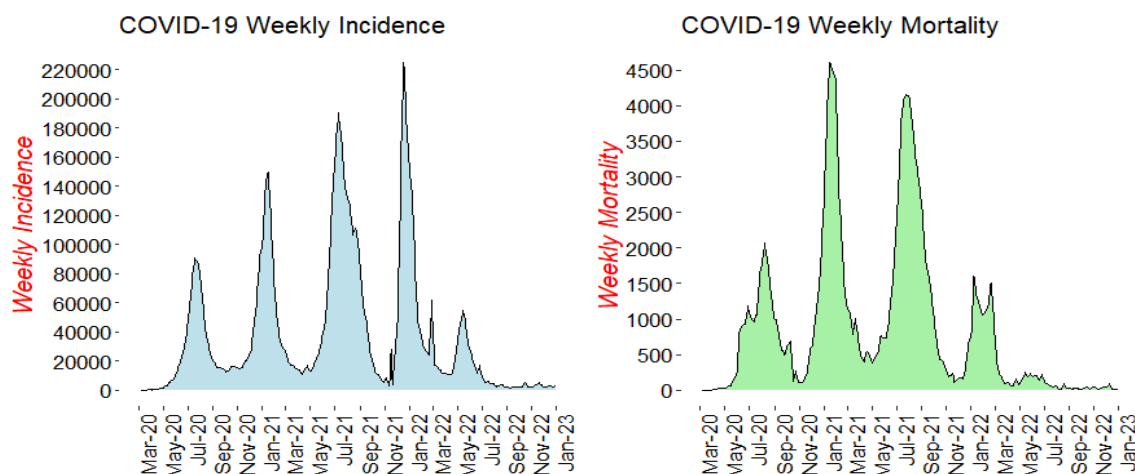


Figure 2.1: Time series plot of the weekly incidence and mortality for COVID-19 in Southern Africa.

Figure 2.1 shows patterns observed in the weekly incidence and mortality in Southern Africa from March 2020 to December 2022. The data reveal that reported incidence and mortality are characterized by peaks and troughs that are regular between periods of rainfall, winter and transition, especially in the months of June, July, August, December, January, and February, where the peaks of the reported cases and deaths correspond to the winter and summer seasons in the Southern Hemisphere (Mckeever, 2023). Notably, in January 2021, the region experienced its highest peak of over 4500 mortalities in a single week.

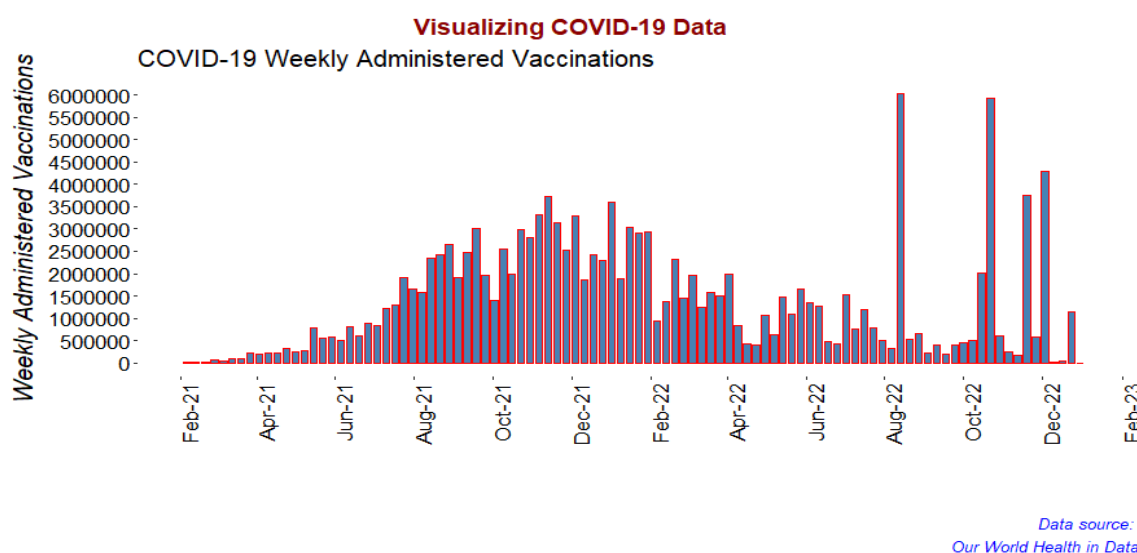


Figure 2.2: Time series plot of the administered vaccinations for COVID-19 in Southern Africa.

Figure 2.2 shows patterns observed in the weekly administered vaccinations in Southern Africa from February 2021 to December 2022. The data revealed an upward trend in reported weekly doses from February 2021 to September 2022, followed by subsequent oscillation patterns. Despite the overall rise in vaccination rates from August 2022 to February 2022, the region encountered high infection rate of about 220,000 in January 2022, as shown in Figure 2.1. However, the highest number of administered doses occurred in the weeks of August 2022 and October 2022, potentially attributed to intensified vaccination campaigns. These efforts have significantly contributed to the reduction in reported cases and deaths across the population of Southern Africa.

2.3.1.1 Seasonal Index

To further understand the weekly incidence and mortality waves shown in Figure 2.1, we calculated the seasonal index (SI) by expressing each monthly average ($\bar{y}_i$) as a percentage of the overall average ($\bar{\bar{y}}$). Thus,

$$\text{Seasonal index for the } i^{\text{th}} \text{ month} = \frac{\bar{y}_i}{\bar{\bar{y}}} \times 100, \text{ for } i = 1, 2, \dots, 12, \quad (2.1)$$

such that an SI value above or below 100 shows whether the observed month had more or fewer cases or deaths than the average level, respectively.

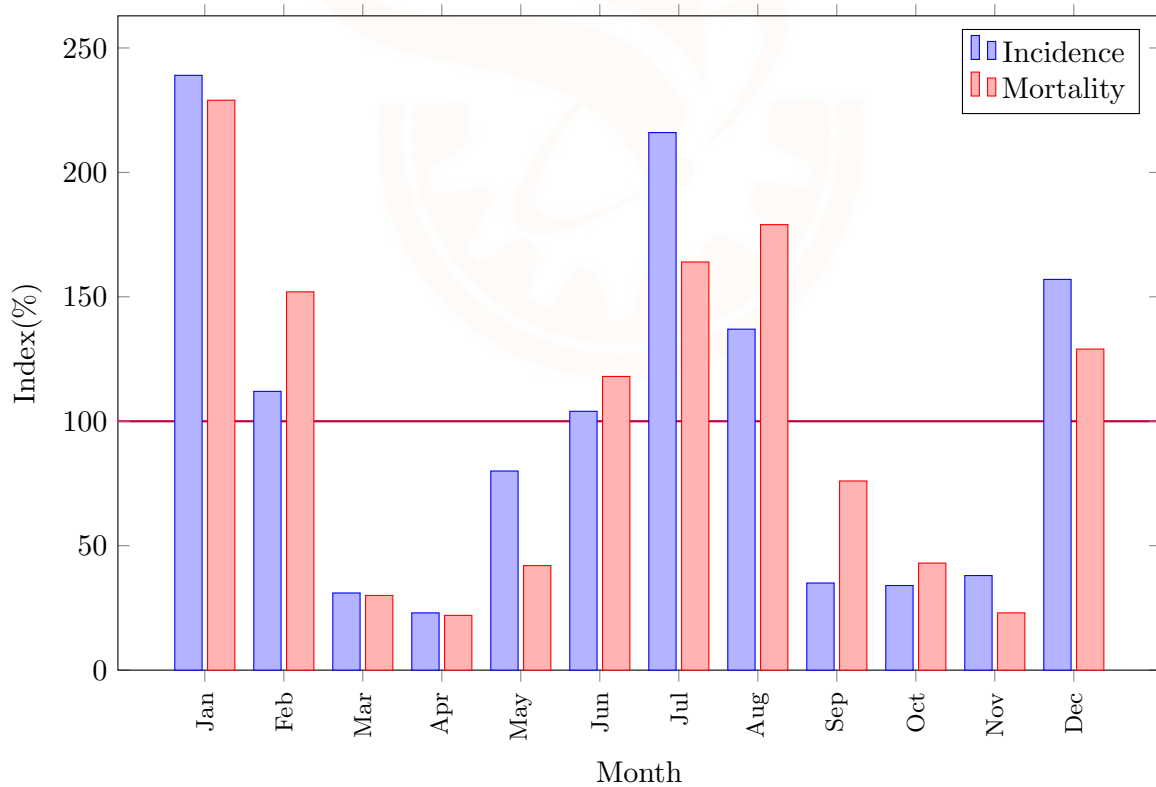


Figure 2.3: Monthly seasonal index plot for weekly COVID-19 incidence and mortality in southern Africa.

Figure 2.3 shows the seasonal index (SI) values for the months in question. The months of January, February, June, July, August, and December have SI values greater than hundred, which indicates these months had higher numbers of cases and deaths across all months, with January and July exhibiting more than 200% of the average for incidence and mortality across

the studied months. This implies a strong seasonal pattern where these months in the Southern African region saw a substantial increase in the number of reported incidence and mortality compared to other months, and these correspond to winter, summer, and transition in the Southern African hemisphere. Hence, based on the observed peaks, the study partitioned the weekly incidence and mortality data into four seasons (autumn, winter, spring, and summer). Accurate peak estimates are crucial for providing policymakers with valuable insights into the distribution of the data.

2.3.2 Box and Q-Q Plots

In this section, we provide visualizations that depict the weekly COVID-19 incidence, mortality, and administered vaccinations in Southern Africa. Note that the normal quantile plot assesses if the data follows a standard normal distribution; the points will fall on the straight line $Y=X$ within the confidence bands ([Wang and Bushman, 1998](#)).

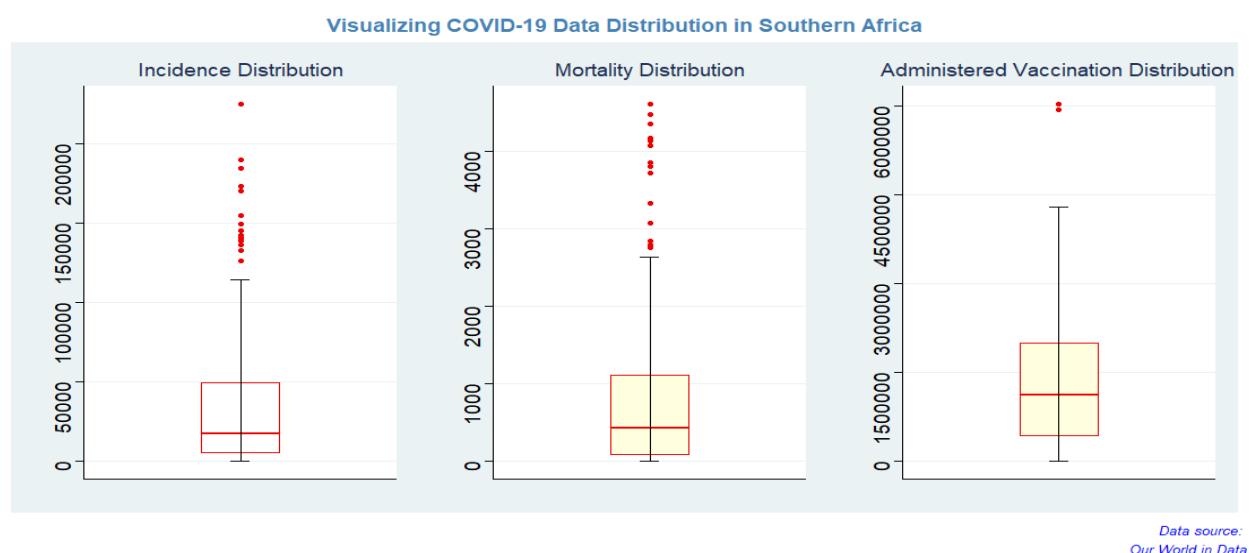


Figure 2.4: Box-plots for weekly COVID-19 reported incidence, mortality and administered vaccinations in Southern Africa.

Figure 2.4 shows the box-plots for the weekly incidence, mortality and administered vaccinations in Southern Africa. The box plots for the datasets under consideration reveal that the median is positioned closer to the lower quartile of the box, and the whiskers are longer on the upper end of the box with the presence of extreme values. This suggests that the data are skewed to the right. Furthermore, the normal quantile plot in Figure 2.5 reflects the reported weekly COVID-19 incidence, mortality, and administered vaccinations. Note that the points do not fall on the straight line and exceed the confidence bands with a U-shaped curve, indicating that the datasets and consideration have an asymmetric distribution that is skewed to the right.

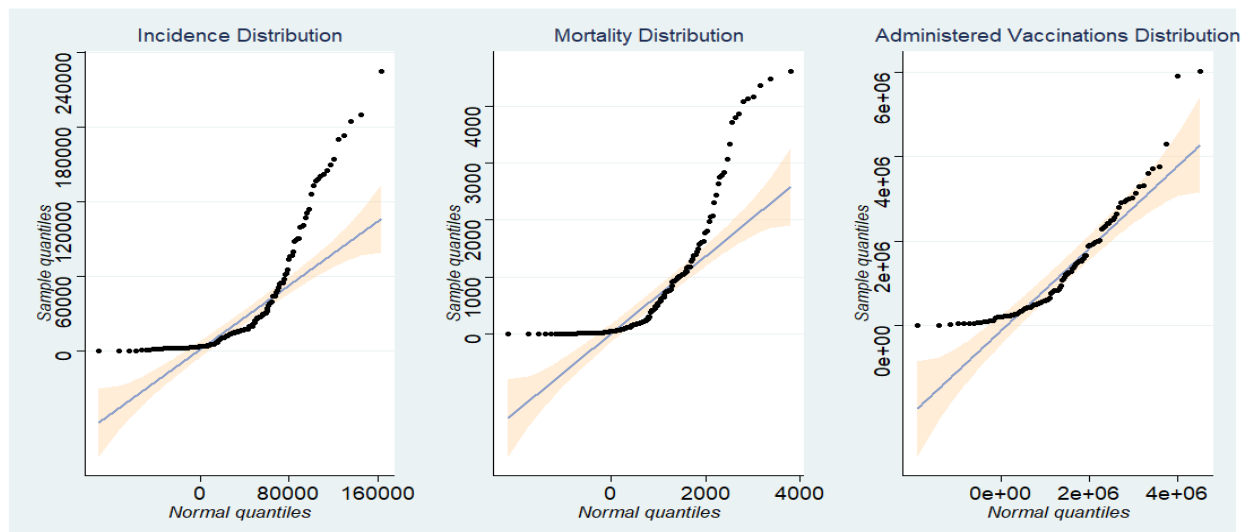
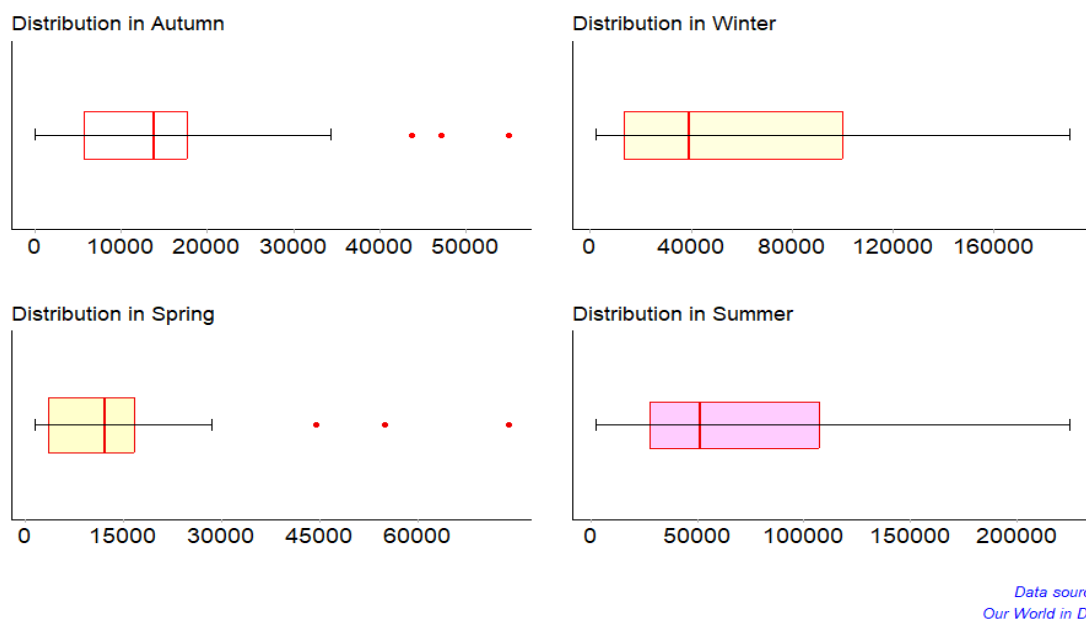


Figure 2.5: Q-Q plots for weekly COVID-19 reported incidence, mortality and administered vaccinations in Southern Africa.

2.3.2.1 Seasonal Plots for COVID-19 Incidence in Southern Africa



Data source:
Our World in Data

Figure 2.6: Box-plots for seasonal COVID-19 incidence in southern Africa.

The box-plots for the seasonal incidence reported in the Southern African region in Figure 2.6 show that the data are skewed to the right. This could be because the upper quartile has longer whiskers as compared to the lower whiskers. Consequently, a majority of the incidence data are concentrated in the lower quartile, resulting in an arithmetic mean value that is higher than the median. The Q-Q plot in Figure 2.7 further confirms that the datasets under consideration are skewed to the right.

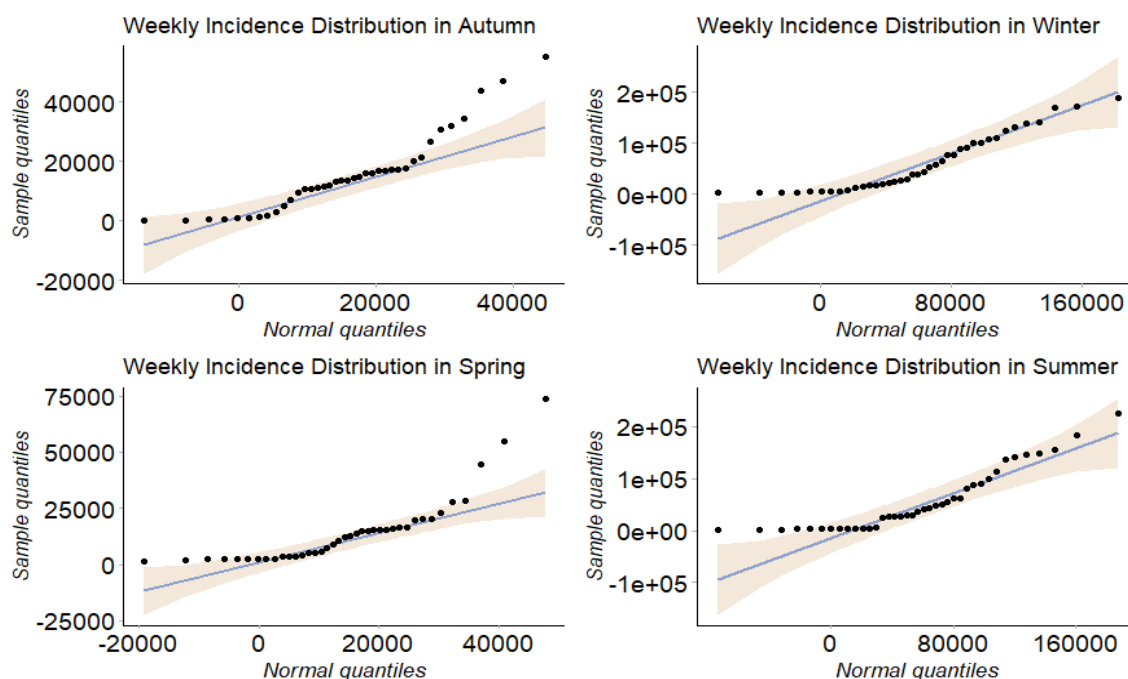


Figure 2.7: Q-Q plots for seasonal COVID-19 reported incidence in southern Africa.

2.3.2.2 Seasonal Plots for COVID-19 Mortality in Southern Africa

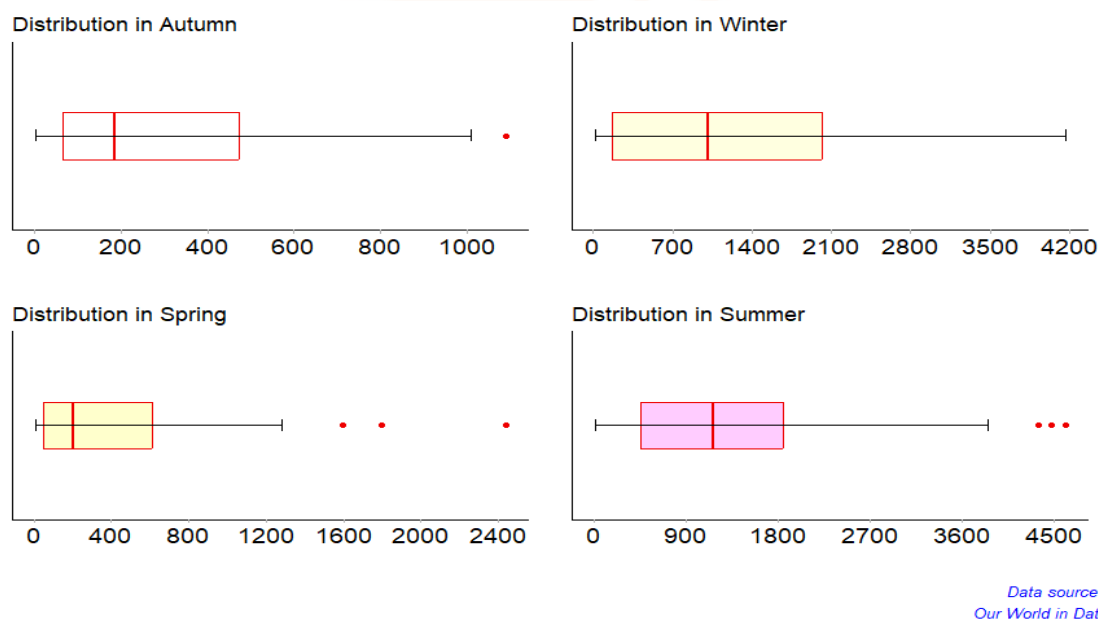


Figure 2.8: Box-plots for seasonal COVID-19 mortality in Southern Africa.

The box-plots for the seasonal mortality reported in the Southern African region in Figure 2.8 show that the data are skewed to the right. This is because the upper quartile has longer whiskers as compared to the lower whisker. Consequently, a majority of the mortality data are concentrated in the lower quartile, resulting in an arithmetic mean value that is higher than the median. Furthermore, the Q-Q plot in Figure 2.9 demonstrates that the data points for seasonal weekly mortality deviate from the line in both the lower and upper tails, providing additional

evidence of the right-skewed nature of the data.

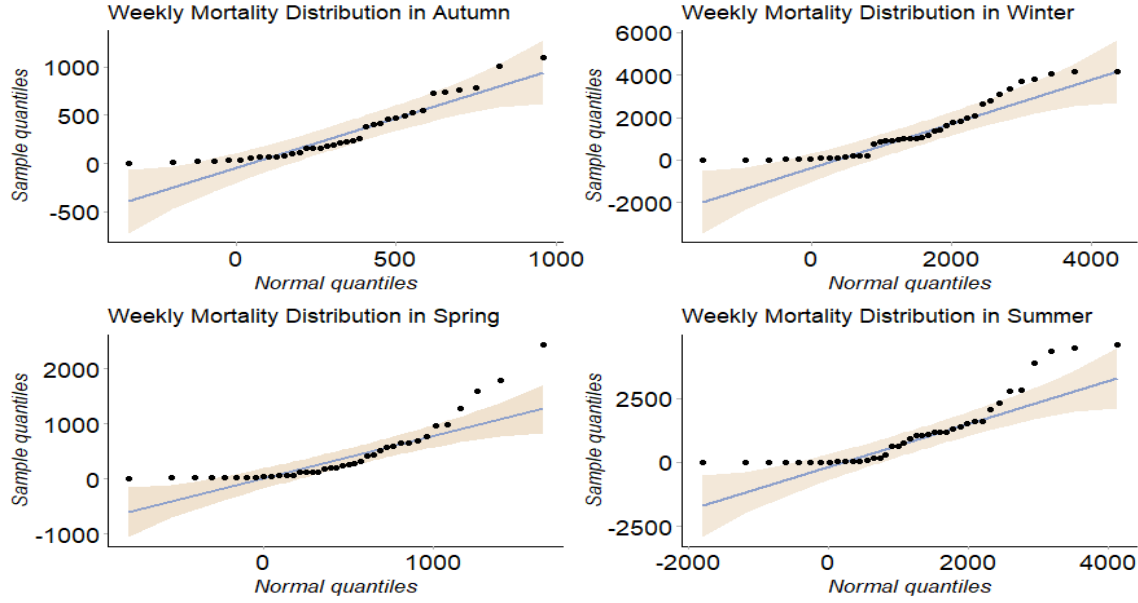


Figure 2.9: Q-Q plots for seasonal COVID-19 mortality in Southern Africa.

2.4 Summary Statistics

According to [Olivier and Norberg \(2010\)](#), data that is symmetrical in nature is usually best represented by using a measure of center (the mean) and a measure of spread (the standard deviation). On the other hand, data that is skewed can be better understood by describing the direction and degree of skewness, which is a concept that is often overlooked. In light of this, the study has chosen to utilize some of these central moments as a statistical characterization method for COVID-19 data. The following summary statistics were computed:

$$\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k \quad ; \text{for } k = 1, \dots, n, \quad (2.2)$$

$$Skew = \frac{\sum_{k=1}^n (x_k - \bar{x})^3}{n \times S^3}, \quad (2.3)$$

$$Kurt = \frac{\sum_{k=1}^n (x_k - \bar{x})^4}{n \times S^4}, \quad (2.4)$$

where :

- $\bar{x}$: the average number of COVID-19 weekly incidence, mortality and administered vaccinations, respectively, over the period under consideration.
- S : a measure of the variability(standard deviation) of the weekly COVID-19 incidence, mortality and administered vaccinations.

- *Skew*: an assessment of the asymmetry of the distribution of the weekly COVID-19 incidence, mortality and administered vaccinations for Southern African region and seasons under study. A positive skew indicates that the distribution is skewed to the right (i.e., there are more days with fewer mortality and fewer days with many mortality), while a negative skew indicates that the distribution is skewed to the left (i.e., there are more days with many deaths and fewer days with fewer deaths).
- *Kurt*: a measure of the “peakedness” of the distribution of the weekly COVID-19 incidence, mortality and administered vaccinations. A higher kurtosis value signifies that the distribution contains a greater number of extreme values compared to a normal distribution, whereas a lower kurtosis value suggests that the distribution has fewer extreme values compared to a normal distribution. Furthermore kurtosis is particularly sensitive to extreme fluctuations, making it an effective indicator of intermittency in pandemic research (Plocoste et al., 2020). Hence, the distribution of the data can be accurately estimated for in-depth statistical descriptions. The expected kurtosis of the normal distribution is $Kurt = 3$. However, the Psych (William Revelle, 2024) package in $R^{\odot}$ Studio returns the excess kurtosis value, which is obtained by subtracting three from the kurtosis value ($Kurt - 3$), yielding the excess kurtosis of zero for a normal distribution.

Subsection 2.4.1 and Sub-subsection 2.4.1.1 and 2.4.1.2 present several summary statistics. In these sections, $\bar{X}$ represents the average of the data, min is the smallest value, Q_1 shows where 25% of the data are at or below a specific value, the median indicates the middle point of the data, Q_3 marks where 75% of the data are at or below a particular value, and max represents the largest value. *Skew* measures the asymmetry in the data, and *Kurt* describes the shape of the distribution of the data.

2.4.1 Summary Statistics for COVID-19 Data in Southern Africa

Table 2.1 shows the overall results of the summary statistics for the weekly COVID -19 incidence, mortality and administered vaccinations in Southern Africa.

Table 2.1: Summary statistics for COVID-19 incidence, mortality and administered vaccinations in Southern Africa.

Season	$\bar{X}$	Min	Q_1	Median	Q_3	Max	Skew	Kurt
Incidence	38338.61	1.00	5307.00	17275.55	49359.00	224890.00	1.83	2.75
Mortality	877.20	3.00	94.00	446.00	1151.00	4608.00	1.84	2.78
Vaccinations	1412784.00	316.00	428302.00	1142481.00	2000526.00	6028223.00	1.38	2.14

The results presented in Table 2.1 show that the Southern Africa region recorded a weekly average of 38,338.61 incidences, 877.20 fatalities, and 1,412,784.00 administered vaccinations for the studied period. The median values for these variables are lower than their respective means, indicating that the data is skewed to the right. The positive coefficients of skewness and excess kurtosis values further indicate that these datasets are right-skewed with more outliers.

Thus, COVID-19 datasets for these periods exhibit heavy-tailed distributions. These findings are consistent with Figure 2.4, which demonstrated that the data are skewed to the right, with extended tails on the right.

2.4.1.1 Summary Statistics for Seasonal COVID-19 Incidence

Table 2.2: Summary statistics for seasonal COVID-19 incidence in Southern Africa.

Season	$\bar{X}$	Min	Q_1	Median	Q_3	Max	Skew	Kurt
Autumn	15437.47	24.00	5671.00	13777.00	17665.00	55050.00	1.22	0.95
Winter	60131.03	2175.00	13338.00	39037.00	100301.00	189892.00	0.73	-0.73
Spring	14265.00	1555.00	3510.00	12198.00	16744.00	74017.00	2.14	5.01
Summer	70517.65	2052.00	27505.00	51145.00	107296.00	224890.00	0.90	-0.63

Table 2.2 summarizes COVID-19 seasonal weekly incidence in the Southern African region. The reported statistics indicate that there is variation in the seasonal time series data, with the summer period having the highest mean number of weekly reported COVID-19 incidence, while the spring and autumn have the lowest weekly confirmed incidence. All the seasons under study exhibit a positive skew, implying that the distribution of the weekly incidence is skewed to the right. The Q-Q plots and box-plots in Figures 2.6 and 2.7 support this observation, indicating asymmetry in all seasons with right displacement. Furthermore, the positive excess kurtosis values for autumn and spring indicate that the incidence of these seasons contains more outliers than would be expected in a normal distribution. On the other hand, the negative excess kurtosis for the winter and summer periods indicates that the incidence datasets have less mass in the tail than the normal distribution.

2.4.1.2 Summary Statistics for Seasonal COVID-19 Mortality

Table 2.3: Summary statistics for seasonal COVID-19 mortality in Southern Africa.

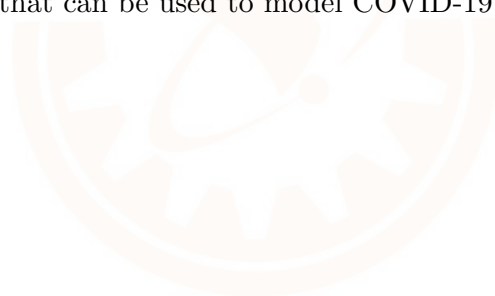
Season	$\bar{X}$	Min	Q_1	Median	Q_3	Max	Skew	Kurt
Autumn	275.80	2.00	37.00	161.00	457.00	1091.00	1.10	0.62
Winter	1403.87	14.00	169.00	1013.00	2016.00	4163.00	0.78	-0.69
Spring	431.23	11.00	48.50	206.00	614.50	2442.00	1.85	3.21
Summer	1458.32	9.00	461.50	1170.00	11844.00	4608.00	1.25	0.70

Table 2.3 shows the seasonal mortality in the region for the period under consideration. The summer season had the highest average mortality of 1458.32, with a median of 1170.00. The winter season followed with an average of 1403.87 confirmed deaths and a median of 1013.00,

while the spring season had an average of 431.23 deaths and a median of 206. Lastly, the autumn period had an average of 275.80 confirmed deaths, with a median of 161. The higher means compared to their medians extrapolate that the mortality in these seasons are right-skewed. Additionally, the skewness values for each season are greater than zero, indicating a positively skewed dataset. This observation aligns with the box and Q-Q plots in Figures 2.8 and 2.9. Furthermore, the autumn and spring seasons have kurtosis values greater than zero, indicating more peaked mortality distributions with more extreme values compared to a normal distribution. Conversely, the winter and summer seasons have negative kurtosis values, suggesting fewer outliers. Therefore, it can be inferred that weekly seasonal mortality datasets are skewed to the right and can be modelled using flexible distributions that can handle both light and heavy-tailed datasets.

2.5 Chapter Summary

This chapter comprehensively described and presented the graphical and numerical summaries of the Southern African COVID-19 datasets. A general outlook reveals a reduction in weekly incidence and fatalities in the Southern African region. This reduction is likely due to accelerated vaccination campaigns. The high skewness and kurtosis in the distribution of incidence, mortality and administrated vaccinations suggest that the data are highly skewed to the right. Hence, the use of models that can accommodate right-skewed data is proposed. The next chapter presents possible distributions that can be used to model COVID-19 datasets.



3. Generalized Families of Distributions

3.1 Introduction

Chapter 2 established that data does not always satisfy the necessary assumptions for normal distributions, despite being a fundamental statistical distribution. As a result, generalized probability distributions that accurately fit and describe the diverse features of the data have been developed, and this process is still ongoing. This chapter describes the generalized families of distributions that captures different shapes and hence are suitable for the different COVID-19 datasets. We also explore the statistical estimation techniques as well as the goodness-of-fit test statistics as applied in Chapter 4 of the thesis. The modeling procedures will also be thoroughly explained, including all the necessary and relevant steps.

3.2 Generalized Families of Distributions

For many years, researchers have been studying generalized families of distributions in various fields of statistics. This is because empirical datasets exhibit traits such as skewness, unimodality, and excessively heavy and light tails (Zhao et al., 2021). The Topp-Leone(TL) (Topp and Leone, 1955) and Half Logistic(HL) (Balakrishnan, 1985), among other classical distributions, are some of the most popular and widely used in the analysis of many scenarios, including modeling survival data, biology, ecology, and economics, just to mention a few. Despite this, it has been discovered that the Topp-Leone (TL) or the Half Logistic (HL) and other classical distributions often fall short of adequately fitting the unique characteristics exhibited by the data (Zhao et al., 2021; Lekono et al., 2024). For example the one parameter TL and HL are limited to a bathtub hazard function and a uni-modal or reversed-J density function shape, respectively. Due to these limitations, researchers have turned to the generalization and development of new families of distributions that offer better fits to various fields and address the drawbacks of classical distributions. Several researchers, including Alzaatreh et al. (2013), Zografos and Balakrishnan (2009), Cordeiro et al. (2013), have proposed techniques for making standard distributions more adaptable to the unique characteristics of the data. Zhao et al. (2020a) proposed the Type I heavy-tailed distribution, and Al-Shomrani et al. (2016) extended the TL distribution to obtain the Topp-Leone-G (TL-G) distribution using the T-X family technique (Alzaatreh et al., 2013). In their article, Zhao et al. (2020a) advocated that the proposed class of generator (Alzaatreh et al., 2013) is very flexible for the any skewed data. Thus, a number of extensions have been put forth, such as the Topp-Leone Odd Exponential Half Logistic-G Family of distributions by Chipepa and Oluyede (2021); the Type I Half Logistic Skew-T distribution: An inverted bathtub-shaped hazard rate heavy-tail model by (Adubisi et al., 2021); and the Topp-Leone Half Logistic Generalized-G Family of distributions by (Moakofi et al., 2021), among others. These new distributions represent a significant improvement in addressing the shortcomings of classical distributions and can offer better fits for real-world data in various applications.

However, [Husak et al. \(2007\)](#) noted that there are contradictions in the literature concerning the optimal distribution best suited to accurately represent real-world data. To analyze data on COVID-19, the families of Topp-Leone-Exponentiated Half Logistic-Odd Burr X-G Family (TLEHLOBX-G) and Type-I Heavy-Tailed Exponentiated Generalized-G (TIHTEG-G) distributions proposed by [Tlhaloganyang et al. \(2022\)](#) and [Lekono et al. \(2024\)](#) were explored with the baseline distribution as the Log-Logistic distribution. The Log-Logistic distribution stands out among the various classical distributions that have been widely applied to many aspects of analysis because of its appealing properties. For instance, the ability of the Log-logistic hazard function to capture the upside-down bathtub shape, as highlighted by [De Santana et al. \(2012\)](#). This characteristic makes it a preferable alternative to other classical distributions like Weibull or exponential distributions. These special families of distributions are a good choice for describing COVID-19 data because of their tremendous flexibility in capturing a variety of distribution shapes, as illustrated in Figures 3.1a and 3.1b. Furthermore, [Lekono et al. \(2024\)](#) and [Tlhaloganyang et al. \(2022\)](#) alluded to the fact that these distributions can be used to describe datasets in many different areas.

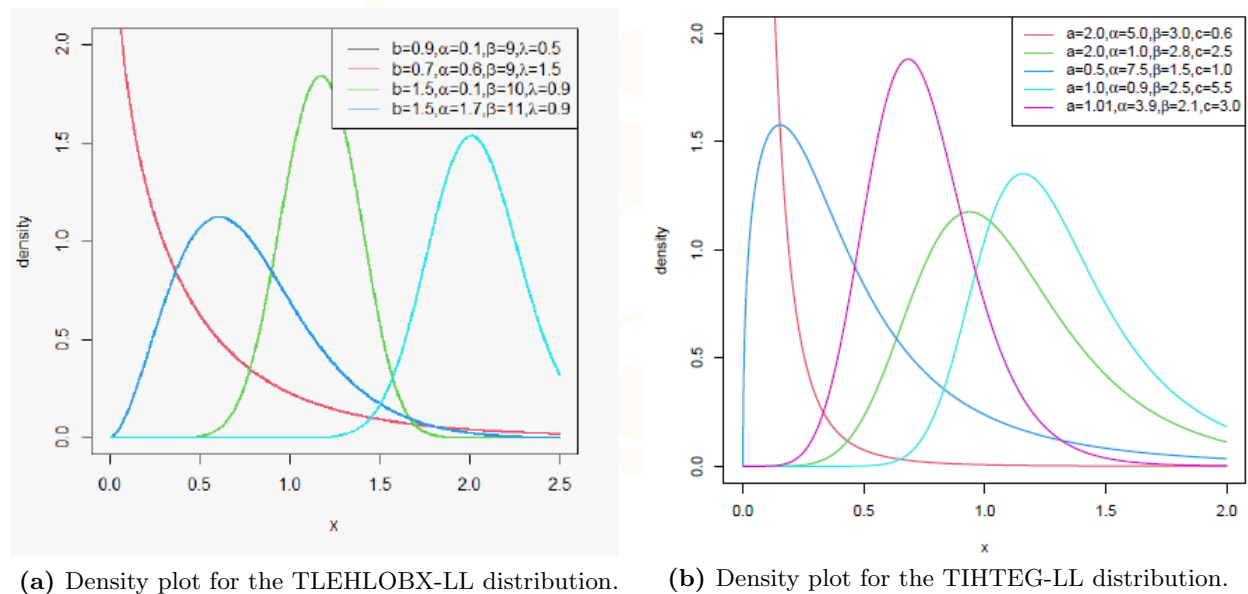


Figure 3.1: Density Plots for the TLEHLOBX-LL and TIHTEG-LL Distributions.

For some statistical properties, such as quantile functions and moments and probability-weighted moments, derivations are provided by [Lekono et al. \(2024\)](#) and [Tlhaloganyang et al. \(2022\)](#).

3.3 Fitting Weekly COVID-19 Incidence, Mortality and Administered Vaccinations Data

To improve the flexibility of fitting real-world data, our analysis is based on distributions specifically developed for this purpose, and that allows for a thorough examination of the likelihood of COVID-19 data on the acquired sample. We employed probability and cumulative density functions as the main tools to characterize the weekly COVID-19 data. Thus, we present the

cumulative distribution function (cdf) and probability density function (pdf) of the recently developed distributions that considered the special baseline distribution of the log-logistic distribution.

3.3.1 Type-I Heavy-Tailed Exponentiated Generalized-Log-Logistic (TIHTEG-LL) Distribution

The pdf and cdf of the TIHTEG-LL distribution by [Lekono et al. \(2024\)](#) is given by:

$$f_{TIHTEG-LL}(x; \rho, \alpha, \beta, \theta) = \frac{\rho^2 \alpha \beta \theta x^{(\theta-1)} [1 + x^\theta]^{-(\alpha+1)} (1 - [1 + x^\theta]^{-\alpha})^{\beta-1}}{\left(1 - \left[(1 - \rho)(1 - [1 + x^\theta]^{-\alpha})^\beta\right]\right)^2} \times \left[\frac{1 - (1 - [1 + x^\theta]^{-\alpha})^\beta}{1 - \left[(1 - \rho)(1 - [1 + x^\theta]^{-\alpha})^\beta\right]} \right]^{\rho-1} \quad (3.1)$$

and

$$F_{TIHTEG-LL}(x; \rho, \alpha, \beta, \theta) = 1 - \left[\frac{1 - (1 - [1 + x^\theta]^{-\alpha})^\beta}{1 - \left[(1 - \rho)(1 - [1 + x^\theta]^{-\alpha})^\beta\right]} \right]^\rho, \quad (3.2)$$

for $x \geq 0$, and $\rho, \alpha, \beta, \theta > 0$.

- When $\beta = 1$, the TIHTEG-LL distribution reduces to the Type I Heavy-Tailed Generalized Log-Logistic (TIHTG-LL) distribution, which has the following pdf and cdf:

$$f_{TIHTG-LL}(x; \rho, \alpha, \theta) = \frac{\rho^2 \alpha \theta x^{(\theta-1)} [1 + x^\theta]^{-(\alpha+1)}}{\left([1 + x^\theta]^{-\alpha} + \rho (1 - [1 + x^\theta]^{-\alpha})\right)^2} \times \left[\frac{[1 + x^\theta]^{-\alpha}}{[1 + x^\theta]^{-\alpha} + \rho (1 - [1 + x^\theta]^{-\alpha})} \right]^{\rho-1}$$

and

$$F_{TIHTG-LL}(x; \rho, \alpha, \theta) = 1 - \left[\frac{[1 + x^\theta]^{-\alpha}}{[1 + x^\theta]^{-\alpha} + \rho (1 - [1 + x^\theta]^{-\alpha})} \right]^\rho, \quad (3.3)$$

for $x \geq 0$, $\rho, \alpha, \theta > 0$.

- By setting $\alpha = 1$, a Type I Heavy-Tailed Log-Logistic (TIHT-LL) distribution was obtained. The pdf and cdf can be expressed as follows:

$$f_{TIHT-LL}(x; \rho, \beta, \theta) = \frac{\rho^2 \beta \theta x^{(\theta-1)} [1 + x^\theta]^{-2} (x^\theta [1 + x^\theta]^{-1})^{\beta-1}}{\left(1 - (1 - \rho) (x^\theta [1 + x^\theta]^{-1})^\beta\right)^2} \times \left[\frac{1 - (x^\theta [1 + x^\theta]^{-1})^\beta}{1 - (1 - \rho) (x^\theta [1 + x^\theta]^{-1})^\beta} \right]^{\rho-1} \quad (3.4)$$

and

$$F_{TIHT-LL}(x; \rho, \beta, \theta) = 1 - \left[\frac{1 - (x^\theta [1 + x^\theta]^{-1})^\beta}{1 - (1 - \rho) (x^\theta [1 + x^\theta]^{-1})^\beta} \right]^\rho, \quad (3.5)$$

for $x \geq 0$, and $\rho, \beta, \theta > 0$.

■ When $\rho = \alpha = 1$, the TIHTEG-LL distribution reduces to the Exponentiated Log-Logistic(E-LL) distribution, which has the following pdf and cdf:

$$f_{E-LL}(x; \beta, \theta) = \beta \theta x^{(\theta-1)} [1 + x^\theta]^{-2} (1 - [1 + x^\theta]^{-1})^{\beta-1}$$

and

$$F_{E-LL}(x; \beta, \theta) = (1 - [1 + x^\theta]^{-1})^\beta, \quad (3.6)$$

for $x \geq 0$, and $\beta, \theta > 0$.

3.3.2 The Topp-Leone-Exponentiated Half Logistic-Odd Burr X-Log-Logistic (TLEHLOBX-LL) Distribution

Thaloganyang et al. (2022) expressed the pdf and cdf of the TLEHLOBX-LL distribution as;

$$f_{TLEHLOBX-LL}(x; \nu, \alpha, \beta, \theta) = \frac{8\nu\alpha\beta\theta x^{2\theta-1}}{\exp(x^{2\theta})} [1 - \exp(-x^{2\theta})]^{\alpha\beta-1} \times \left[2 - (1 - \exp(-x^{2\theta}))^\beta \right]^{-\alpha-1} \left\{ 1 - \left[\frac{(1 - \exp(-x^{2\theta}))^\beta}{2 - (1 - \exp(-x^{2\theta}))^\beta} \right]^\alpha \right\} \times \left\{ 1 - \left[1 - \left[\frac{(1 - \exp(-x^{2\theta}))^\beta}{2 - (1 - \exp(-x^{2\theta}))^\beta} \right]^\alpha \right]^2 \right\}^{\nu-1} \quad (3.7)$$

and

$$F_{TLEHLOBX-LL}(x; \nu, \alpha, \beta, \theta) = \left\{ 1 - \left[1 - \left[\frac{(1 - \exp(-x^{2\theta}))^\beta}{2 - (1 - \exp(-x^{2\theta}))^\beta} \right]^\alpha \right]^2 \right\}^\nu, \quad (3.8)$$

respectively, for $x \geq 0$, and $\nu, \alpha, \beta, \theta > 0$.

- By setting $\beta = 1$, the reduced Topp-Leone-Exponentiated Half Logistic Odd Burr X-Log-Logistic (TLEHLOBX-LL) distribution was obtained and can be expressed as follows:

$$\begin{aligned}
 f_{TLEHLOBX-LL}(x; \nu, \alpha, \theta) &= \frac{8\nu\alpha\theta x^{2\theta-1}}{\exp(x^{2\theta})} \left[1 - \exp(-x^{2\theta})\right]^{\alpha-1} \\
 &\times \left[1 + \exp(-x^{2\theta})\right]^{-\alpha-1} \left\{1 - \left[\frac{(1 - \exp(-x^{2\theta}))}{1 + \exp(-x^{2\theta})}\right]^{\alpha}\right\} \\
 &\times \left\{1 - \left[1 - \left[\frac{(1 - \exp(-x^{2\theta}))}{1 + \exp(-x^{2\theta})}\right]^{\alpha}\right]^2\right\}^{\nu-1}
 \end{aligned} \tag{3.9}$$

and

$$F_{TLEHLOBX-LL}(x; \nu, \alpha, \theta) = \left\{1 - \left[1 - \left[\frac{(1 - \exp(-x^{2\theta}))}{1 + \exp(-x^{2\theta})}\right]^{\alpha}\right]^2\right\}^{\nu}, \tag{3.10}$$

for $x \geq 0$ and $\nu, \alpha, \theta > 0$.

- By setting $\alpha = 1$, The TLHLOBX-LL distribution reduces to the TLHLOBX-LL distribution. The pdf and cdf can be expressed as follows:

$$\begin{aligned}
 f_{TLHLOBX-LL}(x; \nu, \beta, \theta) &= \frac{8\nu\beta\theta x^{2\theta-1}}{\exp(x^{2\theta})} \left[1 - \exp(-x^{2\theta})\right]^{\beta-1} \\
 &\times \left[2 - (1 - \exp(-x^{2\theta}))^{\beta}\right]^{-2} \left\{1 - \left[\frac{(1 - \exp(-x^{2\theta}))^{\beta}}{2 - (1 - \exp(-x^{2\theta}))^{\beta}}\right]\right\} \\
 &\times \left\{1 - \left[1 - \left[\frac{(1 - \exp(-x^{2\theta}))^{\beta}}{2 - (1 - \exp(-x^{2\theta}))^{\beta}}\right]\right]^2\right\}^{\nu-1}
 \end{aligned} \tag{3.11}$$

and

$$F_{TLHLOBX-LL}(x; \nu, \beta, \theta) = \left\{1 - \left[1 - \left[\frac{(1 - \exp(-x^{2\theta}))^{\beta}}{2 - (1 - \exp(-x^{2\theta}))^{\beta}}\right]\right]^2\right\}^{\nu}, \tag{3.12}$$

for $x \geq 0$ and $\nu, \beta, \theta > 0$.

- By setting $b = \alpha = 1$, The TLHLOBX-LL distribution reduces to the Half Logistic Odd Burr X-Log Logistic (HLOBX-LL) distribution. The pdf and cdf can be expressed as follows:

$$\begin{aligned}
 f_{HLOBX-LL}(x; \beta, \theta) &= \frac{8\beta\theta x^{2\theta-1}}{\exp(x^{2\theta})} \left[1 - \exp(-x^{2\theta})\right]^{\beta-1} \\
 &\times \left[2 - (1 - \exp(-x^{2\theta}))^{\beta}\right]^{-2} \left\{1 - \left[\frac{(1 - \exp(-x^{2\theta}))^{\beta}}{2 - (1 - \exp(-x^{2\theta}))^{\beta}}\right]\right\}
 \end{aligned}$$

and

$$F_{HLOBX-LL}(x; \beta, \theta) = 1 - \left[1 - \left(\frac{(1 - \exp(-x^{2\theta}))^\beta}{2 - [1 - \exp(-x^{2\theta})]^\beta} \right) \right]^2, \quad (3.13)$$

respectively, for $x \geq 0$, and $\beta, \theta > 0$.

3.3.3 Log-Logistic

The pdf and cdf of the Log-Logistic distribution are expressed as follows:

$$f_{LL}(x; \theta) = \theta x^{\theta-1} (1 + x^\theta)^{-2} \quad (3.14)$$

and

$$F_{LL}(x; \theta) = 1 - (1 + x^\theta)^{-1}, \quad (3.15)$$

respectively, for $x \geq 0$ and $\theta > 0$.

3.3.4 Weibull

The pdf and cdf of the Weibull distribution are expressed as;

$$f_W(x; \lambda) = \lambda x^{\lambda-1} e^{-x^\lambda} \quad (3.16)$$

and

$$F_W(x; \lambda) = 1 - e^{-x^\lambda}, \quad (3.17)$$

respectively, for $x \geq 0$ and $\lambda > 0$.

3.4 Parameter Estimation

The parameters for the aforementioned distributions were estimated using the maximum likelihood estimation (MLE) technique. The MLE technique was utilized because it has good statistical properties for large samples.

3.4.1 Maximum Likelihood Estimation (MLE)

Let $X_1, X_2, \dots, X_n$, represent an observable sample of size n from the TIEHTG-LL and TLEHLOBX-LL families of pdfs. The likelihood function is defined as follows:

$$L(\Delta) = \prod_{i=1}^n f(x_i; \phi), \quad (3.18)$$

where ϕ represents the vectors of parameters in subsections 3.3.1 and 3.3.2. The parameters were obtained by evaluating the partial derivatives of the logarithmic likelihood function with

respect to each element of the parameter vector. The resulting system of non-linear equations, $\left(\frac{\partial \log L(\Delta)}{\partial \phi}\right)^T = 0$, was solved using iterative techniques such as the Newton-Raphson procedure in the $R^\odot$ software packages namely stats4, stats (R Core Team, 2022), bbmle (Bolker and R Development Core Team, 2022), numDeriv (Gilbert and Varadhan, 2019). This is because these partial derivatives cannot be expressed in closed form.

3.5 Statistical Analysis

We used several numerical including goodness-of-fit statistics, and graphical criteria to compare and evaluate the families of distributions under consideration.

3.5.1 Goodness-of-fit Statistics

In this study, we applied some well-known goodness-of-fit (GoF) statistics such as -2Log-Likelihood function (-2log $\mathcal{L}$), Akaike Information Criterion (AIC), Consistent Akaike Information Criterion (CAIC), Bayesian Information Criterion (BIC)), Anderson-Darling (A-D), and Kolmogorov-Smirnov (K-S) statistic and its p-value to compare the distributions considered. Chen and Balakrishnan (1995) defined these criteria as follows:

- Kolmogorov-Smirnov:

$$K - S = \sup_x \left| \frac{1}{n} \sum_{i=1}^n I\{x_i \leq x\} - \hat{F}(x) \right|, \quad (3.19)$$

where $I\{\cdot\}$ represents the indicator function and $\hat{F}(x)$ denotes the estimate of the cdf $F(x)$.

- Anderson-Darling:

$$A^* = -n - \frac{1}{n} \sum_{i=1}^n (2i - 1) [\ln(F(x_{(i)})) + \ln(1 - F(X_{(n-i+1)}))] , \quad (3.20)$$

where n is the sample size, $F(x_{(i)})$ represents the cdf of the specified distribution, and $x_{(i)}$ denotes the i^{th} sample in ascending order.

- Akaike Information Criterion:

$$AIC = 2q - 2 \log \mathcal{L}, \quad (3.21)$$

where q is the number of estimated parameters in the model and $\mathcal{L}$ is the model maximized value.

- Bayesian Information Criterion:

$$BIC = q \log(n) - 2 \log \mathcal{L}, \quad (3.22)$$

where n is the number of data points observed.

- Consistent Akaike Information Criterion

$$CAIC = -2 \log \mathcal{L} + q [\log(n) + 1]. \quad (3.23)$$

These goodness-of-fit statistics are useful for determining how well the models fit the weekly reported cases, deaths, and administered vaccinations in Southern Africa. In general, an acceptable model or one that outperforms others possesses the smallest value of all GoF statistics and a large K-S p-value.

3.5.2 Graphical Assessments

The effectiveness of the theoretical model was evaluated by superimposing an appropriate distribution curve onto the COVID-19 frequency data. The GoF of the utilized models was graphically assessed using fitted probability densities and Kaplan-Meier survival curves. The survival function for the suitable distribution are as follows:

$$S_{TIHT-LL}(x; \rho, \beta, \theta) = \left[\frac{[1 + x^\theta]^{-\alpha}}{[1 + x^\theta]^{-\alpha} (1 - \rho) + \rho} \right]^\rho, \quad (3.24)$$

for $x \geq 0$, $\rho, \beta, \theta > 0$,
and

$$S_{TLHLOBX-LL}(x; \nu, \beta, \theta) = 1 - \left\{ 1 - \left[1 - \left[\frac{(1 - \exp(-x^{2\theta}))^\beta}{2 - (1 - \exp(-x^{2\theta}))^\beta} \right] \right]^2 \right\}^\nu, \quad (3.25)$$

for $x \geq 0$ and $\nu, \beta, \theta > 0$.

The hazard function, also known as the instantaneous rate of occurrence of the event, is another way to characterize the distribution. It is defined as:

$$h(t) = \frac{f(t)}{1 - F(t)}. \quad (3.26)$$

The hazard function for the TIHT-LL and TLHLOBX-LL models are expressed as :

$$\begin{aligned} h_{TIHT-LL}(x; \rho, \beta, \theta) &= \frac{\rho^2 \alpha \beta \theta x^{(\theta-1)} [1 + x^\theta]^{-(\alpha+1)} (1 - [1 + x^\theta]^{-\alpha})^{\beta-1}}{\left((1 - (1 - \rho) (1 - [1 + x^\theta]^{-\alpha})^\beta) \right)^2} \\ &\quad \times \left[1 - \left[\frac{[1 + x^\theta]^{-\alpha}}{[1 + x^\theta]^{-\alpha} (1 - \rho) + \rho} \right]^\rho \right]^{-1}, \end{aligned} \quad (3.27)$$

for $x \geq 0, \rho, \beta, \theta > 0$,

and

$$\begin{aligned}
 h_{TLHLOBX-LL}(x; \nu, \beta, \theta) &= \frac{8\nu\beta\theta x^{2\theta-1}}{\exp(x^{2\theta})} \left[1 - \exp(-x^{2\theta})\right]^{\beta-1} \\
 &\times \left[2 - \left(1 - \exp(-x^{2\theta})\right)^\beta\right]^{-2} \left\{1 - \left[\frac{(1 - \exp(-x^{2\theta}))^\beta}{2 - (1 - \exp(-x^{2\theta}))^\beta}\right]\right\} \\
 &\times \left\{1 - \left[1 - \left[\frac{(1 - \exp(-x^{2\theta}))^\beta}{2 - (1 - \exp(-x^{2\theta}))^\beta}\right]\right]^2\right\}^{b-1} \\
 &\times \left(1 - \left\{1 - \left[1 - \left[\frac{(1 - \exp(-x^{2\theta}))^\beta}{2 - (1 - \exp(-x^{2\theta}))^\beta}\right]\right]^2\right\}^\nu\right)^{-1}, \quad (3.28)
 \end{aligned}$$

for $x \geq 0, \nu, \beta, \theta > 0$.

3.6 Chapter Summary

In this chapter, we discussed several generalized families of distributions for skewed data. Although classical distributions like the Topp Leone and Half Logistic distribution have been widely used in many areas, they often fall short in fitting diverse datasets, which underscores the importance of developing new distributions. Hence, advanced models such as the Topp-Leone-Exponentiated Half Logistic-Odd Burr X-G and Type-I Heavy-Tailed Exponentiated Generalized-G distributions have made great strides in solving some of these problems, particularly in the context of analysing COVID-19 data. As such, researchers continue to refine and modify these models, and their potential for improved understanding and application of statistical analysis in tackling complex phenomena remains valuable and impactful. Furthermore, we also discussed the estimation technique and the goodness-of-fit measures that were employed to analyse the data.

4. Results and Discussions

4.1 Introduction

This chapter provides a comprehensive analysis and presentation of results based on the procedures described in Chapters 2 and 3. The study utilized secondary data obtained from the number of COVID-19 incidence, mortality, and administered vaccinations reported on a weekly basis in Southern Africa (as stated in Section 2.2). The special families of the Type-I Heavy-Tailed Exponentiated Generalized-Log-Logistics Distribution (3.3.1) and the Topp-Leone-Exponentiated Half Logistic-Odd Burr X-Log-Logistic Distribution (3.3.2) as well as the two classical distributions (Log-Logistic and Weibull) were fitted to the data. The data were analysed using the $R^{\odot}$ software packages namely stats4, stats (R Core Team, 2022), AdequacyModel (Diniz Marinho et al., 2016), bbmle (Bolker and R Development Core Team, 2022), numDeriv (Gilbert and Varadhan, 2019). These models were evaluated using selection criteria such as the Akaike information criterion (AIC), Bayesian information criterion (BIC), and -2 Log-likelihood function ($-2\log \mathcal{L}$). The model with the lowest GoF statistics value was considered the most flexible and the best fit for the data. To provide further clarity and draw more compelling conclusions we utilised the fitted density curve, the empirical cumulative distribution plots, and Kaplan-Meier survival plots.

4.2 Results for Incidence in Southern African Region

This section presents the results obtained from the COVID-19 data for each probability distribution and is presented numerically and graphically for the overall Southern African region.

Table 4.1: Estimated parameters for the fitted distributions.

Distribution	Estimates and Standard Errors			
	ρ	α	β	θ
TIHTEG-LL($\rho, \alpha, \beta, \theta$)	223.890(2.7×10^{-5})	7.149(0.059)	2987.100(2.8×10^{-5})	0.017(0.001)
TIHTG-LL(ρ, α, θ)	0.230(0.037)	0.480(0.512)		1.086(1.141)
TIHT-LL(ρ, β, θ)	6601.900(4.2×10^{-4})		38.574(1.030)	0.053(0.003)
E-LL(β, θ)			34.523(5.225)	0.404(0.018)
	ν	α	β	θ
TLEHLOBX-LL($\nu, \alpha, \beta, \theta$)	0.726(0.081)	9.469(0.006)	38.071(0.001)	0.025(5.9×10^{-4})
TLEHLOBX-LL(ν, α, θ)	0.219(0.090)	257.430(2.5×10^{-4})		0.021(5.1×10^{-4})
TLHLOBX-LL(ν, β, θ)	2.073(0.251)		167.430(0.003)	0.022(4.9×10^{-4})
HLOBX-LL(β, θ)			359.210(4.2×10^{-9})	0.029(3.6×10^{-4})
				θ
LL(θ)				0.159(0.010)
				λ
W(λ)				0.082(0.004)

Table 4.2: GoF statistics for COVID-19 incidence in Southern Africa.

Distribution	Model Selection Criteria						
	$-2\log \mathcal{L}$	AIC	$CAIC$	BIC	A^*	K-S	P-Value
TIHTEG-LL($\rho, \alpha, \beta, \theta$)	3403.158	3411.150	3411.437	3423.146	0.633	0.060	0.671
TIHTG-LL(ρ, α, θ)	3738.406	3744.406	3744.573	3753.398	5.926	0.380	2.2×10^{-16}
TIHT-LL(ρ, β, θ)	3400.966	3406.966	3407.133	3415.958	0.655	0.052	0.825
E-LL(β, θ)	3530.302	3534.302	3534.385	3540.297	7.304	0.217	1.8×10^{-6}
TLEHLOBX-LL($\nu, \alpha, \beta, \theta$)	3435.261	3443.261	3443.541	3455.250	1.971	0.118	0.032
TLEHLOBX-LL(ν, α, θ)	3446.744	3452.744	3452.911	3461.736	2.562	0.129	0.015
TLHLOBX-LL(ν, β, θ)	3442.580	3448.580	3448.747	3457.572	2.297	0.125	0.020
HLOBX-LL(β, θ)	3445.099	3449.099	3449.182	3455.094	2.370	0.134	0.010
LL(θ)	3970.049	3972.049	3972.076	3975.046	2.335	0.715	2.2×10^{-16}
W(λ)	4017.695	4019.695	4019.722	4022.692	1.379	0.791	2.2×10^{-16}

The estimated parameters (standard error in parentheses) and the summary of the GoF statistics are shown in 4.1 and 4.2, respectively. According to the GoF statistics, the TIHT-LL(ρ, β, θ) distribution appears to be the most appropriate model for characterizing COVID-19 incidence in Southern Africa. Furthermore, based on the GoF statistics such as A^* , K-S, and its p-value, the TIHTEG-LL($\rho, \alpha, \beta, \theta$) model emerges as a strong contender and is recommended as another better model. To add on to that, visual illustrations were employed to draw more compelling conclusions about the best-chosen distributions in this study.

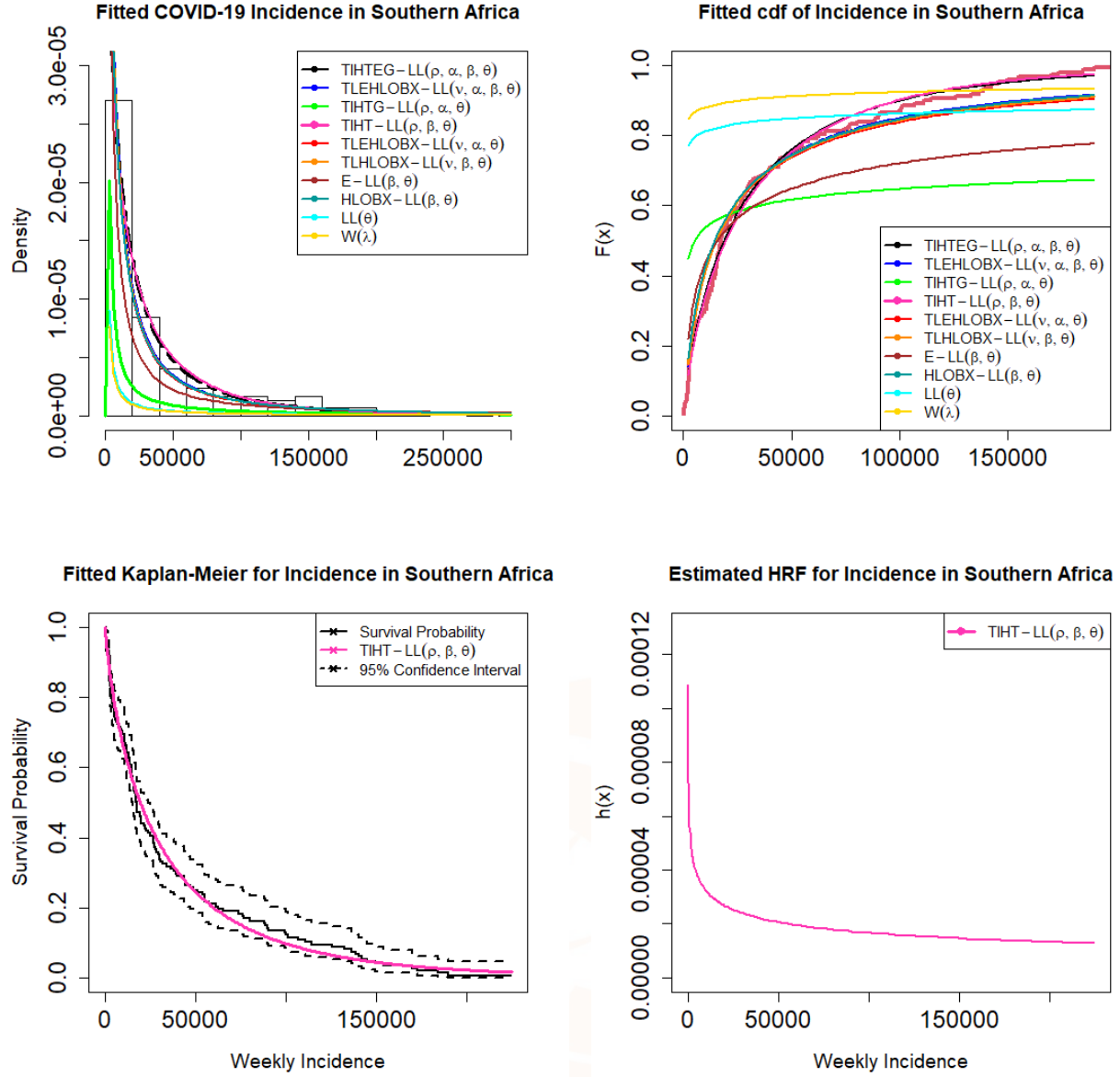


Figure 4.1: Fitted densities, Ecdf, Kaplan-Meier Survival Curve, and Hazard Rate Function Plot for Incidence in Southern Africa.

The fitted pdf and estimated cdf plots of the six distributions under consideration for weekly COVID-19 incidences in Southern Africa are shown in Figure 4.1. We observed that the models are performing well in describing and capturing the characteristics of the COVID-19 cases data. Additionally, the survival function of the TIHT-LL(ρ, β, θ) model, superimposed on the Kaplan-Meier plot, demonstrates that the model closely aligns with the empirical data. Furthermore, the estimated hazard rate function exhibits a decreasing shape, demonstrating the flexibility and ability of the TIHT-LL(ρ, β, θ) model to accurately describing the incidence data in Southern Africa.

4.2.1 Estimates and Model Selection Criteria for Incidence in Autumn

Table 4.3: Estimated parameters for the fitted distributions.

Distribution	Estimates and Standard Errors			
	ρ	α	β	θ
TIHTEG-LL($\rho, \alpha, \beta, \theta$)	204.400(6.6×10^{-5})	6.218(0.133)	1636.900(0.002)	0.023(0.003)
TIHTG-LL(ρ, α, θ)	0.225 (0.074)	0.139 (0.041)		4.157 (0.004)
TIHT-LL(ρ, β, θ)	3.7×10^4 (0.002)		45.067(2.384)	0.054(0.006)
E-LL(β, θ)			43.188(15.239)	0.467(0.047)
	ν	α	β	θ
TLEHLOBX-LL($\nu, \alpha, \beta, \theta$)	0.328(0.074)	12.116(0.002)	63.059(0.004)	$0.026(1.5 \times 10^{-3})$
TLEHLOBX-LL(ν, α, θ)	0.647(0.086)	$495.880(5.8 \times 10^{-5})$		$0.024(1.1 \times 10^{-3})$
TLHLOBX-LL(ν, β, θ)	1.243(0.286)		$281.090(1.1 \times 10^{-3})$	$0.025(1.1 \times 10^{-3})$
HLOBX-LL(β, θ)			$355.690(9.8 \times 10^{-9})$	$0.025(7.6 \times 10^{-4})$
LL(θ)				$0.171(0.022)$
W(λ)				$0.088(0.0091)$

Table 4.4: GoF statistics for COVID-19 incidence in autumn.

Distribution	Model Selection Criteria						
	$-2 \log \mathcal{L}$	AIC	CAIC	BIC	A*	K-S	P-Value
TIHTEG-LL($\rho, \alpha, \beta, \theta$)	812.698	820.698	821.910	827.248	1.577	0.227	0.033
TIHTG-LL(ρ, α, θ)	901.024	907.0239	907.730	911.937	3.483	0.340	0.0002
TIHT-LL(ρ, β, θ)	810.475	816.475	817.181	821.387	1.365	0.206	0.069
E-LL(β, θ)	846.476	850.476	850.819	853.751	3.851	0.265	0.008
TLEHLOBX-LL($\nu, \alpha, \beta, \theta$)	825.259	833.259	834.472	839.810	2.605	0.270	0.006
TLEHLOBX-LL(ν, α, θ)	828.382	834.382	835.088	839.295	2.816	0.277	0.005
TLHLOBX-LL(ν, β, θ)	826.937	832.937	833.643	837.850	2.711	0.276	0.005
HLOBX-LL(β, θ)	827.023	831.023	831.366	834.298	2.711	0.280	0.004
LL(θ)	960.901	962.901	963.012	964.539	2.820	0.683	2.2×10^{-16}
W(λ)	973.169	975.169	975.280	976.807	2.446	0.768	2.2×10^{-16}

The estimated parameters (standard error in parentheses) shown in Table 4.3 provide sufficient evidence for the validity of the distributions considered in this study. Table 4.4 displays the GoF measures. These measures indicate that the three-parameter TIHT-LL(ρ, β, θ) distribution yields the best fit for modelling the COVID-19 incidence in autumn. The conclusion is based on this distribution having the lowest values for $-2 \log \mathcal{L}$, AIC, CAIC, BIC, and K-S statistics. Furthermore, the K-S p-value of 0.069 further supports the conclusion that this nested model outperforms the other models under consideration. Furthermore, Figure 4.2 displays the fitted densities, estimated cdf plots, K-M survival plots, and estimated hazard plots that were employed. It can be observed that the fitted densities and estimated cdf plots closely follow the empirical plots. Additionally, the hazard plots demonstrate a slight increase, followed by a decreasing hazard rate function for the reported incidence in autumn. These plots collectively provide evidence that the TIHT-LL(ρ, β, θ) model accurately captures the characteristics of the

data set.

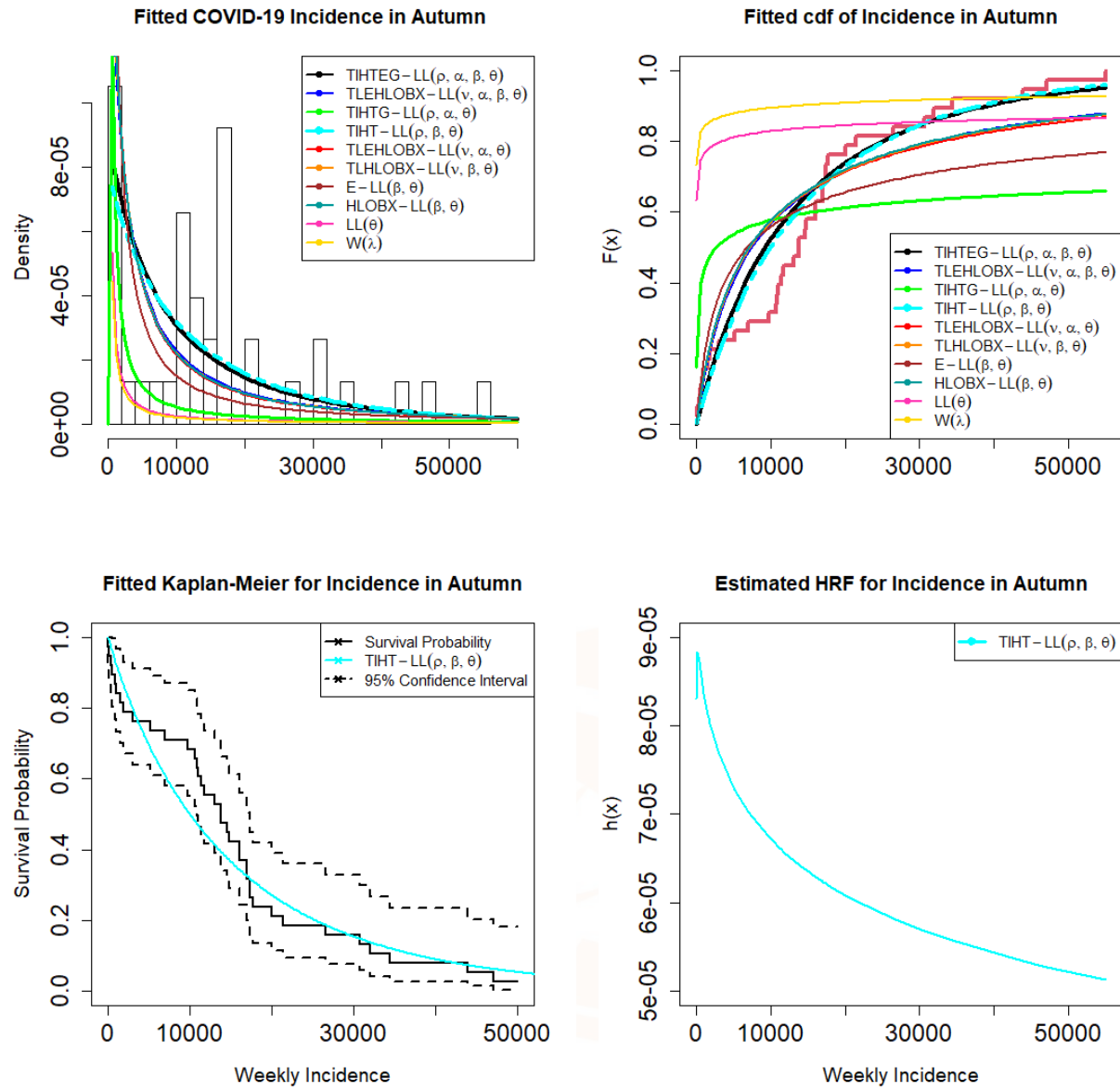


Figure 4.2: Fitted densities, Ecdf, Kaplan-Meier Survival Curve, and Hazard Rate Function Plot for Incidence in Autumn.

4.2.2 Estimates and Model Selection Criteria for Incidence in Winter

Table 4.5: Estimated parameters for the fitted distributions.

Distribution	Estimates and Standard Errors			
	ρ	α	β	θ
TIHTEG-LL($\rho, \alpha, \beta, \theta$)	140.412(0.040)	0.066(0.026)	13.543(3.029)	0.911(0.350)
TIHTG-LL(ρ, α, θ)	0.225(0.069)	0.259(0.068)		1.901 (0.009)
TIHT-LL(ρ, β, θ)	4922.000(0.002)		42.955(3.043)	0.066(0.008)
E-LL(β, θ)			1338.200(1.4×10^{-6})	0.745(0.017)
	ν	α	β	θ
TLEHLOBX-LL($\nu, \alpha, \beta, \theta$)	0.173(0.047)	31.030(0.002)	154.102(3.6×10^{-5})	0.027(0.001)
TLEHLOBX-LL(ν, α, θ)	31.162(4.1×10^{-8})	30.584(7.7×10^{-8})		0.021(0.003)
TLHLOBX-LL(ν, β, θ)	0.572(0.158)		2473.000(2.6×10^{-5})	0.027(0.001)
HLOBX-LL(β, θ)			1208.500(1.6×10^{-9})	0.026 (5.6×10^{-4})
				θ
LL(θ)				0.149(0.018)
				λ
W(λ)				0.077(0.008)

Table 4.6: GoF statistics for COVID-19 incidence in winter.

Distribution	Model Selection Criteria						
	$-2 \log \mathcal{L}$	AIC	CAIC	BIC	A*	K-S	P-Value
TIHTEG-LL($\rho, \alpha, \beta, \theta$)	937.410	945.410	946.586	952.064	0.708	0.118	0.607
TIHTG-LL(ρ, α, θ)	1045.042	1051.042	1051.728	1056.033	1.158	0.412	1.6×10^{-6}
TIHT-LL(ρ, β, θ)	936.480	942.480	943.165	947.470	0.618	0.110	0.693
E-LL(β, θ)	948.866	952.866	953.199	956.193	1.616	0.145	0.354
TLEHLOBX-LL($\nu, \alpha, \beta, \theta$)	940.749	948.749	949.925	955.403	0.994	0.131	0.478
TLEHLOBX-LL(ν, α, θ)	944.945	950.945	951.631	955.936	1.307	0.130	0.481
TLHLOBX-LL(ν, β, θ)	940.303	946.303	946.989	951.294	0.995	0.134	0.443
HLOBX-LL(β, θ)	940.649	944.649	944.982	947.976	0.993	0.130	0.488
LL(θ)	1107.172	1109.172	1109.280	1110.835	1.006	0.758	7.8×10^{-16}
W(λ)	1119.475	1121.475	1121.584	1123.139	0.906	0.836	7.8×10^{-16}

Table 4.5 presents parameter estimates and standard errors in parentheses for the distributions, while Table 4.6 displays the GoF statistics for the probability density functions. The GoF statistics in Table 4.6 indicate that the models under consideration are close competitors, except for the TIHTG-LL(ρ, α, θ), LL(θ) and W(λ) distributions, as the models have the largest GoF statistic and the least K-S p-value in comparison to the other models. However, the TIHT-LL(ρ, β, θ) model is the most suitable choice for describing the COVID-19 incidence reported in winter season, as the model exhibits least values for the model selection criteria ($-2 \log \mathcal{L}$, AIC, BIC, and A*) and the highest K-S p-value among the models. In Figure 4.3, the fitted pdfs and estimated cdf plots closely fit the empirical data. The fitted parametric survival function of the TIHT-LL(ρ, β, θ) model aligns closely with the Kaplan-Meier survival plot within the confidence interval. The estimated hazard rate function plots for the winter COVID-19 incidence demonstrate a decreasing shape. We can conclude that the TIHT-LL(ρ, β, θ) model is the most

appropriate and reliable choice for describing the COVID-19 weekly incidence in winter.

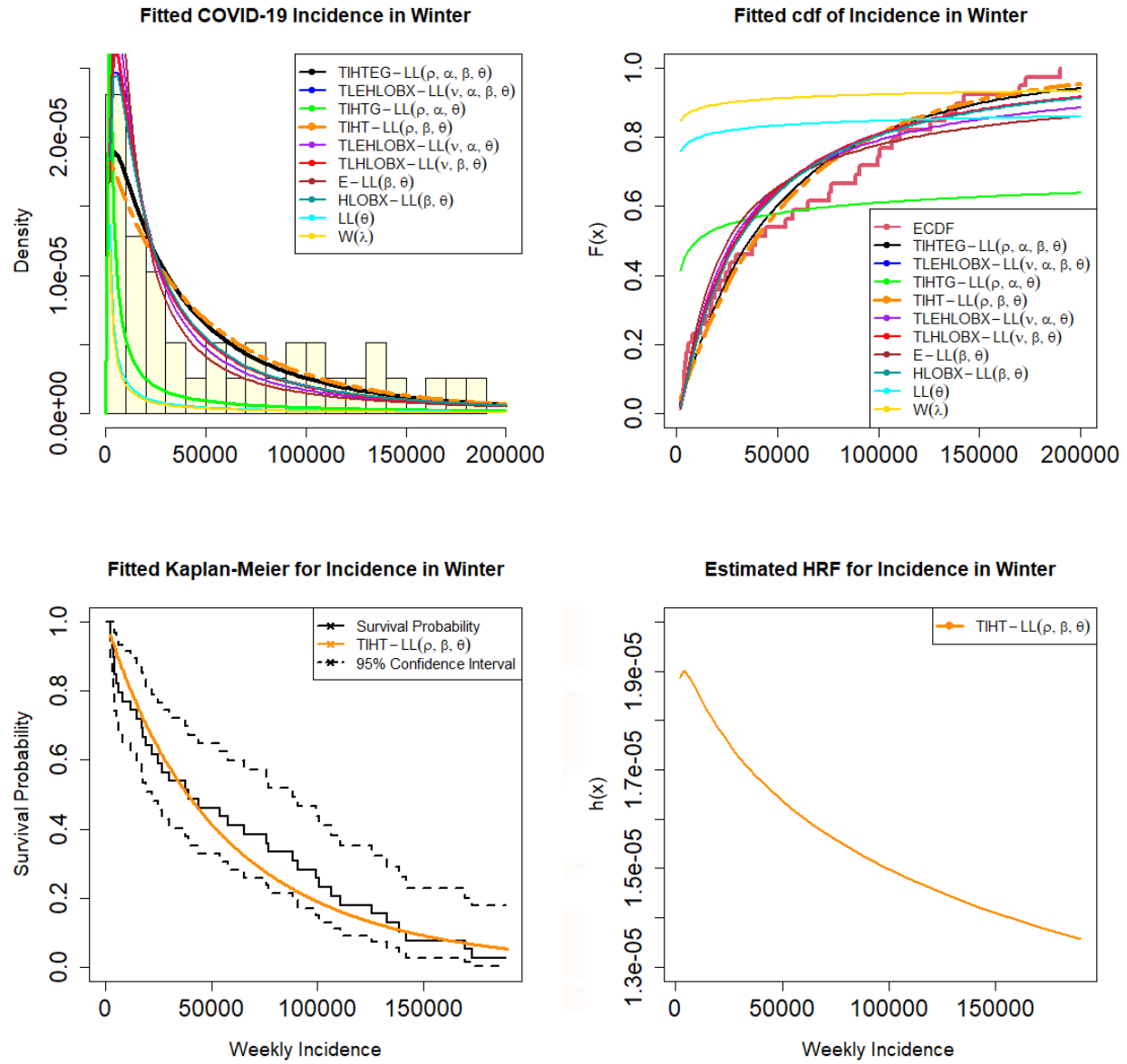


Figure 4.3: Fitted densities, Ecdf, Kaplan-Meier Survival Curve, and Hazard Rate Function Plot for Incidence in Winter.

4.2.3 Estimates and Model Selection Criteria for Incidence in Spring

Table 4.7: Estimated parameters for the fitted distributions.

Distribution	Estimates and Standard Errors			
	ρ	α	β	θ
TIHTEG-LL($\rho, \alpha, \beta, \theta$)	5.776 (1.727)	0.168 (0.091)	61.047 (0.123)	1.828(0.985)
TIHTG-LL(ρ, α, θ)	0.226(0.068)	0.194(0.051)		2.863 (0.003)
TIHT-LL(ρ, β, θ)	7.857(3.090)		59.579(0.576)	0.280 (0.022)
E-LL(β, θ)			1410.300(1.6×10^{-7})	1.112(0.019)
	ν	α	β	θ
TLEHLOBX-LL($\nu, \alpha, \beta, \theta$)	10.958(0.974)	9.988(1.488)	11.523 (1.254)	0.027 (0.002)
TLEHLOBX-LL(ν, α, θ)	122.560(1.1×10^{-8})	24.076(1.0×10^{-7})		0.026(4.7×10^{-4})
TLHLOBX-LL(ν, β, θ)	16.032(8.9×10^{-8})		107.039(2.3×10^{-8})	0.028(5.0×10^{-4})
HLOBX-LL(β, θ)			2315.300 (9.032×10^{-10})	0.032 (5.6×10^{-4})
	θ			
LL(θ)	0.169 (0.021)			
	λ			
W(λ)	0.088 (0.009)			

Table 4.8: GoF statistics for COVID-19 incidence in spring.

Distribution	Model Selection Criteria						
	$-2 \log \mathcal{L}$	AIC	CAIC	BIC	A^*	K-S	P-Value
TIHTEG-LL($\rho, \alpha, \beta, \theta$)	820.452	828.455	829.629	835.109	0.976	0.137	0.416
TIHTG-LL(ρ, α, θ)	936.042	942.042	942.728	947.033	1.000	0.452	8.8×10^{-8}
TIHT-LL(ρ, β, θ)	820.466	826.466	827.152	831.457	0.912	0.128	0.504
E-LL(β, θ)	821.833	825.833	826.166	829.160	1.298	0.181	0.138
TLEHLOBX-LL($\nu, \alpha, \beta, \theta$)	820.447	828.447	829.623	835.101	1.104	0.164	0.217
TLEHLOBX-LL(ν, α, θ)	820.578	826.578	827.263	831.568	1.147	0.170	0.185
TLHLOBX-LL(ν, β, θ)	820.449	826.449	27.135	831.440	1.090	0.162	0.232
HLOBX-LL(β, θ)	819.203	823.203	823.536	826.530	0.925	0.141	0.382
LL(θ)	998.216	1000.216	1000.324	1001.88	0.935	0.776	7.8×10^{-16}
W(λ)	1010.421	1012.421	1012.53	1014.085	0.890	0.852	7.8×10^{-16}

Table 4.7 displays the parameter estimates (standard error in parentheses), and Table 4.8 presents some model selection criteria. The results show that the TIHT-LL(ρ, β, θ) model consistently has the lowest values for various GoF statistics, as shown in Table 4.8. These results indicate that the TIHT-LL(ρ, β, θ) model provides the best fit for describing the COVID-19 situation during the spring season. Furthermore, the model outperforms the others based on the high K-S p-value. Thus, we can infer that the model better fits the dataset under consideration.

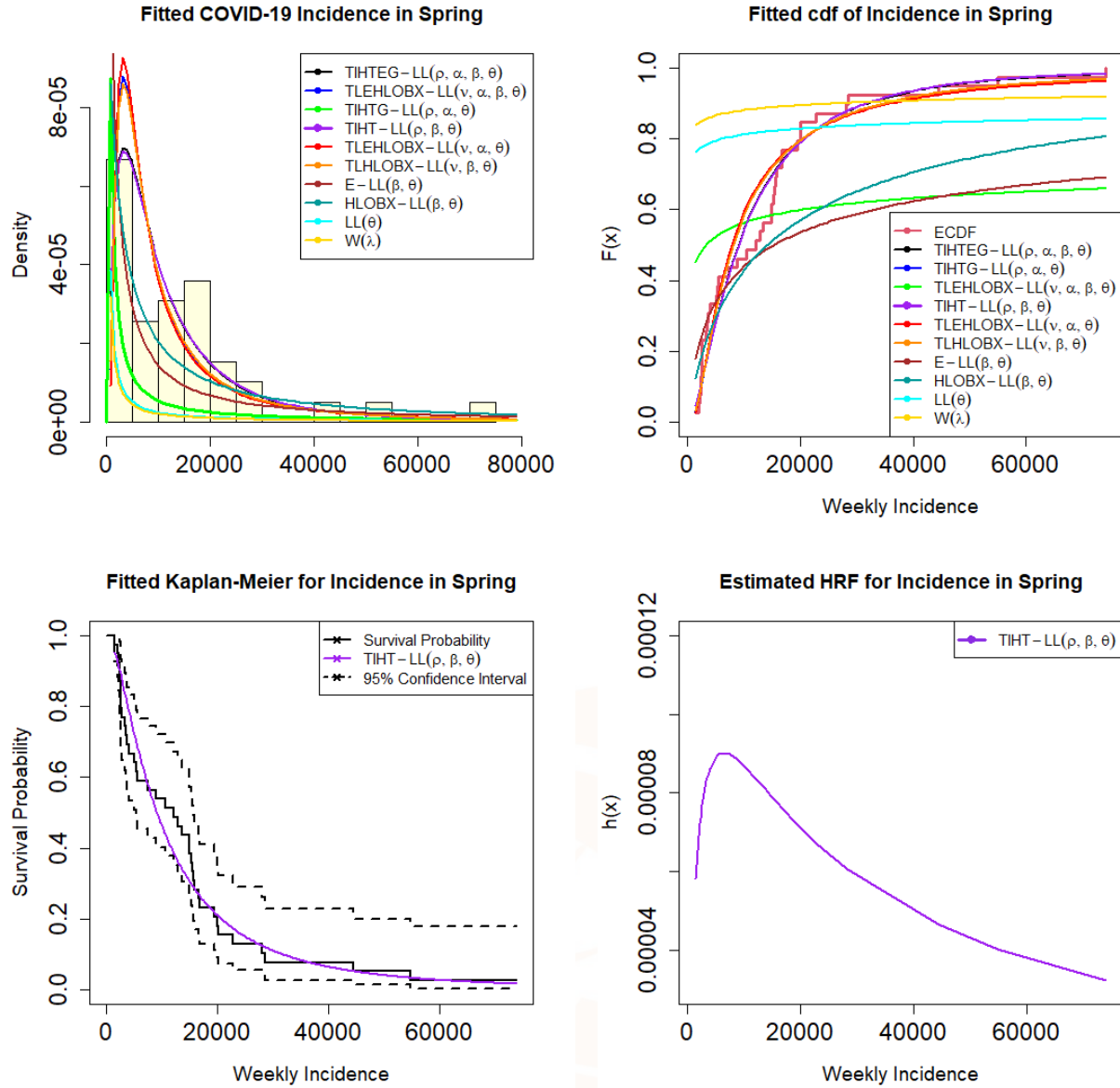


Figure 4.4: Fitted densities, Ecdf, Kaplan-Meier Survival Curve, and Hazard Rate Function Plot for Incidence in Spring.

Figure 4.4 shows the fitted densities, Kaplan-Meier survival plots, and the hazard rate function plots for the incidence in spring. The fitted Kaplan-Meier plots show that all of the incidence data points fall within the confidence region, implying that this model adequately describes the characteristics of weekly reported incidence in the spring. The hazard rate function plot indicates that the incidence hazard increased to a peak and then decreased over time.

4.2.4 Estimates and Model Selection Criteria for Incidence in Summer

Table 4.9: Estimated parameters for the fitted distributions.

Distribution	Estimates and Standard Errors			
	ρ	α	β	θ
TIHTEG-LL($\alpha, \alpha, \beta, \theta$)	309.820(3.1×10^{-5})	7.683(0.172)	5789.600(4.0×10^{-5})	0.020(0.003)
TIHTG-LL(ρ, α, θ)	0.225(0.077)	0.249(0.074)		1.930(0.010)
TIHT-LL(ρ, β, θ)	1.29×10^4 (0.001)		47.908(3.904)	0.065(0.009)
E-LL(β, θ)			645.670(3.2×10^{-6})	0.656(0.019)
	ν	α	β	θ
TLEHLOBX-LL($\nu, \alpha, \beta, \theta$)	0.258(0.072)	4.562(0.004)	1084.450(1.0×10^{-5})	0.027(0.004)
TLEHLOBX-LL(ν, α, θ)	0.300(0.094)	$1847.800(1.2 \times 10^{-5})$		0.025(0.001)
TLHLOBX-LL(ν, β, θ)	0.980(0.294)		1144.600(0.002)	0.025(0.001)
HLOBX-LL(β, θ)			$112.930(1.8 \times 10^{-9})$	$0.025(5.9 \times 10^{-4})$
LL(θ)				θ 0.145(0.020)
W(λ)				λ 0.075(0.009)

Table 4.10: GoF statistics for COVID-19 incidence in summer.

Distribution	Model Selection Criteria						
	$-2 \log \mathcal{L}$	AIC	$CAIC$	BIC	A^*	K-S	P-Value
TIHTEG-LL($\rho, \alpha, \beta, \theta$)	755.596	763.596	765.134	769.332	0.806	0.129	0.63
TIHTG-LL(ρ, α, θ)	846.625	852.625	853.514	856.927	2.102	0.399	5.9×10^{-5}
TIHT-LL(ρ, β, θ)	754.819	760.819	761.708	765.121	0.706	0.126	0.659
E-LL(β, θ)	774.366	952.866	953.199	956.193	1.616	0.145	0.354
TLEHLOBX-LL($\nu, \alpha, \beta, \theta$)	761.822	769.822	771.361	775.558	0.232	0.206	0.138
TLEHLOBX-LL(ν, α, θ)	764.519	770.519	771.408	774.821	1.882	0.219	0.118
TLHLOBX-LL(ν, β, θ)	763.492	778.366	778.795	781.234	2.885	0.261	0.024
HLOBX-LL(β, θ)	763.492	767.492	767.921	770.360	1.780	0.214	0.101
LL(θ)	895.969	897.969	898.107	899.403	1.744	0.752	3.3×10^{-16}
W(λ)	905.720	907.720	907.858	909.154	1.511	0.831	3.3×10^{-16}

The parameter estimates (standard error in parenthesis) are shown in Table 4.9, and the results of some model selection criteria are shown in Table 4.10. We note that the TIHT-LL(ρ, β, θ) model has the lowest GoF statistic values and the highest K-S p-value. Consequently, we can conclude that the TIHT-LL(ρ, β, θ) model is the best model for describing the COVID-19 incidence reported during the summer season. Furthermore, Figure 4.5 displays the fitted densities, estimated cdf plots, K-M survival plots, and estimated hazard plots that were employed. It can be observed that the fitted densities and estimated cdf plots closely follow the empirical plots. Additionally, the hazard plot demonstrate a decreasing hazard rate function for the reported incidence in summer. Thus, it is evident from these plots that the three-parameter distribution is the most suitable for describing the weekly incidence reported during the summer period.

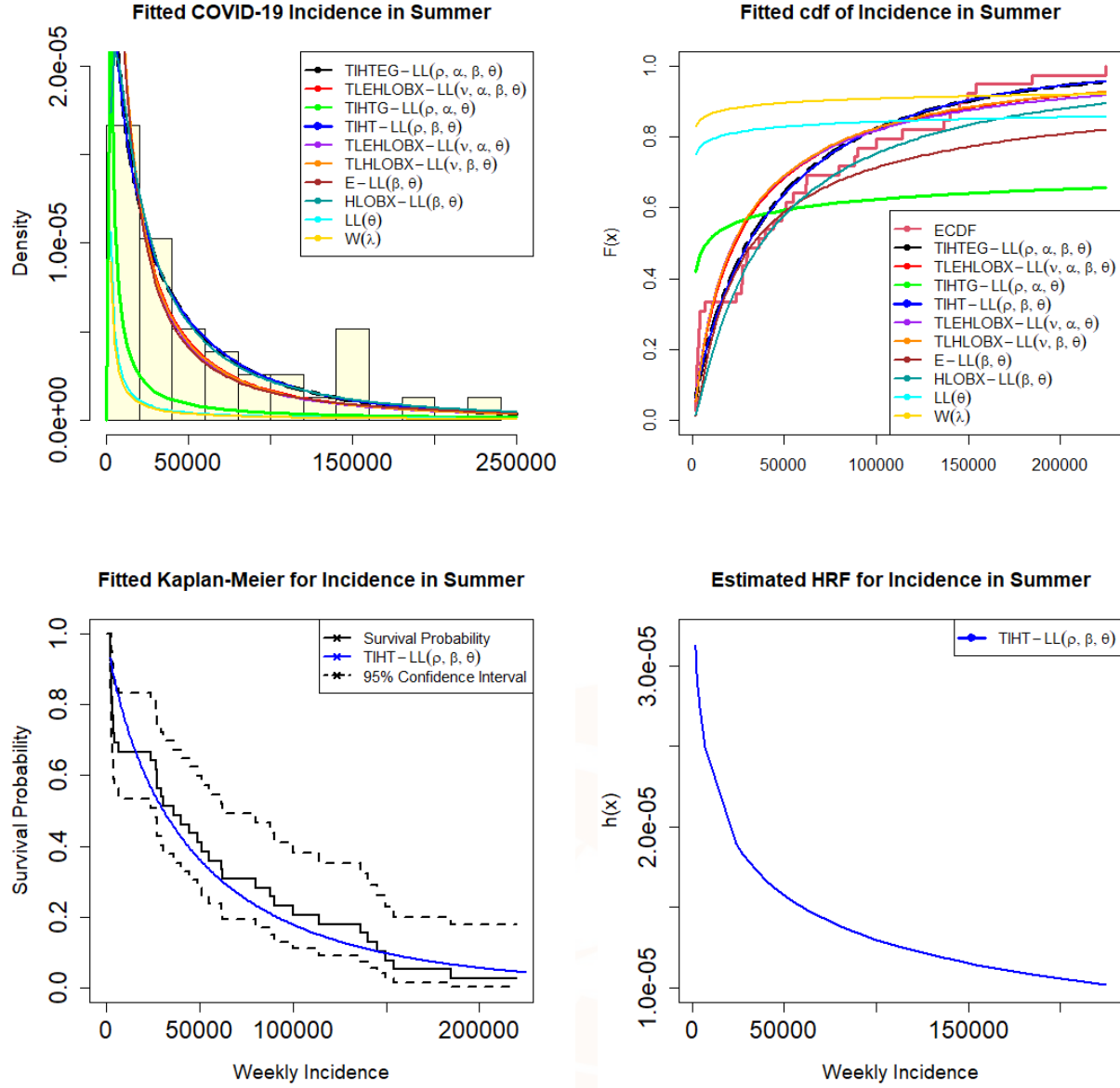


Figure 4.5: Fitted densities, Ecdf, Kaplan-Meier Survival Curve, and Hazard Rate Function Plot for Incidence in Summer.

4.3 Results for Mortality in Southern African Region

This section presents the results obtained from the COVID-19 mortality data for each probability distribution and is presented numerically and graphically for the overall Southern African region and at the seasonal level.

Table 4.11: Estimated parameters for the fitted distributions.

Distribution	Estimates and Standard Errors			
	ρ	α	β	θ
TIHTEG-LL($\rho, \alpha, \beta, \theta$)	299.830(1.9×10^{-5})	7.704(0.036)	3430.500(1.5×10^{-5})	0.015(0.001)
TIHTG-LL(ρ, α, θ)	0.226(0.042)	0.386(0.430)		2.393(2.591)
TIHT-LL(ρ, β, θ)	155.000(0.001)		28.226(0.554)	0.058(0.004)
E-LL(β, θ)			12.921(1.648)	0.528(0.028)
	ν	α	β	θ
TLEHLOBX-LL($\nu, \alpha, \beta, \theta$)	0.260(0.016)	8.658(5.136)	41.594(0.525))	0.031(0.002)
TLEHLOBX-LL(ν, α, θ)	0.278(0.029)	247.300(3.2×10^{-5})		0.029(8.8×10^{-4})
TLHLOBX-LL(ν, β, θ)	11.201(0.600)		14.668(0.614)	0.024(9.0×10^{-4})
HLOBX-LL(β, θ)			129.510(3.76×10^{-8})	0.029(6.8×10^{-4})
				θ
LL(θ)				0.265(0.017)
				λ
W(λ)				0.132(0.007)

Table 4.12: GoF statistics for COVID-19 mortality in Southern Africa.

Distribution	Model Selection Criteria						
	$-2 \log \mathcal{L}$	AIC	CAIC	BIC	A*	K-S	P-Value
TIHTEG-LL($\rho, \alpha, \beta, \theta$)	2259.988	2267.988	2268.267	2279.977	0.774	0.072	0.422
TIHTG-LL(ρ, α, θ)	2423.273	2429.273	2429.440	2438.265	4.744	0.241	7.0×10^{-8}
TIHT-LL(ρ, β, θ)	2259.367	2265.367	2265.534	2274.359	0.738	0.066	0.535
E-LL(β, θ)	2316.589	2320.589	2320.672	2326.583	4.051	0.139	0.006
TLEHLOBX-LL($\nu, \alpha, \beta, \theta$)	2270.062	2278.062	2278.342	2290.051	1.389	0.095	0.136
TLEHLOBX-LL(ν, α, θ)	2275.571	2281.571	2281.737	2290.562	1.728	0.103	0.088
TLHLOBX-LL(ν, β, θ)	2285.540	2291.539	2291.706	2300.531	2.386	0.115	0.039
HLOBX-LL(β, θ)	2273.400	2374.168	2374.246	2380.280	1.906	0.083	0.225
LL(θ)	2647.605	2743.932	2743.958	2746.988	2.542	0.597	2.2×10^{-16}
W(λ)	2700.623	2801.762	2801.788	2804.818	1.676	0.693	2.2×10^{-16}

Table 4.11 displays the parameters and standard errors in parentheses, and Table 4.12 displays the results of some GoF statistics used in this study. Note that the TIHT-LL(ρ, β, θ) distribution is the most suitable model for describing the confirmed weekly COVID-19 mortality in Southern Africa. This is due to the model having the lowest GoF statistics as well as the highest K-S p-value. In Figure 4.6, the fitted pdfs and estimated cdf plots closely fit the empirical data. Furthermore, the fitted parametric survival function of the TIHT-LL(ρ, β, θ) model aligns closely with the Kaplan-Meier survival plot within the confidence interval. Lastly, the estimated hazard rate function plot depict the decreasing hazard rate of the weekly mortality in Southern Africa.

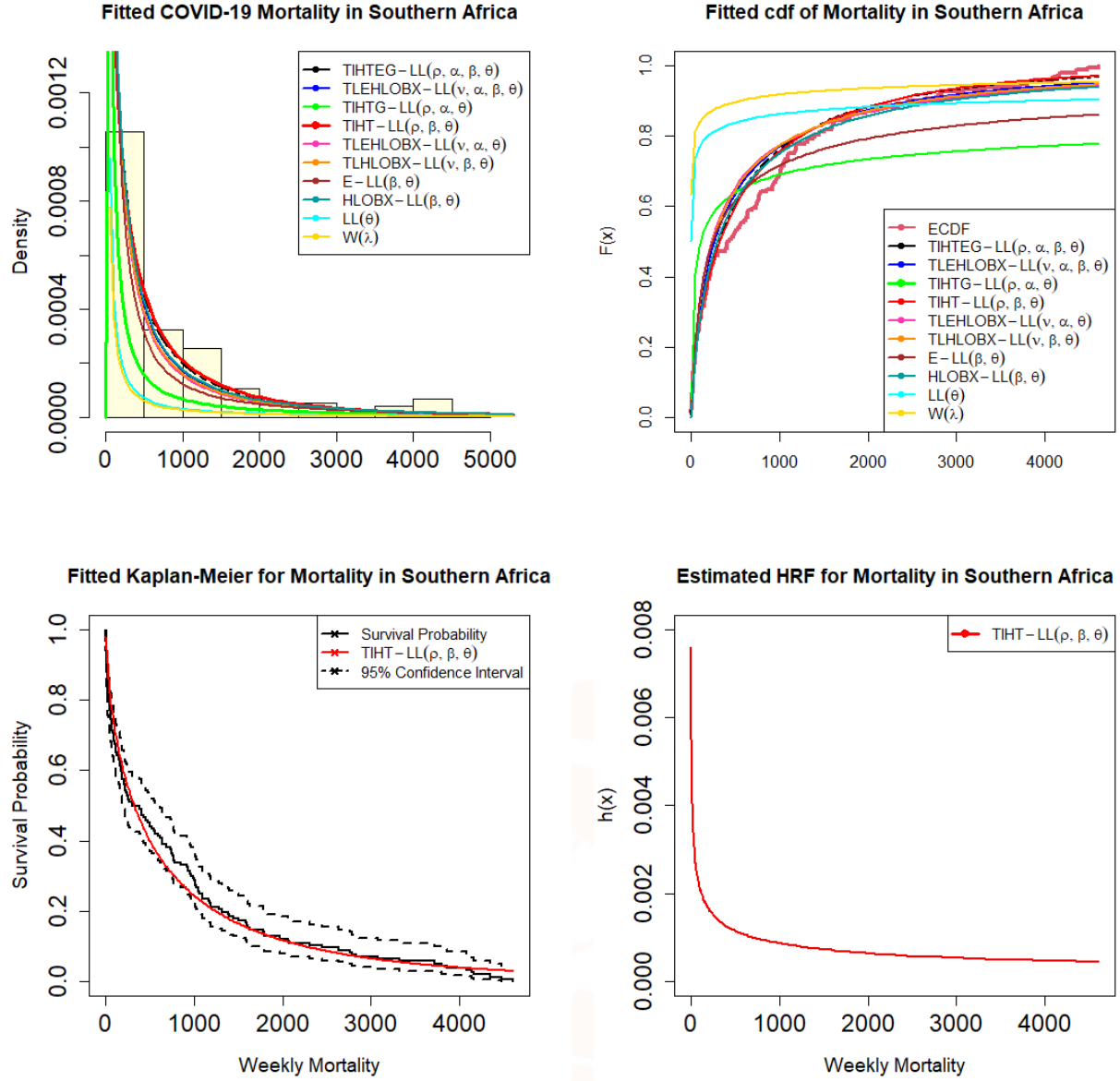


Figure 4.6: Fitted densities, Ecdf, Kaplan-Meier Survival Curve, and Hazard Rate Function Plot for Mortality in Southern Africa.

Based on the comprehensive analysis presented by the numerical and graphical measures in Table 4.14 and Figure 4.6, respectively, we can infer that the $\text{TIHT-LL}(\rho, \beta, \theta)$ provides a good fit to the COVID-19 mortality in Southern Africa.

4.3.1 Estimates and Model Selection Criteria for Mortality in Autumn

Table 4.13: Estimated parameters for the fitted distributions.

Distribution	Estimates and Standard Errors			
	ρ	α	β	θ
TIHTEG-LL($\rho, \alpha, \beta, \theta$)	0.614(0.293)	0.041(0.016)	4.546(1.429)	16.767(4.5×10^{-5})
TIHTG-LL(ρ, α, θ)	4122.231(0.001)		30.188(0.942)	0.056(0.007)
TIHT-LL(ρ, β, θ)	0.258(0.099)	0.051(0.019)		19.853(4.8×10^{-6})
E-LL(β, θ)			8.501(1.807)	0.553(0.057)
	ν	α	β	θ
TLEHLOBX-LL($\nu, \alpha, \beta, \theta$)	0.174(0.142)	4.923(3.973)	100.777(0.095)	0.037(0.003)
TLEHLOBX-LL(ν, α, θ)	0.202(0.037)	253.456(2.6×10^{-5})		0.034(0.001)
TLHLOBX-LL(ν, β, θ)	0.695(0.127)		140.201(0.006)	0.033(0.002)
HLOBX-LL(β, θ)			96.807(1.4×10^{-7})	0.032 (0.004)
LL(θ)				θ 0.319(0.039)
W(λ)				λ 0.158(0.016)

Table 4.14: GoF statistics for COVID-19 mortality in autumn.

Distribution	Model Selection Criteria						
	$-2 \log \mathcal{L}$	AIC	CAIC	BIC	A*	K-S	P-Value
TIHTEG-LL($\rho, \alpha, \beta, \theta$)	561.112	569.112	570.223	575.966	2.856	0.210	0.054
TIHTG-LL(ρ, α, θ)	573.754	579.754	580.403	584.895	2.713	0.254	0.010
TIHT-LL(ρ, β, θ)	537.645	543.645	544.293	548.785	0.545	0.103	0.774
E-LL(β, θ)	556.057	560.057	560.373	563.484	2.221	0.166	0.208
TLEHLOBX-LL($\nu, \alpha, \beta, \theta$)	541.422	549.422	550.533	556.276	1.001	0.145	0.359
TLEHLOBX-LL(ν, α, θ)	544.346	550.346	550.994	555.486	1.239	0.152	0.296
TLHLOBX-LL(ν, β, θ)	543.354	549.354	550.003	554.495	1.163	0.149	0.321
HLOBX-LL(β, θ)	543.539	547.539	547.855	550.966	1.160	0.144	0.364
LL(θ)	634.001	636.001	636.103	637.714	1.511	0.562	1.1×10^{-11}
W(λ)	649.214	651.214	651.316	652.927	0.990	0.672	2.2×10^{-16}

The estimated parameter results are presented in Table 4.13, while Table 4.14 demonstrates that the TIHT-LL(ρ, β, θ) model showcases the most favorable outcomes for the $-2 \log \mathcal{L}$, AIC, BIC, A*, and K-S measures. Additionally, it exhibits the highest K-S p-value statistics. These findings suggest that the model provides the most accurate representation of the autumn COVID-19 mortality data.

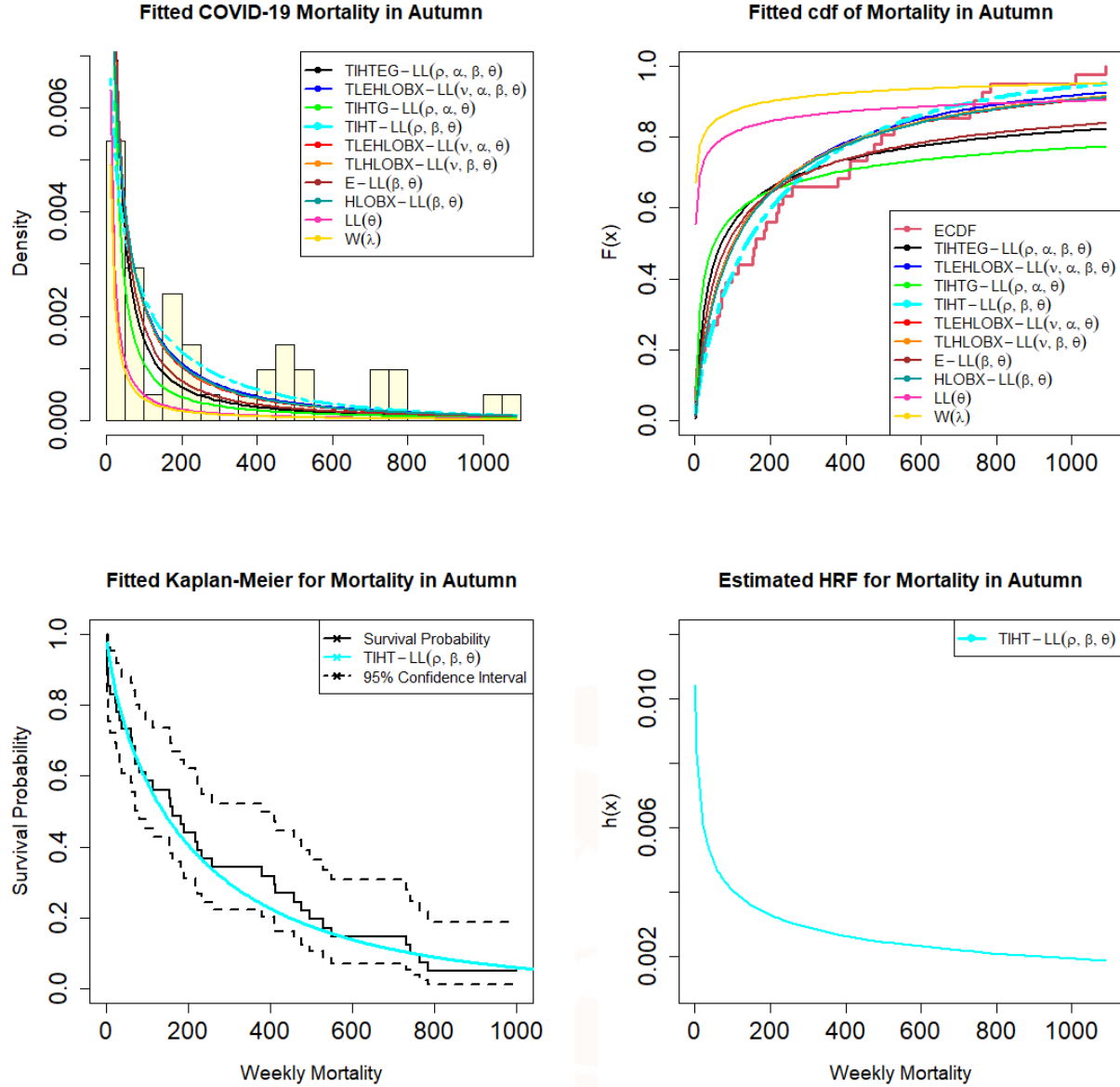


Figure 4.7: Fitted densities, Ecdf, Kaplan-Meier Survival Curve, and Hazard Rate Function Plot for Mortality in Autumn.

Presented in Figure 4.7 are the fitted densities, estimated cdf, Kaplan-Meier survival function, and estimated hazard rate function plots for the weekly mortality in autumn. It can be seen from fitted densities and estimated cdf plots that all the considered models are capturing the characteristics of the data with the TIHT-LL(ρ, β, θ) distribution being the closest to the empirical data. Hence, it can be used to make inferences about the mortality reported in autumn period. The fitted survival function on the Kaplan-Meier survival plot on the best-chosen model closely follows the empirical data. The flexibility of the best-chosen model is also shown in its ability to capture the decreasing hazard rate, which indicates the rate at which the number of individuals dying from the disease was decreasing.

4.3.2 Estimates and Model Selection Criteria for Mortality in Winter

Table 4.15: Estimated parameters for the fitted distributions.

Distribution	Estimates and Standard Errors			
	ρ	α	β	θ
TIHTEG-LL($\rho, \alpha, \beta, \theta$)	423.120(2.8×10^{-5})	8.313(0.084)	5993.232(2.0×10^{-5})	0.015(0.002)
TIHTG-LL(ρ, α, θ)	0.213(0.076)	0.158(0.051)		5.431(0.012)
TIHT-LL(ρ, β, θ)	7305.000(0.001)		34.906(1.453)	0.057(0.007)
E-LL(β, θ)			24.471(7.874)	0.572(0.062)
	ν	α	β	θ
TLEHLOBX-LL($\nu, \alpha, \beta, \theta$)	0.189(0.156)	3.689(3.129)	290.423(0.016)	0.033(0.002)
TLEHLOBX-LL(ν, α, θ)	0.201(0.043)	512.441(1.6×10^{-5})		0.030(0.021)
TLHLOBX-LL(ν, β, θ)	0.675(0.142)		298.205(0.001)	0.031(0.002)
HLOBX-LL(β, θ)			194.090(3.8×10^{-8})	0.030 (0.001)
				θ
LL(θ)				0.237(0.029)
				λ
W(λ)				0.120(0.012)

Table 4.16: GoF statistics for COVID-19 mortality in winter.

Distribution	Model Selection Criteria						
	$-2\log \mathcal{L}$	AIC	$CAIC$	BIC	A^*	K-S	P-Value
TIHTEG-LL($\rho, \alpha, \beta, \theta$)	642.566	650.566	651.742	657.220	1.072	0.172	0.178
TIHTG-LL(ρ, α, θ)	701.928	707.928	708.614	712.919	2.278	0.304	0.001
TIHT-LL(ρ, β, θ)	641.885	647.885	648.571	652.876	1.004	0.158	0.256
E-LL(β, θ)	660.813	667.052	667.385	670.379	2.368	0.259	0.009
TLEHLOBX-LL($\nu, \alpha, \beta, \theta$)	646.694	654.694	655.871	661.348	1.477	0.227	0.030
TLEHLOBX-LL(ν, α, θ)	649.516	655.516	656.201	660.506	1.680	0.233	0.024
TLHLOBX-LL(ν, β, θ)	648.472	654.472	655.157	659.462	1.605	0.230	0.027
HLOBX-LL(β, θ)	648.684	652.684	653.018	656.011	1.614	0.223	0.035
LL(θ)	763.134	765.134	765.242	766.798	1.795	0.651	7.8×10^{-16}
W(λ)	776.463	778.463	778.571	780.127	1.509	0.747	7.8×10^{-16}

Table 4.15 displays parameter estimates as well as the standard errors, which are presented in parentheses, and Table 4.16 displays the results of model selection criteria. The model fit evaluation statistics presented in Table 4.16 show that the TIHT-LL(ρ, β, θ) distribution exhibits the lowest GoF statistics and the highest Kolmogorov-Smirnov p-value. Therefore, we can conclude that the TIHT-LL(ρ, β, θ) distribution is the most appropriate model that can describe COVID-19 mortality in winter. Furthermore, the fitted pdfs and estimated cdfs in Figure 4.8 display that, with the exception of the TIHTG-LL(ν, α, θ), LL(θ) and W(λ) distributions which deviates from the empirical data, all the fitted models are closely aligned with the observed data. The survival function plots further confirm that the model fits the data as it falls within the Kaplan-Meier confidence interval. Thus, it is clear from this plot that the three-parameter distribution is best suited for describing the weekly confirmed mortality during winter season.

The hazard rate function plot displays a decreasing shape.

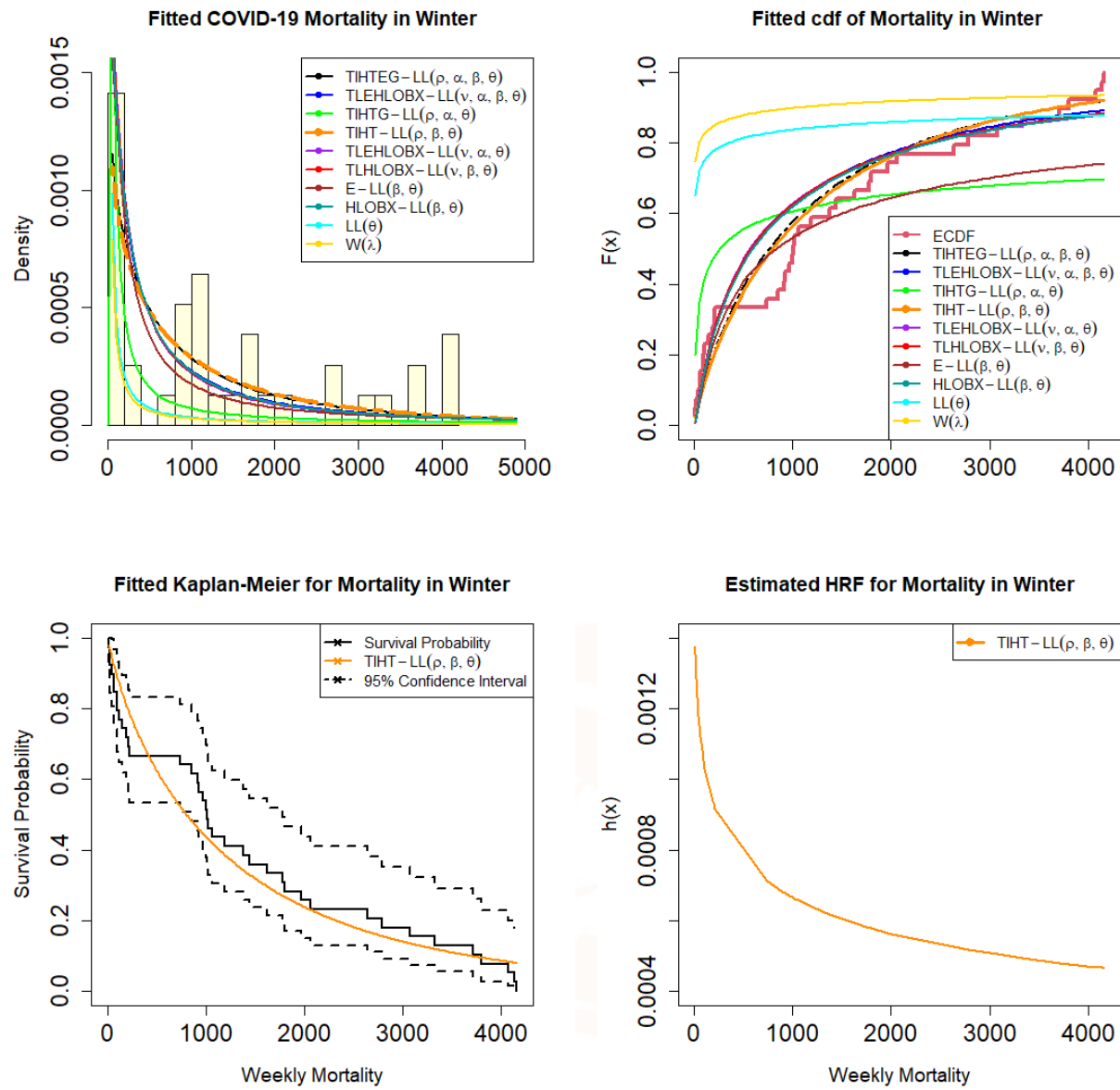


Figure 4.8: Fitted densities, Ecdf, Kaplan-Meier Survival Curve, and Hazard Rate Function Plot for Mortality in Winter.

4.3.3 Estimates and Model Selection Criteria for Mortality in Spring

Table 4.17: Estimated parameters for the fitted distributions.

Distribution	Estimates and Standard Errors			
	ρ	α	β	θ
TIHTEG-LL($\rho, \alpha, \beta, \theta$)	5.289(4.268)	1.953(1.134)	33.43(30.263)	0.152 (0.086)
TIHTG-LL(ρ, α, θ)	0.204(0.074)	0.191(0.062)		5.846(0.002)
TIHT-LL(ρ, β, θ)	52.531(0.059)		19.110(1.235)	0.115(0.012)
E-LL(β, θ)			23.520(7.935)	0.713(0.083)
	ν	α	β	θ
TLEHLOBX-LL($\nu, \alpha, \beta, \theta$)	0.151(0.031)	23.472(0.003)	28.27(0.001)	0.036(0.002)
TLEHLOBX-LL(ν, α, θ)	0.147(0.031)	570.3(7.6×10^{-6})		0.035(2.1×10^{-3})
TLHLOBX-LL(ν, β, θ)	0.356(0.074)		482.7(4.8×10^{-5})	0.037(2.0×10^{-3})
HLOBX-LL(β, θ)			154.700 (7.46×10^{-8})	0.034 (0.001)
LL(θ)				θ 0.294(0.037)
W(λ)				λ 0.147(0.015)

Table 4.18: GoF statistics for COVID-19 mortality in spring.

Distribution	Model Selection Criteria						
	$-2 \log \mathcal{L}$	AIC	CAIC	BIC	A*	K-S	P-Value
TIHTEG-LL($\rho, \alpha, \beta, \theta$)	545.801	553.801	554.978	560.455	0.599	0.104	0.756
TIHTG-LL(ρ, α, θ)	586.958	592.958	593.643	597.948	0.921	0.254	0.010
TIHT-LL(ρ, β, θ)	544.628	550.628	551.313	555.618	0.458	0.118	0.604
E-LL(β, θ)	549.390	553.390	553.724	556.717	1.024	0.121	0.580
TLEHLOBX-LL($\nu, \alpha, \beta, \theta$)	544.319	552.319	553.496	558.973	0.528	0.101	0.784
TLEHLOBX-LL(ν, α, θ)	544.166	550.166	550.852	555.157	0.554	0.099	0.800
TLHLOBX-LL(ν, β, θ)	543.517	549.517	550.203	554.508	0.520	0.102	0.777
HLOBX-LL(β, θ)	544.338	548.338	548.671	551.665	0.542	0.107	0.728
LL(θ)	648.529	650.529	650.637	652.192	0.648	0.670	7.8×10^{-16}
W(λ)	662.501	664.501	664.609	666.165	0.517	0.759	7.8×10^{-16}

Table 4.17 presents estimated parameters and standard errors in brackets for the distributions, while Table 4.18 displays the results for the model selection criteria. The GoF statistics in Table 4.18 indicate that the six models considered are close competitors, except for the TIHTG-LL(ρ, α, θ) model as it exhibits the highest GoF statistics and the lowest K-S p-value. However, the TLHLOBX-LL(ν, β, θ) model seem to be the best fit for describing the COVID-19 fatalities in spring since the model has the lowest values for the GoF criteria ($-2 \log \mathcal{L}$, AIC, BIC, and A*) and the highest K-S p-value compared to the other competing models. In Figure 4.9, the fitted pdfs and cdfs closely match the empirical data with the TLHLOBX-LL(ν, β, θ) distribution being the closest to the data. The fitted parametric survival function of the TLHLOBX-LL(ν, β, θ) model aligns closely with the Kaplan-Meier survival plot and falls within the confidence interval. The estimated hazard rate function plot for the spring COVID-19 mortality data demonstrates

a decreasing shape. By diligently examining the presented figures and statistical measures, we can conclude that the TLHLOBX-LL(ν, β, θ) model is the most appropriate and reliable choice for describing the COVID-19 mortality confirmed in spring season.

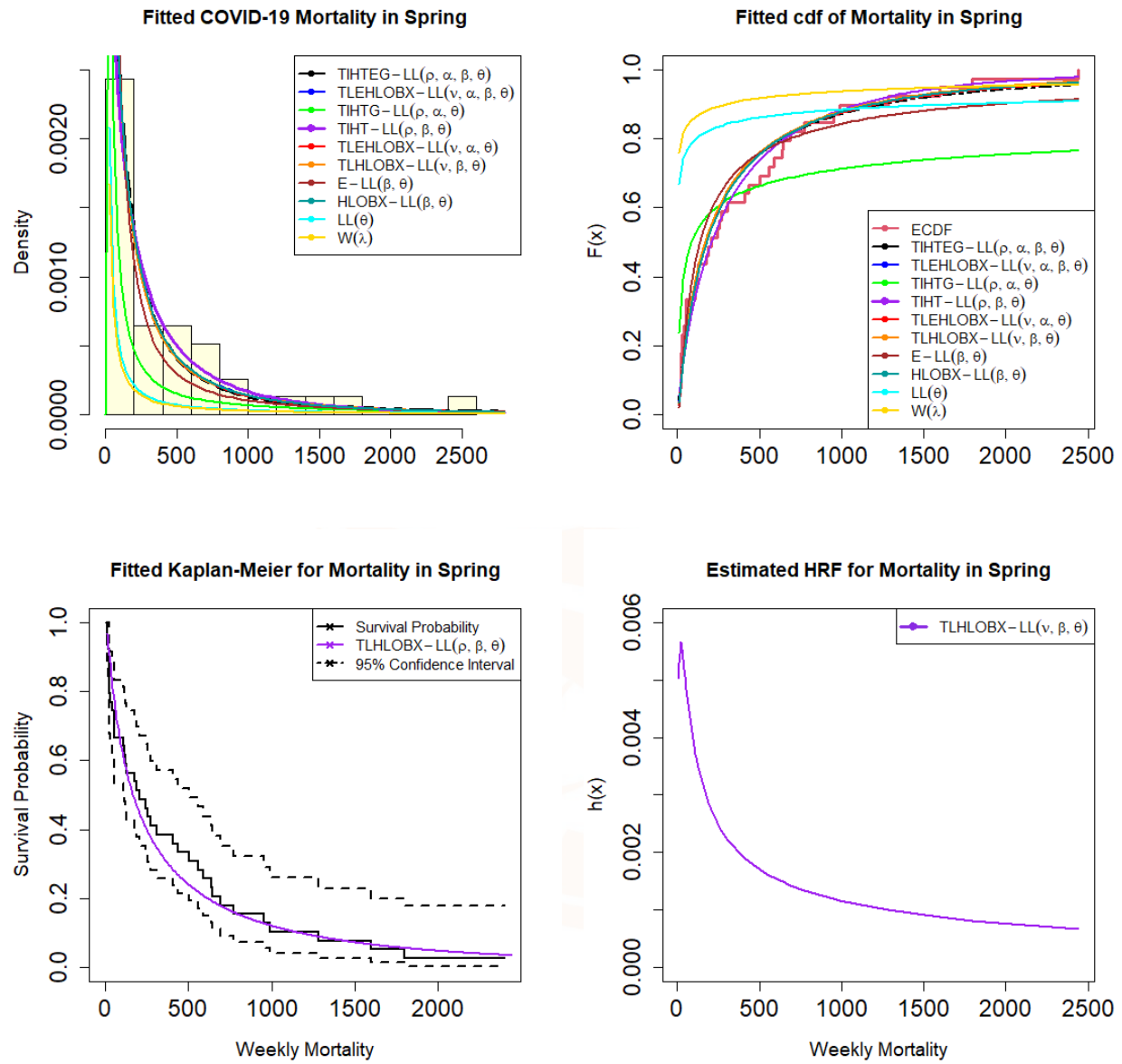


Figure 4.9: Fitted densities, Ecdf, Kaplan-Meier Survival Curve, and Hazard Rate Function Plot for Mortality in Spring.

4.3.4 Estimates and Model Selection Criteria for Mortality in Summer

Table 4.19: Estimated parameters for the fitted distributions.

Distribution	Estimates and Standard Errors			
	ρ	α	β	θ
TIHTEG-LL($\rho, \alpha, \beta, \theta$)	295.810(5.0×10^{-5})	7.291(0.105)	2931.500(5.1×10^{-5})	0.0181(0.003)
TIHTG-LL(ρ, α, θ)	0.222(0.090)	0.130(0.048)		6.253(0.001)
TIHT-LL(ρ, β, θ)	14931.000(0.003)		37.684(1.736)	0.057(0.008)
E-LL(β, θ)			19.256(6.136)	0.520(0.060)
	ν	α	β	θ
TLEHLOBX-LL($\nu, \alpha, \beta, \theta$)	0.234(0.241)	8.511(8.912)	75.448(0.716)	0.031(0.003)
TLEHLOBX-LL(ν, α, θ)	0.281(0.068)	349.190(5.1×10^{-5})		0.0284(0.002)
TLHLOBX-LL(ν, β, θ)	0.972(0.226)		194.120(0.001)	0.029(0.002)
HLOBX-LL(β, θ)			188.330(4.3×10^{-8})	0.029 (0.001)
LL(θ)				θ 0.233(0.032)
W(λ)				λ 0.118(0.013)

Table 4.20: Distribution fitting to COVID-19 mortality in summer.

Distribution	Model Selection Criteria						
	$-2 \log \mathcal{L}$	AIC	CAIC	BIC	A*	K-S	P-Value
TIHTEG-LL($\rho, \alpha, \beta, \theta$)	515.116	523.116	524.655	528.852	0.946	0.179	0.242
TIHTG-LL(ρ, α, θ)	566.299	572.299	573.188	576.601	2.741	0.306	0.005
TIHT-LL(ρ, β, θ)	513.997	519.999	520.888	524.302	0.827	0.166	0.427
E-LL(β, θ)	536.287	540.287	540.715	543.155	2.777	0.255	0.029
TLEHLOBX-LL($\nu, \alpha, \beta, \theta$)	521.576	529.576	531.115	535.312	1.606	0.223	0.079
TLEHLOBX-LL(ν, α, θ)	524.258	530.258	531.147	534.560	1.852	0.237	0.068
TLHLOBX-LL(ν, β, θ)	523.210	529.21	530.099	533.21	1.763	0.233	0.079
HLOBX-LL(β, θ)	523.211	527.211	527.640	530.079	1.763	0.232	0.060
LL(θ)	614.426	616.426	616.564	617.860	1.972	0.625	3.2×10^{-12}
W(λ)	624.894	626.894	627.032	628.328	1.562	0.727	3.3×10^{-16}

The parameter estimates with standard errors in parentheses and GoF statistics are shown in Tables 4.19 and 4.20, respectively. The model adequacy statistics show that the TIHT-LL(ρ, β, θ) model consistently exhibits the lowest possible values across various statistical measures under consideration. These findings indicate that the TIHT-LL(ρ, β, θ) model best describes the COVID-19 situation during the summer season. Furthermore, as evidenced by significantly higher K-S p-value, the TIHT-LL(ρ, β, θ) model outperforms the other models. Based on these results, we can conclude that the TIHT-LL(ρ, β, θ) model is the most suitable choice for accurately representing the dataset under consideration.

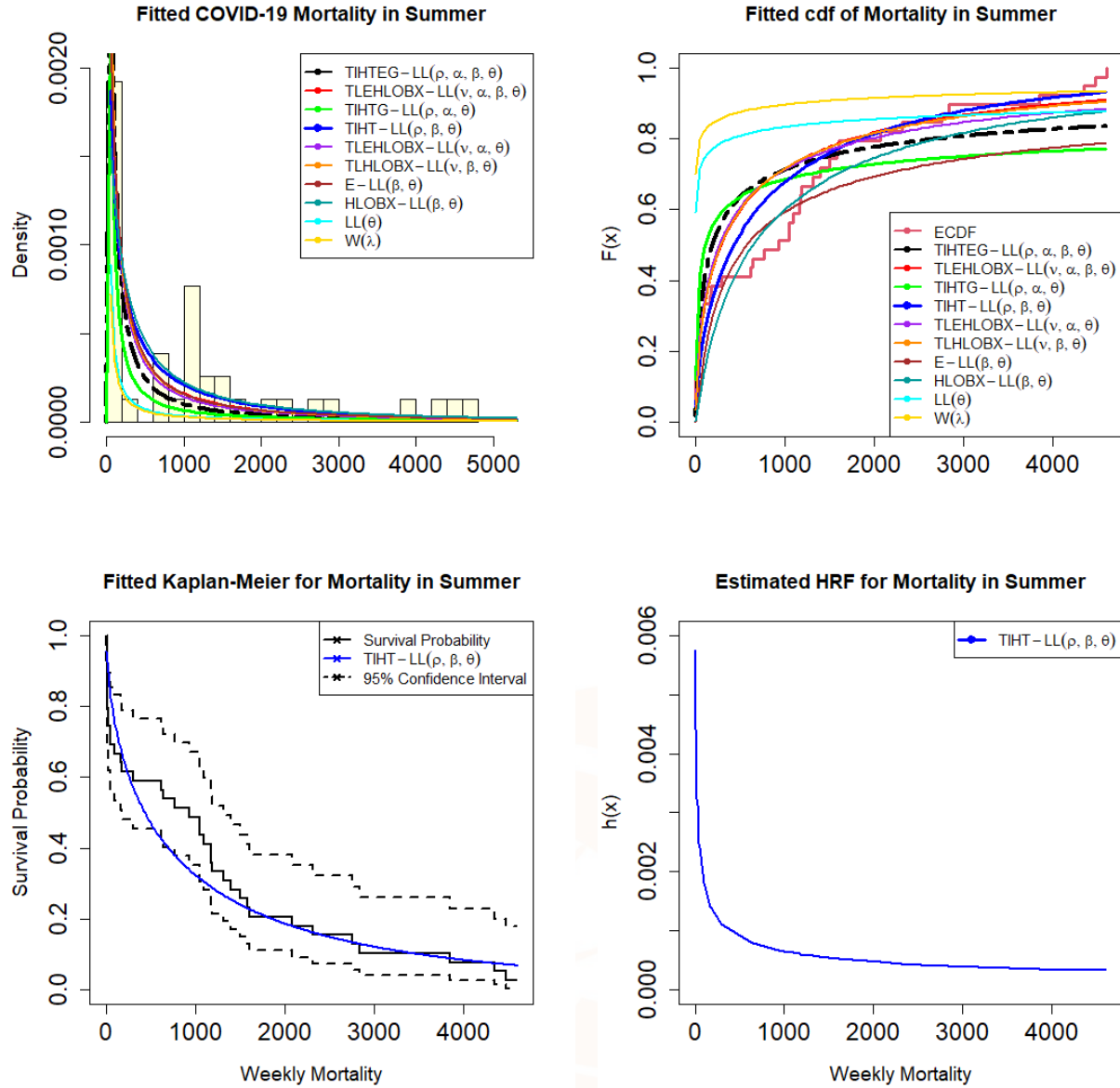


Figure 4.10: Fitted densities, Ecdf, Kaplan-Meier Survival Curve, and hazard rate function plot for mortality in summer.

Figure 4.10 shows the fitted densities, estimated cdf plots, Kaplan-Meier survival plot, and hazard rate function plot for the summer data. The fitted Kaplan-Meier plots show that all of the mortality data points fall within the confidence region, implying that this model adequately describe the characteristics of deaths reported in summer. The hazard rate function plot shows that the hazard of the deaths decreased over time. This is because, as the days progressed, fewer weekly deaths were being reported.

4.4 Results for Administered Vaccinations in Southern African Region

In this section, we present the numerical and graphical results of the administered vaccinations in the Southern African region.

Table 4.21: Estimated parameters for the fitted distributions.

Distribution	Estimates and Standard Errors			
	ρ	α	β	θ
TIHTEG-LL($\rho, \alpha, \beta, \theta$)	51.902(0.669)	0.086(0.025)	27.716(4.720)	1.156(0.339)
TIHTG-LL(ρ, α, θ)	0.227(0.043)	0.075(0.013)		4.962(0.002)
TIHT-LL(ρ, β, θ)	16032.000(0.004)		55.256(2.974)	0.061(0.004)
E-LL(β, θ)			265.660(2.5×10^{-6})	0.437(0.008)
	ν	α	β	θ
TLEHLOBX-LL($\nu, \alpha, \beta, \theta$)	0.689(0.119)	20.576(0.004)	52.887(0.001)	0.020(4.1×10^{-4})
TLEHLOBX-LL(ν, α, θ)	0.693(0.123)	882.290(8.2×10^{-5})		0.019(4.3×10^{-4})
TLHLOBX-LL(ν, β, θ)	2.250(0.384)		543.010(0.001)	0.019(4.0×10^{-4})
HLOBX-LL(β, θ)			1344.200(5.1×10^{-10})	0.020 (5.6×10^{-4})
LL(θ)				θ 0.113(0.009)
W(λ)				λ 0.059(0.004)

Table 4.22: GoF statistics for COVID-19 administered vaccinations in Southern Africa.

Distribution	Model Selection Criteria						
	$-2 \log \mathcal{L}$	AIC	CAIC	BIC	A*	K-S	P-Value
TIHTEG-LL($\rho, \alpha, \beta, \theta$)	2991.779	2999.779	3000.209	3010.118	1.414	0.098	0.282
TIHTG-LL(ρ, α, θ)	3314.464	3320.464	3320.719	3328.219	4.686	0.416	4.4×10^{-16}
TIHT-LL(ρ, β, θ)	2976.355	2982.355	2982.610	2990.110	0.602	0.072	0.658
E-LL(β, θ)	3093.909	3097.909	3098.036	3103.079	7.941	0.268	1.0×10^{-7}
TLEHLOBX-LL($\nu, \alpha, \beta, \theta$)	3025.655	3033.655	3034.085	3043.995	3.621	0.148	0.024
TLEHLOBX-LL(ν, α, θ)	3031.990	3037.990	3038.246	3045.745	4.023	0.156	0.015
TLHLOBX-LL(ν, β, θ)	3027.883	3033.883	3034.139	3041.638	3.756	0.146	0.028
HLOBX-LL(β, θ)	3030.767	3034.767	3034.894	3039.937	3.886	0.159	0.012
LL(θ)	3470.273	3472.273	3472.315	3474.858	3.2426	0.741	4.4×10^{-16}
W(λ)	3500.727	3502.727	3502.769	3505.312	2.566	0.817	4.4×10^{-16}

Table 4.21 displays the parameter estimates with standard errors in parentheses, while Table 4.22 presents the results of the model evaluation criteria. Based on the GoF statistics shown in Table 4.22, it is clear that the TIHT-LL(ρ, β, θ) model has the lowest GoF statistics and the highest K-S p-value. As a result, we conclude that the TIHT-LL(ρ, β, θ) distribution provides the most accurate representation for describing COVID-19 vaccinations in Southern Africa. Furthermore, the modelled densities and cdf plots show that, with the exception of the TIHTG-LL(ρ, α, θ) model, all of the other fitted models closely align with the observed data. The survival function plots further confirm that the selected model accurately fits the data, as it falls within the Kaplan-Meier confidence interval. Hence, it is evident from this plot that the three-parameter TIHT-LL(ρ, β, θ) distribution is the most appropriate for analyzing the number of administered vaccinations. Furthermore, the hazard rate function plot depicts an increasing and decreasing hazard function. In summary, the analysis and visualizations presented in the Table 4.22 and Figure 4.11 provide clarity regarding the selection of the TIHT-LL (ρ, β, θ) model for describing COVID-19 administered vaccinations in Southern Africa.

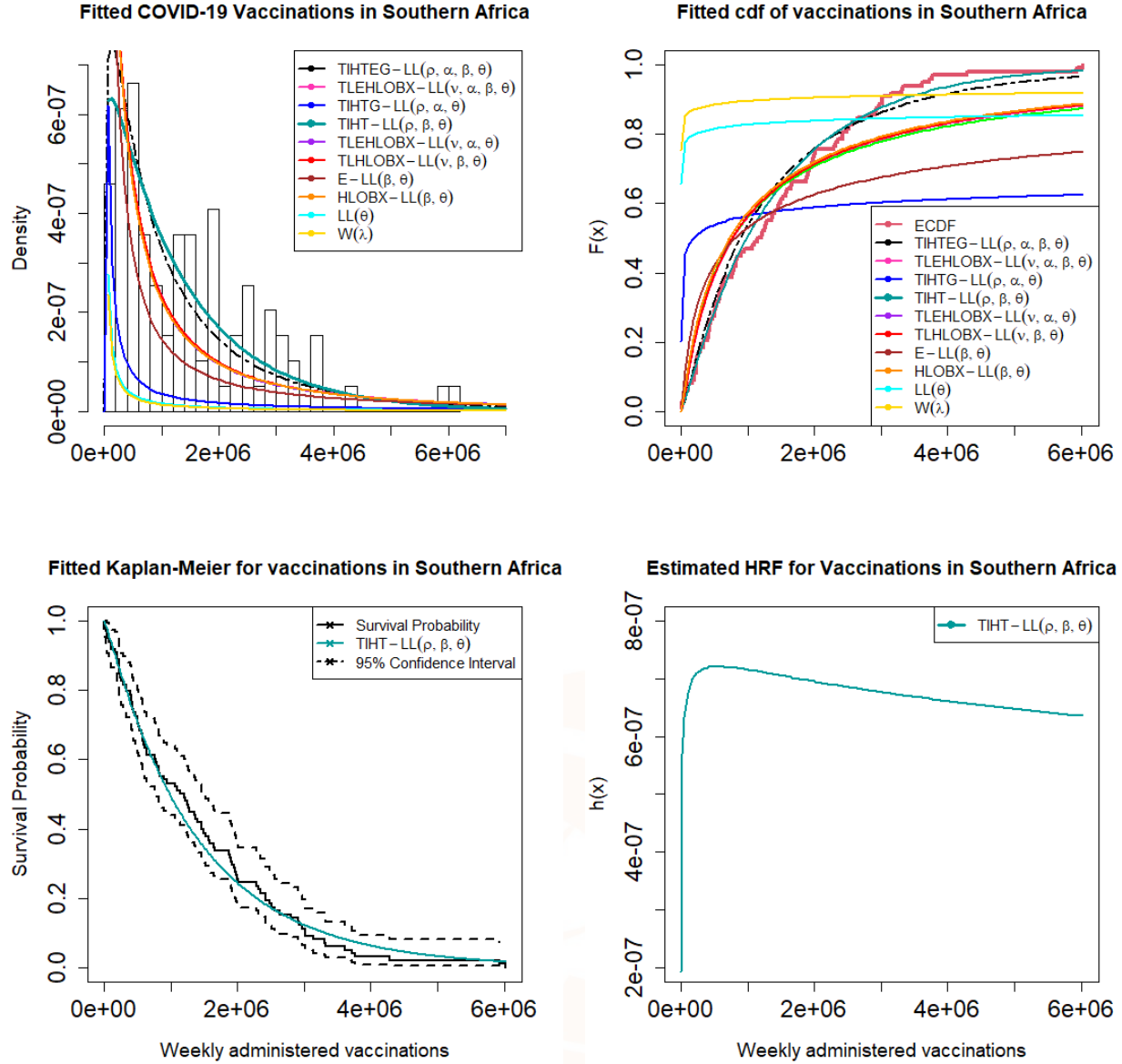


Figure 4.11: Fitted densities, Ecdf, Kaplan-Meier Survival Curve, and Hazard Rate Function Plot for Administered Vaccinations in Southern Africa.

4.5 Discussions

This chapter presented a comprehensive analysis of the fitted models, bearing in mind the principle of model parsimony. From the results presented, it is worth mentioning that the parameter estimates that were obtained through the technique of maximum likelihood estimation are significant. The evaluation of the model fit statistics, including the $-2\log \mathcal{L}$, AIC, CAIC, and BIC, A^* , K-S, and its p-value, showed that the TIHT-LL(ρ, β, θ) distribution adequately approximated the COVID-19 weekly incidence, mortality, and administered vaccinations in the Southern African region. Furthermore, the GoF statistics also showed that TIHT-LL(ρ, β, θ) distribution was best suited for approximating the seasons datasets under consideration, while the TLHLOBX-LL(ν, β, θ) distribution was found to be suited for describing mortality reported during the spring season. Overall, these findings indicate that the chosen models effectively

capture the observed patterns and offer insightful information about the evolution of COVID-19 in the Southern African region.



5. Conclusions

5.1 Introduction

This dissertation presented and analysed various statistical models for modelling weekly COVID-19 incidence, mortality, and administered vaccinations in the Southern Africa region. Furthermore, the study also presented and analysed the weekly seasonal COVID-19 Incidence and mortality data sets in the Southern African region. The importance of monitoring COVID-19, particularly during its peaks, makes modelling weekly COVID-19 data crucial for decision-makers in the epidemiology sector. This chapter presents the conclusions based on the findings from Chapters 2 and 4. Furthermore, the conclusions drawn from this project provide a platform for further research on different countries in the SADC region.

5.2 Conclusions

The presentation of the weekly COVID-19 dataset using graphical (line plots, Q-Q plots, and box plots) and numerical (summary statistics) summaries was discussed in Chapter 2. Section 2.3.1 revealed that the Southern African region recorded high numbers of weekly confirmed cases and deaths during the winter and summer periods. Therefore, it was of paramount importance to have a clear understanding of the disease pattern, which would assist epidemiologists in closely monitoring these seasons. Hence, COVID-19 incidence and mortality data were partitioned into four seasons according to the Southern hemisphere. The normality of the COVID-19 data was examined using the Q-Q and box plots (Section 2.3.2) and numerical descriptive summaries, including values for the coefficients of skewness and kurtosis (Section 2.4.1). Visual inspection of the box and Q-Q plots determined that the data was not normally distributed but right-skewed, with long tails and outliers. The values of the coefficients of skewness and kurtosis further confirmed that the COVID-19 data for the time period under consideration was not normally distributed but rather positively skewed to the right with long tails. The determination of distributions suitable for modelling COVID-19 was an important aspect of this study.

Chapter 4 extended the comparative analysis discussed in Chapter 2 and the methods in Chapter 3 by fitting the probabilistic distributions, and the results demonstrated that the estimated parameters and goodness-of-fit statistics provided sufficient evidence for the validity of these generalized distributions. The TIHT-LL(ρ, β, θ) distribution was found to be appropriate in modelling incidence, mortality and administered vaccinations in Southern Africa. Furthermore, the TIHT-LL(ρ, β, θ) distribution was also found to be appropriate in modelling seasonal COVID-19 incidence and mortality in the region, while the TLHLOBX-LL(ν, β, θ) distribution was found to be better suited for modelling incidence in spring. The decision was made based on the models that gave lower values of the $-2\log \mathcal{L}$, AIC, CAIC, BIC, A^* , and K-S statistics. A similar conclusion was reached based on the high Kolmogorov-Smirnov p-value of the goodness-of-fit

statistic, which further confirmed that the distribution was appropriate for the data under consideration. Furthermore, the fitted densities and the superimposed Kaplan-Meier survival curves also confirm that the models fit the data very well. Moreover, the hazard rate function plots for these distributions displayed a general decreasing rate for COVID-19 data over time. However, it is important to note that a decreasing hazard does not necessarily mean that there are no incidence or mortality from COVID-19, and monitoring the situation closely is still necessary to prevent any potential surges in incidence and mortality and to provide insights into the trend of the vaccination rate. Overall, the decreasing hazard rate of COVID-19 incidence and mortality in Southern Africa was an encouraging trend.

5.3 Limitations

One of the limitations of this dissertation is that we only used aggregated weekly data for the overall Southern African region. It would have been interesting to analyse the disaggregated COVID-19 data for each country in the region. Generally, the lack of key covariates such as age, sex, type of vaccine, comorbidity status, etc.. was a major drawback of this study.

5.4 Areas for Further Research

Different areas for further research have also been considered, and these include modelling COVID-19 in Southern African countries as well as at the provincial level since the patterns are expected to vary, developing the TIHT-LL and TLHLOBX-LL regression models, incorporating covariates such as age, gender, types of vaccination, and comorbidity status, just to mention a few, and the use of other estimation techniques, such as Bayesian estimation.

A. R Code for Chapter Two on Southern Africa Data

```
# Graphical Presentation of the Data
library(lubridate)
library(ggplot2)
library(summarytools)
library(psych)
library(ggthemes)
Data<-read.csv("Data.csv")
Data$Date<-dmy(Data$Date)
  ggplot(aes(x= Date,y= data,color=variable))+
  geom_area(colour="black",size=0.2,alpha=0.8,fill="lightblue")
  +labs(title = " ",x = "",y="",caption = "")+
theme(axis.title.x = element_text(size = 15,face = "bold",colour="black"),
axis.title.y = element_text(size = 10,color = "red",face = "italic"),
axis.text = element_text(size = 8,color = "black"),
  plot.title = element_text(size = 15),
panel.background=element_rect("white"),
  legend.position="none",
strip.background=element_blank(),strip.text = element_blank(),
axis.text.x = element_text(angle = 90,hjust = 1))
#
# Boxplots
Boxplot<-Data
ggplot(aes(y= data,fill= variable)) +
  geom_boxplot(width=.4,show.legend = FALSE,
  colour="red2",fill="burlywood1")+
  stat_boxplot(geom = "errorbar",width = 0.1)+
  scale_x_discrete(name = "") +
  theme(axis.line=element_line(colour="black"),
  strip.background=element_blank(),
  strip.text.x = element_blank(),
  strip.text.y = element_blank()) +
  theme_light()+
  theme(plot.background = element_blank(),
        panel.grid.major = element_blank(),
        panel.border = element_blank())+
```

```
theme(axis.line = element_line(color="black"))
Boxplot
#-----
#####QQplots
QQ_Plots<-Data
  ggplot(aes(sample = Varibale)) +
  stat_qq_band(bandType = "pointwise", fill = "", alpha = 0.4) +
  stat_qq_line(colour = "#8DA0CB") +
  stat_qq_point() +
  xlab("Normal quantiles") +
  ylab("Sample quantiles") +
  theme_light()
QQ_Plots
#-----
#####Summary Statistics
describe(Data)
#-----
```

Date	Cases	Deaths	Date	Cases	Deaths	Date	Cases	Deaths	Date	Cases
7/3/2020	1	5	6/2/2021	51145	2837	8/1/2022	136443	1611	10/12/2022	2677
14/3/2020	24	0	13/2/2021	36236	2312	15/1/2022	88035	1317	17/12/2022	3265
21/3/2020	184	0	20/2/2021	29479	1583	22/1/2022	48764	1178	24/12/2022	2545
28/3/2020	1015	3	27/2/2021	27518	1185	29/1/2022	39857	1051	31/12/2022	2791
4/4/2020	379	9	6/3/2021	20001	1091	5/2/2022	30370	1094	7/1/2023	1968
11/4/2020	542	22	13/3/2021	17381	783	12/2/2022	26791	1170	14/1/2023	3712
18/4/2020	832	26	20/3/2021	16992	1010	19/2/2022	23987	1500	21/1/2023	6882
25/4/2020	1557	32	27/3/2021	14744	741	26/2/2022	61575	934	28/1/2023	4130
2/5/2020	1852	37	3/4/2021	13769	476	5/3/2022	17352	413	4/2/2023	4227
9/5/2020	3078	64	10/4/2021	10670	409	12/3/2022	16059	222	11/2/2023	3386
16/5/2020	5233	72	17/4/2021	14584	550	19/3/2022	13785	183	18/2/2023	3027
23/5/2020	6986	153	24/4/2021	16963	495	26/3/2022	11412	97	25/2/2023	2299
30/5/2020	9710	218	1/5/2021	13074	380	2/4/2022	11778	114		
6/6/2020	14694	851	8/5/2021	16072	457	9/4/2022	10904	61		
13/6/2020	19191	918	15/5/2021	21409	528	16/4/2022	11056	70		
20/6/2020	26561	927	22/5/2021	26510	763	23/4/2022	17760	161		
27/6/2020	37924	1186	29/5/2021	34339	730	30/4/2022	31934	82		
4/7/2020	54110	1024	5/6/2021	43978	740	7/5/2022	43822	157		
11/7/2020	75901	965	12/6/2021	65421	1002	14/5/2022	55050	258		
18/7/2020	90603	1060	19/6/2021	100960	1442	21/5/2022	47107	189		
25/7/2020	88557	1626	26/6/2021	138473	1968	28/5/2022	30703	233		
1/8/2020	76841	1779	3/7/2021	169884	2638	4/6/2022	24873	191		
8/8/2020	57663	2063	10/7/2021	189892	3713	11/6/2022	17893	208		
15/8/2020	39037	1800	17/7/2021	172999	4074	18/6/2022	11982	147		
22/8/2020	29879	1381	24/7/2021	142162	4163	25/6/2022	17199	217		
29/8/2020	21912	1013	31/7/2021	132732	4133	2/7/2022	8320	113		
5/9/2020	19481	988	7/8/2021	125830	3798	9/7/2022	5120	93		
12/9/2020	15319	773	14/8/2021	106820	3326	16/7/2022	6364	64		
19/9/2020	15524	561	21/8/2021	110794	3073	23/7/2022	4327	47		
26/9/2020	14980	503	28/8/2021	99642	2788	30/7/2022	4459	67		
3/10/2020	12869	645	4/9/2021	74017	2442	6/8/2022	2521	14		
10/10/2020	13566	690	11/9/2021	54923	1795	13/8/2022	3491	16		
17/10/2020	16769	127	18/9/2021	44465	1599	20/8/2022	3926	98		
24/10/2020	16719	270	25/9/2021	28586	1281	27/8/2022	2175	25		
31/10/2020	15663	120	2/10/2021	20159	953	3/9/2022	2560	31		
7/11/2020	14903	110	9/10/2021	12198	639	10/9/2022	1555	27		
14/11/2020	15952	174	16/10/2021	10595	435	17/9/2022	2329	25		
21/11/2020	20048	255	23/10/2021	7506	409	24/9/2022	2393	24		
28/11/2020	22815	590	30/10/2021	5332	307	1/10/2022	2106	22		
5/12/2020	27487	622	6/11/2021	8934	188	8/9/2022	2414	13		
12/12/2020	44446	1049	13/11/2021	3436	206	15/10/2022	5635	53		
19/12/2020	62485	1395	20/11/2021	4046	114	22/10/2022	2670	15		
26/12/2020	90029	2077	17/11/2021	28095	242	29/10/2022	2726	56		
2/1/2021	100301	2758	4/12/2021	54927	176	5/11/2022	2654	55		
9/1/2021	140322	3856	11/12/2021	145181	168	12/11/2022	3676	11		
16/1/2021	149420	4608	18/12/2021	224890	301	19/11/2022	5134	26		
23/1/2021	114291	4478	25/12/2021	184462	642	26/11/2022	3583	44		
30/1/2021	79913	4351	1/1/2022	154363	772	3/12/2022	2052	46		

Deaths	Date	Vacc	Date	Vacc	Date	Vacc
92	20/2/2021	10453	22/1/2022	2911905	24/12/2022	1142481
24	27/2/2021	70085	29/1/2022	2929205	31/12/2022	316
12	6/3/2021	50559	5/2/2022	948286	7/1/2023	274243
9	13/3/2021	100965	12/2/2022	1365831	14/1/2023	511607
5	20/3/2021	99238	19/2/2022	2313425	21/1/2023	66324
12	27/3/2021	222044	26/2/2022	1456813	28/1/2023	42740
43	3/4/2021	200871	5/3/2022	1966564	4/2/2023	50857
22	10/4/2021	229780	12/3/2022	1253415	11/2/2023	108161
14	17/4/2021	226876	19/3/2022	1566741	18/2/2023	242040
6	24/4/2021	332241	26/3/2022	1498501	25/2/2023	1532686
9	1/5/2021	239517	2/4/2022	1984869		
10	8/5/2021	286997	9/4/2022	831261		
	15/5/2021	777357	16/4/2022	433761		
	22/5/2021	564308	23/4/2022	415080		
	29/5/2021	574070	30/4/2022	1067638		
	5/6/2021	502946	7/5/2022	637445		
	12/6/2021	822058	14/5/2022	1470093		
	19/6/2021	608769	21/5/2022	1096428		
	26/6/2021	898539	28/5/2022	1662206		
	3/7/2021	828555	4/6/2022	1353527		
	10/7/2021	1219437	11/6/2022	1260184		
	17/7/2021	1286855	18/6/2022	479734		
	24/7/2021	1922305	25/6/2022	422843		
	31/7/2021	1649144	2/7/2022	1525619		
	7/8/2021	1582971	9/7/2022	756141		
	14/8/2021	2348592	16/7/2022	1203576		
	21/8/2021	2428980	23/7/2022	795782		
	28/8/2021	2644448	30/7/2022	511641		
	4/9/2021	1901237	6/8/2022	337519		
	11/9/2021	2485161	13/8/2022	6028223		
	18/9/2021	3011780	20/8/2022	537330		
	25/9/2021	1957941	27/8/2022	647314		
	2/10/2021	1409882	3/9/2022	233576		
	9/10/2021	2556494	10/9/2022	411224		
	16/10/2021	1994470	17/9/2022	202513		
	23/10/2021	2975006	24/9/2022	397605		
	30/10/2021	2802718	1/10/2022	444667		
	6/11/2021	3319989	8/10/2022	499789		
	13/11/2021	3720915	15/10/2022	2006581		
	20/11/2021	3138527	22/10/2022	5923606		
	27/11/2021	2516875	29/10/2022	600154		
	4/12/2021	3302874	5/11/2022	261821		
	11/12/2021	1870050	12/11/2022	174425		
	18/12/2021	2416785	19/11/2022	3760702		
	25/12/2021	2285799	26/11/2022	580535		
	1/1/2022	3610960	3/12/2022	4291285		
	8/1/2022	1892096	10/12/2022	22678		
	15/1/2022	3023546	17/12/2022	46445		

B. R Code for Chapter Two on Southern Africa Data

```
#-----
#APPLICATION
rm(list=ls())
library(stats4)
library(bbmle)
library(stats)
library(numDeriv)
library('bbmle')
x<- c(#Insert your data here )
#TIHT-LL
TIHT-LL<- function(a, beta,c) {
  -sum(log(#Insert your pdf here))
}
TIHT-LL.result<-mle2(TIHT-LL
, hessian = NULL, start=list(a=.1 ,beta=.3,c=.01), optimizer="nlminb", lower=0)
summary(TIHT-LL.result)
#-----
#FITTED DENSITIES
par(mfrow=c(0,4))
x<-c(#Insert data here)
##Type I Heavy Tailed Log Logistic distribution
TIHT-LL<-function(a, beta,c,x)
{
#Insert PDF here
}
#You can add models here
hist(x, prob=TRUE, br=15,ylim=c(0,2.0), main='')
curve(TIHT-LL(x,a=6.6019e+03 ,beta=3.8574e+01,c=5.2937e-02 ),
lty=6, col=2, add=TRUE, lwd=3)
par(new=TRUE)
#-----

# Codes for plotting Kaplan-Meier survival plot of the model
x=c(#Insert data here)
library(survival)
delta=rep(1,length(x))
```

```
x<-sort(x)
km = survfit(Surv(x,delta)~1)
plot(km, conf.int=FALSE, ylab="Kaplan-Meier Survival Plot", xlab="x")
## Write down the parameters of the model
and the corresponding estimated values
a=6.6019e+03
beta=3.8574e+01
c=5.2937e-02
ss <-function(x)
{
  #Insert the survival function here
}
lines(seq(0.05,2.72, length.out=100),ss(seq(0.05,
2.72, length.out=100)),col="green", lwd=2)

#
```

References

- T. Achoki, U. Alam, L. Were, T. Gebremedhin, F. Senkubuge, A. Lesego, S. Liu, R. Wamai, and Y. Kinfu. COVID-19 Pandemic in the African Continent: Forecasts of Cumulative Cases, New Infections, and Mortality. *MedRxiv*, 2020.
- O. Adubisi, A. Abdulkadir, H. Chiroma, and U. Abbas. The Type I Half Logistic Skew-T Distribution: A Heavy-Tail Model with Inverted Bathtub Shaped Hazard Rate. *Asian Journal of Probability and Statistics*, 14(4):21–40, 2021.
- M. Ahsan-ul Haq, M. Ahmed, J. Zafar, and P. L. Ramos. Modeling of COVID-19 Cases in Pakistan Using Lifetime Probability Distributions. *Annals of Data Science*, 9(1):141–152, 2022.
- I. u. H. Akhtar. Understanding the COVID-19 Pandemic Curve Through Statistical Approach. *medRxiv*, pages 2020–04, 2020.
- A. Al-Shomrani, O. Arif, A. Shawky, S. Hanif, and M. Q. Shahbaz. Topp-Leone Family of Distributions: Some Properties and Application. *Pakistan Journal of Statistics and Operation Research*, pages 443–451, 2016.
- M. Alsulami, H. Abu-Zinadah, and A. H. Ibrahim. Machine Learning Model and Statistical Methods for COVID-19 Evolution Prediction. *Wireless Communications and Mobile Computing*, 2021, 2021.
- A. Alzaatreh, C. Lee, and F. Famoye. A New Method for Generating Families of Continuous Distributions. *Metron-International Journal of Statistics*, 71(1):63–79, 2013.
- K. Amouzouvi, K. A. Assamagan, S. Azote, S. H. Connell, J. B. F. Fankam, F. Fanomezana, A. Guga, C. E. Haliya, T. S. Mabote, F. F. Macucule, et al. A Model of COVID-19 Pandemic Evolution in African Countries. *Scientific African*, 14:e00987, 2021.
- I. O. Ayenigbara, J. S. Adegboro, G. O. Ayenigbara, O. R. Adeleke, and O. O. Olofintuyi. The Challenges to a Successful COVID-19 Vaccination Programme in Africa. *Germes*, 11(3):427, 2021.
- N. Balakrishnan. Order Statistics From the Half Logistic Distribution. *Journal of statistical computation and simulation*, 20(4):287–309, 1985.
- B. Bolker and R Development Core Team. *bbmle: Tools for General Maximum Likelihood Estimation*, 2022. URL <https://CRAN.R-project.org/package=bbmle>. R package version 1.0.25.
- M. C. Bryson. Heavy-Tailed Distributions: Properties and Tests. *Technometrics*, 16(1):61–68, 1974.

- G. Chen and N. Balakrishnan. A General Purpose Approximate Goodness-of-Fit Test. *Journal of Quality Technology*, 27(2):154–161, 1995.
- D. Chikobvu and C. Sigauke. Statistical Distribution Fitting to the Number of COVID-19 Deaths in South Africa. Preprint on Research Square: <https://www.researchsquare.com/article/rs-32411/v2>, 2020.
- F. Chipepa and B. Oluyede. The Topp-Leone Odd Exponential Half Logistic-G Family of Distributions : Model, Properties and Applications. *Pakistan Journal of Statistics*, 37(3): 253–277, 2021.
- G. M. Cordeiro, E. M. Ortega, and D. C. da Cunha. The Exponentiated Generalized Class of Distributions. *Journal of data science*, 11(1):1–27, 2013.
- O. O. Daniel, A. T. Michael, and O. E. Ayooluwa. Modeling COVID-19 Deaths in Nigeria Using Exponential-Gamma Distribution. *International Journal of Multidisciplinary Research and Analysis*, 5(6):1456–1460, 2022.
- L. De Alencastro, I. Locatelli, C. Clair, M. H. Ebell, and N. Senn. Correlation of Clinical Decision-Making with Probability of Disease: A Web-Based Study Among General Practitioners. *Plos one*, 15(10):e0241210, 2020.
- T. V. F. De Santana, E. M. Ortega, G. M. Cordeiro, and G. O. Silva. The Kumaraswamy-Log-Logistic Distribution. *Journal of Statistical Theory and Applications*, 11(3):265–291, 2012.
- P. R. Diniz Marinho, M. Bourguignon, and C. R. Barros Dias. *AdequacyModel: Adequacy of Probabilistic Models and General Purpose Optimization*, 2016. URL <https://CRAN.R-project.org/package=AdequacyModel>. R package version 2.0.0.
- P. Dolton. The Statistical Challenges of Modelling COVID-19. *National Institute Economic Review*, 257:46–82, 2021.
- T. Dzinamarira, B. Nachipo, B. Phiri, and G. Musuka. COVID-19 Vaccine Roll-Out in South Africa and Zimbabwe: Urgent Need to Address Community Preparedness, Fears and Hesitancy. *Vaccines*, 9(3):250, 2021.
- M. El-Morshedy, E. Altun, and M. Eliwa. A New Statistical Approach to Model the Counts of Novel Coronavirus Cases. *Mathematical Sciences*, 16(1):37–50, 2022.
- E. Gholami, K. Mansori, and M. Soltani-Kermanshahi. Statistical Distribution of Novel Coronavirus in Iran. *International Journal of One Health*, 6(2):143–146, 2020.
- P. Gilbert and R. Varadhan. *numDeriv: Accurate Numerical Derivatives*, 2019. URL <https://CRAN.R-project.org/package=numDeriv>. R package version 2016.8-1.1.
- G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, and M. Colaneri. Modelling the COVID-19 Epidemic and Implementation of Population-Wide Interventions in Italy. *Nature medicine*, 26(6):855–860, 2020.

- G. J. Husak, J. Michaelson, and C. Funk. Use of the Gamma Distribution to Represent Monthly Rainfall in Africa for Drought Monitoring Applications. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 27(7):935–944, 2007.
- F. Kuwonu. As COVID-19 Cases Rise, African Countries Grapple with Safely Easing Lockdowns. <https://www.un.org/africarenewal/magazine/june-2020/coronavirus/covid-19-africa-cases-rise-along-economic-hardship-countries-grapple-safely-easing>, June 2020.
- G. J. Lekono, B. Oluyede, and L. Gabaitiri. A new heavy-tailed exponentiated generalised-g family of distributions: properties and applications. *International Journal of Mathematics in Operational Research*, 27(1):1–34, 2024.
- S. A. Lone and A. Ahmad. COVID-19 Pandemic—An African Perspective. *Emerging microbes & infections*, 9(1):1300–1308, 2020.
- J. Machado and A. M. Lopes. Rare and Extreme Events: the Case of COVID-19 Pandemic. *Nonlinear dynamics*, 100(3):2953–2972, 2020.
- M. Maleki, M. R. Mahmoudi, M. H. Heydari, and K.-H. Pho. Modeling and Forecasting the Spread and Death Rate of Coronavirus (COVID-19) in the World Using Time Series Models. *Chaos, Solitons & Fractals*, 140:110151, 2020.
- A. Mckeever. Every Season Actually Begins Twice. <https://www.nationalgeographic.com/environment/article/history-science-of-meteorological-astronomical-seasons#:~:text=In%20the%20Southern%20Hemisphere%2C%20the,compare%20seasons%20to%20one%20another.>, Mar. 2023.
- T. Moakofi, B. Oluyede, and F. Chipepa. The Topp-Leone Half Logistic Generalized-G Family of Distributions: Model, Properties and Applications. *Eurasian Bulletin of Mathematics (ISSN: 2687-5632)*, 4(2):107–133, 2021.
- N. Mthethwa, R. Chifurira, and K. Chinhamu. Estimating the Risk of SARS-CoV-2 Deaths Using a Markov Switching-Volatility Model Combined with Heavy-Tailed Distributions for South Africa. *BMC Public Health*, 22(1):1–14, 2022.
- M. Ndlovu, R. Moyo, and M. Mpofu. Modelling COVID-19 Infection with Seasonality in Zimbabwe. *Physics and Chemistry of the Earth, Parts A/B/C*, 127:103167, 2022.
- A. Nyabunze and T. Siavhundu. Economic Impact of COVID-19 Induced Lockdown in Zimbabwe. *Diverse Journal of Multidisciplinary Research*, 2(5):1–7, 2020.
- J. Olivier and M. M. Norberg. Positively Skewed Data: Revisiting the Box-Cox Power Transformation. *International Journal of Psychological Research*, 3(1):68–95, 2010.
- B. A. Olumide, M. Saidu, and A. Oluwasesan. Evaluation of Best Fit Probability Distribution Models for the Prediction of Rainfall and Runoff Volume (Case Study Tagwai Dam, Minna-Nigeria). *International Journal of Engineering and Technology*, 2013.

- T. Plocoste, R. Calif, L. Euphrasie-Clotilde, and F.-N. Brute. The Statistical Behavior of PM10 Events Over Guadeloupean Archipelago: Stationarity, Modelling and Extreme Events. *Atmospheric Research*, 241:104956, 2020.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2022. URL <https://www.R-project.org/>.
- A. P. Ramadhan, R. Yendra, and M. M. Rahmadeni. The Modelling Number of Daily Death COVID-19 Data Using Some of Two and One Parameter Distributions in Indonesia. *International Journal of Mathematics Trends and Technology (IJMTT)*, 68:106–111, 2023.
- M. Sinkala, P. Nkhoma, M. Zulu, D. Kafita, R. Tembo, and V. Daka. The COVID-19 Pandemic in Africa: Predictions Using the SIR Model Indicate the Cases are Falling. *medRxiv*, pages 2020–06, 2020.
- N. Stiegler and J.-P. Bouchard. South Africa: Challenges and Successes of the COVID-19 Lockdown. In *Annales Médico-psychologiques, revue psychiatrique*, volume 178, pages 695–698. Elsevier, 2020.
- K. Tawiah, W. A. Iddrisu, and K. Asampana Asosega. Zero-Inflated Time Series Modelling of COVID-19 Deaths in Ghana. *Journal of Environmental and Public Health*, 2021, 2021.
- H. Tegally, E. Wilkinson, C. L. Althaus, M. Giovanetti, J. E. San, J. Giandhari, S. Pillay, Y. Naidoo, U. Ramphal, N. Msomi, et al. Rapid Replacement of the Beta Variant by the Delta Variant in South Africa. *MedRxiv*, pages 2021–09, 2021.
- B. Tlhaloganyang, B. Oluyede, and A. Puoetsile. The New Topp-Leone-Exponentiated Half Logistic-Odd Burr X-G Family of Distributions with Properties and Applications. Under review in journal of Revista Colombiana, 2022. URL <https://revistas.pedagogica.edu.co/index.php/RCE>.
- C. W. Topp and F. C. Leone. A family of J-Shaped Frequency Functions. *Journal of the American Statistical Association*, 50(269):209–219, 1955.
- M. C. Wang and B. J. Bushman. Using the normal quantile plot to explore meta-analytic data sets. *Psychological Methods*, 3(1):46, 1998.
- WHO. WHO Announces COVID-19 Outbreak a Pandemic. Europe: World Health Organization, 2020. <http://www.euro.who.int/en/health-topics/health-emergencies/coronaviruscovid-19/news/news/2020/3/who-announces-COVID-19-outbreak-a-pandemic>, Apr. 2020.
- WHO. COVID-19 Deaths in African Region to Fall by Nearly 94% in 2022: WHO Analysis. <https://www.afro.who.int/news/covid-19-deaths-african-region-fall-nearly-94-2022-who-analysis>, June 2022.
- William Revelle. *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois, 2024. URL <https://CRAN.R-project.org/package=psych>. R package version 2.4.6.

- Z. Zhang. *Statistical Modeling of Daily Confirmed COVID-19 Cases and Deaths in Europe and United States*. PhD thesis, Marquette University, 2021.
- J. Zhao, Z. Ahmad, E. Mahmoudi, E. H. Hafez, and M. M. Mohie El-Din. A New Class of Heavy-Tailed Distributions: Modeling and Simulating Actuarial Measures. *Complexity*, 2021: 1–18, 2021.
- W. Zhao, S. K. Khosa, Z. Ahmad, M. Aslam, and A. Z. Afify. Type-I Heavy Tailed Family with Applications in Medicine, Engineering and Insurance. *PloS one*, 15(8):e0237462, 2020a.
- Z. Zhao, X. Li, F. Liu, G. Zhu, C. Ma, and L. Wang. Prediction of the COVID-19 Spread in African Countries and Implications for Prevention and Control: A Case Study in South Africa, Egypt, Algeria, Nigeria, Senegal and Kenya. *Science of the Total Environment*, 729: 138959, 2020b.
- K. Zografos and N. Balakrishnan. On Families of Beta-and Generalized Gamma-Generated Distributions and Associated Inference. *Statistical methodology*, 6(4):344–362, 2009.

