Analysis of Covid-19 Mortality data : A probabilistic perspective

Hazard Function

Let T be the lifetime of a component in a system. The probability of failure of the system in $[t\ ,\ t+\Delta t]$ given that it had survived up to t is

$$P(t \le T < t+\Delta t \mid T \ge t)$$

The instantaneous failure rate of the system at time point t:

$$\begin{aligned}
\lim_{\Delta t \to 0} \frac{P(t \le T < t+\Delta t \mid T \ge t)}{\Delta t} &= \lim_{\Delta t \to 0} \frac{P(t \le T < t+\Delta t)}{\Delta t\ P(T \ge t)} \\
&= \lim_{\Delta t \to 0} \frac{P(t \le T < t+\Delta t)}{\Delta t\,(1 - F_T(t))} \\
&= \frac{(F_T(t))'}{(1 - F_T(t))} \\
&= \frac{f_T(t)}{S_T(t)}
\end{aligned}$$

Hazard Function: $h(t) = \frac{f_T(t)}{S_T(t)}$

Exponential Distribution

Parameter:

1- Rate: $\lambda > 0$

Distribution Function

$$F_X(x) = 1 - e^{-\lambda x}, \qquad x \geq 0$$

Density Function

$$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

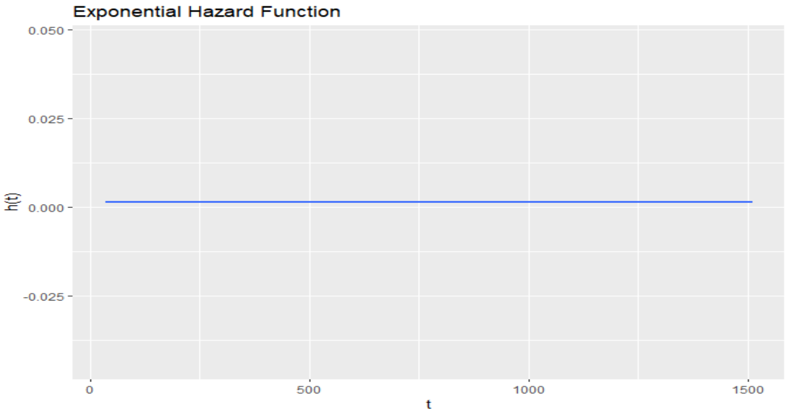
Properties:

- Memoryless: The probability of an event occurring is independent of how much time has elapsed already.

$$P(X > s + t \mid X > s) = P(X > T) \ \forall s, t \geq 0$$

Limitations:

- Constant hazard function



Weibull Distribution

Parameters:

1- Scale: $\alpha > 0$
2- Shape: $\beta > 0$

Distribution Function

$$F_X(x) = 1 - e^{-\alpha x^{\beta}}, \qquad x \geq 0$$

Density Function

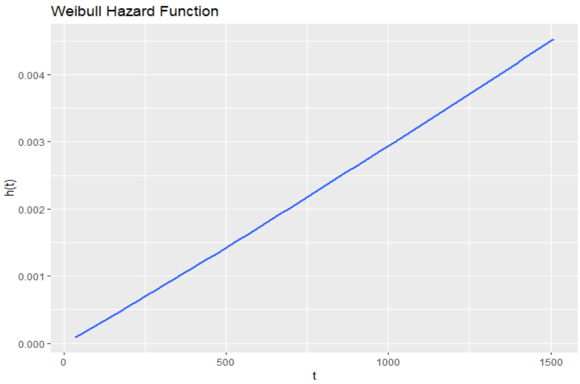
$$f_X(x) = \alpha\beta x^{\beta-1} e^{-\alpha x^{\beta}}, \qquad x \geq 0$$

Properties:

- Flexibility: captures a wide range of shapes of density function
- Shape parameter: models different patterns of the hazard function

Limitations:

- Assumptions: assumes that the hazard function is monotonic
- nple



Generalized Exponential Distribution

Parameters:

1- Scale: $\lambda > 0$
2- Shape: $\alpha > 0$

Distribution Function

$$F_X(x) = \left(1 - \mathrm{e}^{-\lambda \mathrm{x}}\right)^{\alpha}, \qquad x \geq 0$$

Density Function

$$f_X(x) = \alpha\lambda\left(1 - \mathrm{e}^{-\lambda \mathrm{x}}\right)^{\alpha-1} e^{-\lambda x}, \qquad x \geq 0$$

Lehmann Alternatives:

$$G(x) = [\, F(x)\,]^{\alpha}, \qquad \alpha > 0$$

- $\lim\limits_{x \to -\infty} [\, F(x)\,]^{\alpha} = 0\,,\ \lim\limits_{x \to +\infty} [\, F(x)\,]^{\alpha} = 1$
- $\lim\limits_{y \to x^{+}} [\, F(y)\,]^{\alpha} = [\, F(x)\,]^{\alpha}$
- $\dfrac{d}{dx}([\, F(x)\,]^{\alpha}) = \alpha\, [\, F(x)\,]^{\alpha-1} f(x) > 0$

Australian & New Zealand Journal of Statistics

Theory & Methods: Generalized exponential distributions

Rameshwar D. Gupta, Debasis Kundu

First published: 18 December 2002 | https://doi.org/10.1111/1467-842X.00072 | Citations: 642

Generalized Exponential Distribution

Parameters:

1- Scale: $\lambda > 0$
2- Shape: $\alpha > 0$

Distribution Function

$$F_X(x) = \left(1 - \mathrm{e}^{-\lambda \mathrm{x}}\right)^{\alpha}, \qquad x \geq 0$$

Density Function

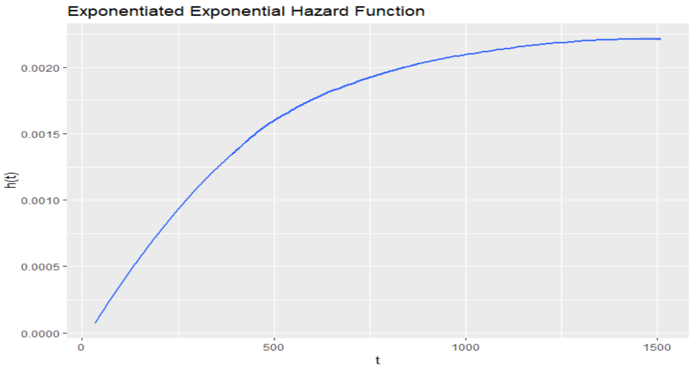
$$f_X(x) = \alpha\lambda\left(1 - \mathrm{e}^{-\lambda \mathrm{x}}\right)^{\alpha-1} e^{-\lambda x}, \qquad x \geq 0$$

Properties:

- Flexibility: captures a wide range of shapes and patterns in the data
- Can be used to analyse skewed data

Limitations:

- Assumptions: hazard function is constant or monotonic
-



Marshall Olkin Distribution Formulation

Suppose $X_1, X_2, \ldots$ are independent and identically distributed random variables with common distribution function $F(\cdot)$ and $N \sim Geo(\theta)$, $0 < \theta < 1$.

Distribution Function : $P(N = n) = \theta(1-\theta)^{n-1} \quad \forall n = 1,2, \ldots$

Consider a random variable $\mathrm{Y} := \min\{X_1, \ldots, X_N\}$

Survival Function: $S_Y(y) = P(\mathrm{Y} \geq y)$

$$= \sum_{n=1}^{\infty} P(y \geq y \mid N = n)\, P(N = n)$$

$$= \sum_{n=1}^{\infty} \left(1 - F(y)\right)^n \theta(1-\theta)^{n-1}$$

$$= \theta \left(1 - F(y)\right) \sum_{n=1}^{\infty} \left(\left(1 - F(y)\right)(1-\theta)\right)^{n-1}$$

The geometric distribution assumption is convenient as it verifies stability property unlike other extreme value distributions (limiting distributions for the minimum/maximum of a large number of identically distributed random variables)

Marshall Olkin Distribution Formulation

Survival Function: $$S_Y(y) = \frac{\theta\left(1 - F(y)\right)}{1 - (1-\theta)\left(1 - F(y)\right)}$$

Distribution Function: $$G_Y(y) = 1 - S_Y(y) = \frac{F(y)}{1 - (1-\theta)\left(1 - F(y)\right)}$$

Density Function: $$g_Y(y) = \frac{d}{dy}\left(G_Y(y)\right) = \frac{\theta\, f(y)}{(1 - (1-\theta)\,(1 - F(y)))^2}$$

A New Method for Adding a Parameter to a Family of Distributions with Application to the Exponential and Weibull Families

Author(s): Albert W. Marshall and Ingram Olkin

Source: *Biometrika*, Sep., 1997, Vol. 84, No. 3 (Sep., 1997), pp. 641-652

Published by: Oxford University Press on behalf of Biometrika Trust

Stable URL: https://www.jstor.org/stable/2337585

Marshall–Olkin Generalized Exponential

Parameters:

1- Scale: $\lambda > 0$
2- Shape 1: $\alpha > 0$
3- Shape 2: $\theta > 0$

Distribution Function

$$F_X(x) = \frac{(1-\ e^{-\lambda x})^{\alpha}}{1-(1-\theta)(1-(1-\ e^{-\lambda x})^{\alpha})}, \qquad x \geq 0$$

Density Function

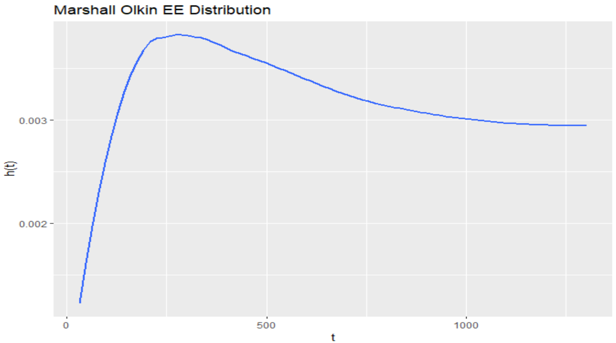
$$f_X(x) = \frac{\alpha\lambda\theta e^{-\lambda x}\,(1-\ e^{-\lambda x})^{\alpha-1}}{(1-(1-\theta)(1-(1-\ e^{-\lambda x})^{\alpha}))^2}, \qquad x \geq 0$$

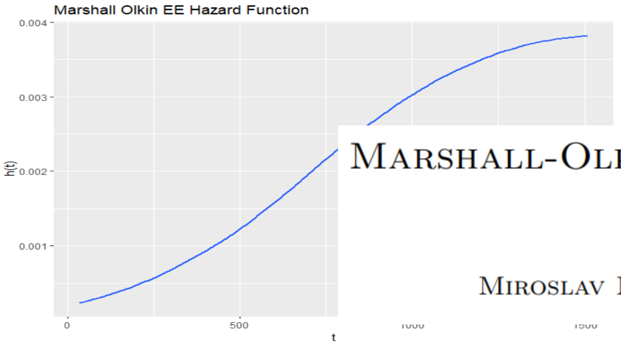
Properties:

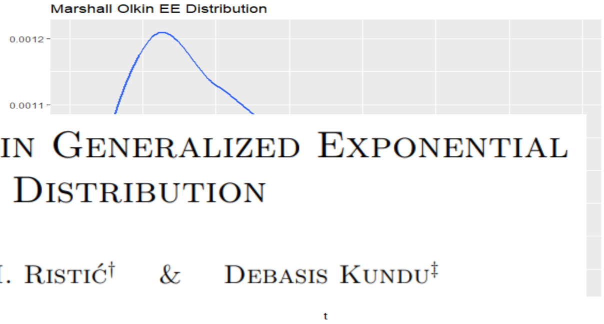
- Flexibility: allows for various forms of dependence between the two variables, it can model positive and negative dependence

Limitations:

- Complexity: relatively complex model that involves multiple parameters that need to be estimated







MARSHALL-OLKIN GENERALIZED EXPONENTIAL DISTRIBUTION

MIROSLAV M. RISTIĆ[†] & DEBASIS KUNDU[‡]

Datasets

01

US

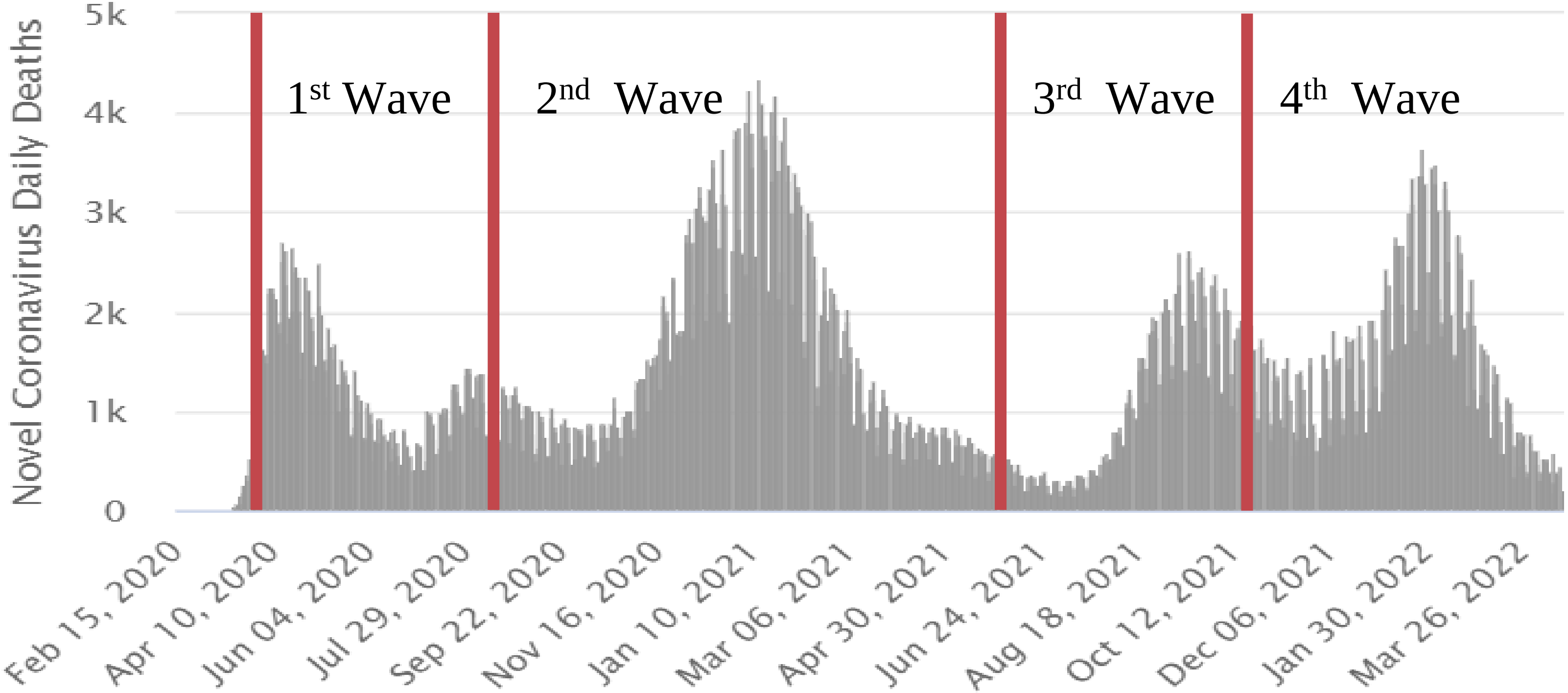
1st Wave: 4/2020 – 8/2020
2nd Wave: 9/2020 – 5/2021
3rd Wave: 6/2021 – 10/2021
4th Wave: 11/2021 – 3/2022

02

Brazil

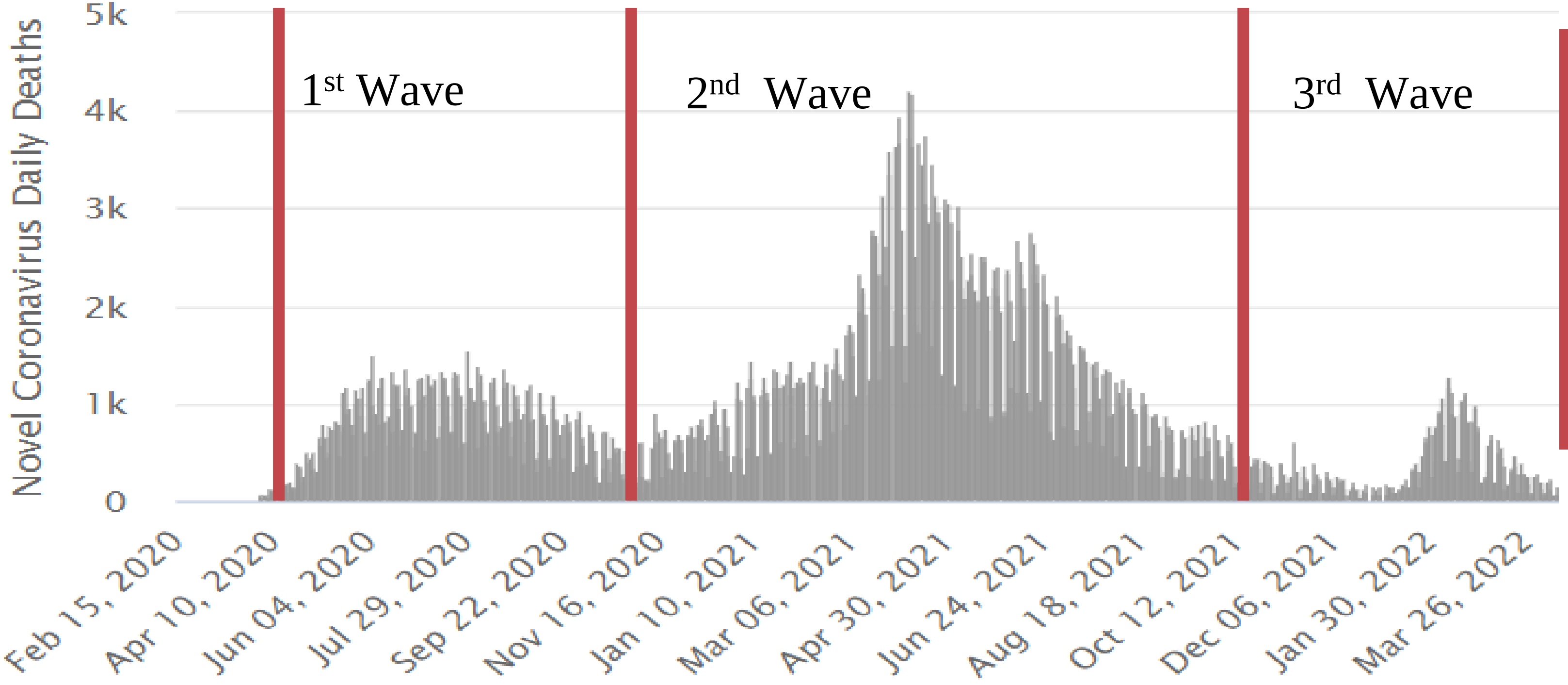
1st Wave: 4/2020 – 10/2020
2nd Wave: 11/2020 – 10/2021
3rd Wave: 11/2021 – 3/2022

US Data



REF: https://www.worldometers.info/coronavirus/country/us/

Brazil Data



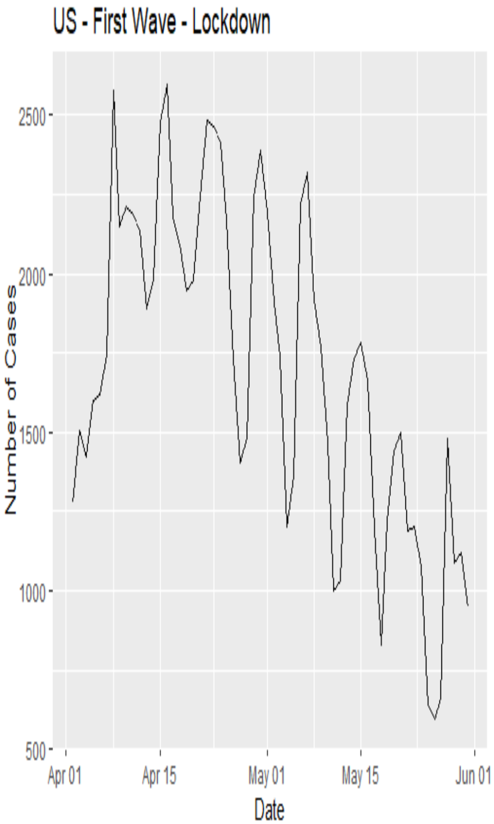
REF: https://www.worldometers.info/coronavirus/country/brazil/

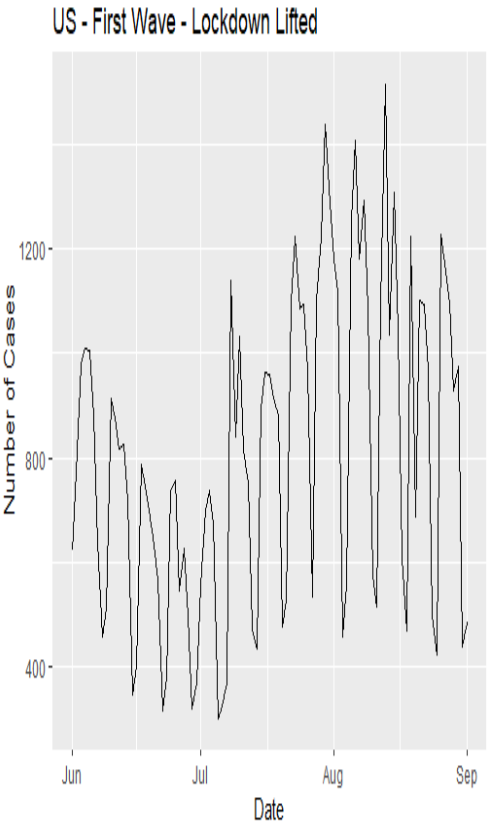
01

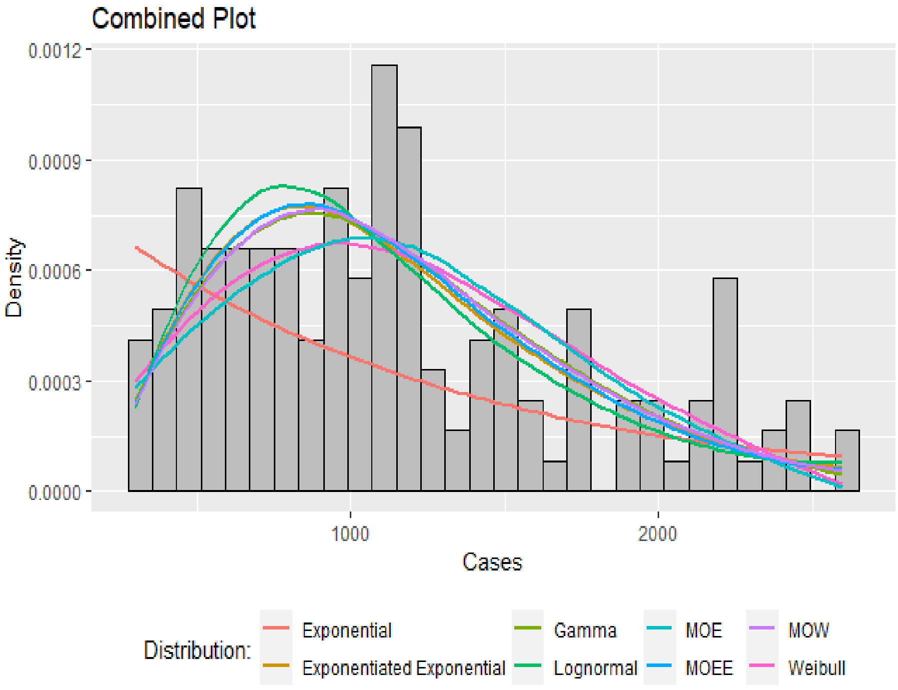
United States

Descriptive Statistics

Minimum	302
Maximum	2598
Median	1086
Mean	1159
Standard Deviation	594.17
Coefficient of Variation	0.51





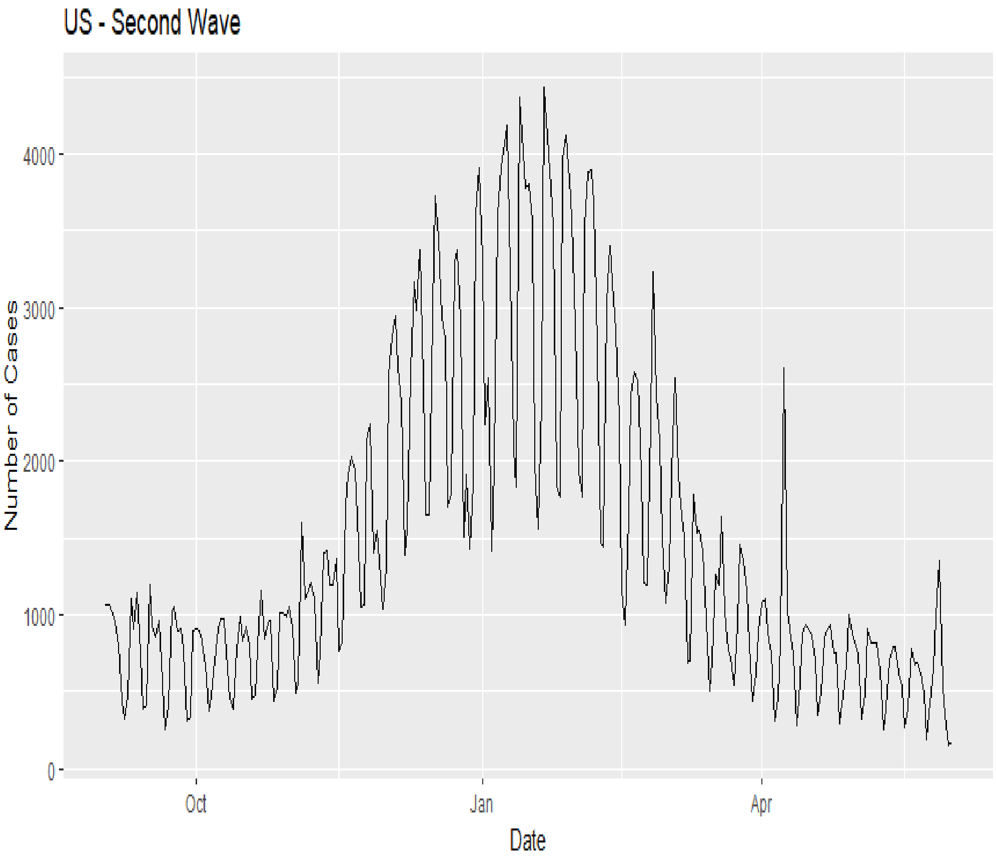


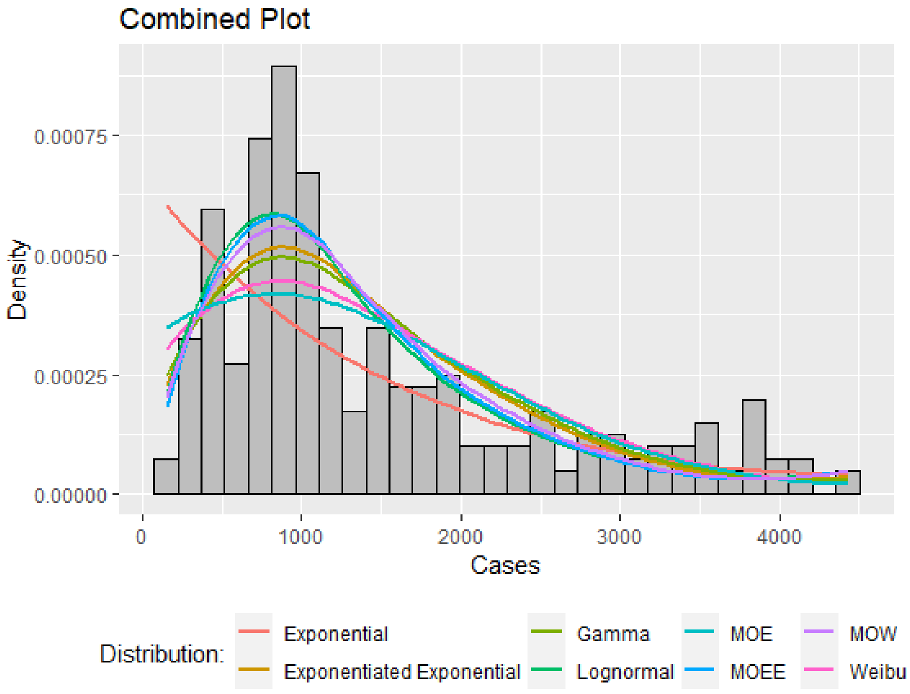
Distribution	Log Likelihood	P - Value
Exponential	-1232.44	1.42e-08
EE	-1180.31	0.9499
Gamma	**-1180.07**	**0.8439**
Weibull	-1182.62	0.2632
Lognormal	-1181.05	0.6835
MOE	-1188.73	0.4489
MOEE	-1180.20	0.8882
MOW	**-1179.90**	**0.7956**

Parameters	Exponential	EE	Gamma	Weibull	Lognormal	MOE	MOEE	MOW
1	$\lambda = 0.00086$	$\alpha = 4.608$	$\alpha = 3.80039$	$k = 2.09485$	$\mu = 6.91785$	$\lambda = 0.0026$	$\alpha = 4.7074$	$\alpha = 2.716$
2	NA	$\lambda = 0.00189$	$\beta = 0.00328$	$\lambda = 1314.19$	$\sigma = 0.53930$	$\theta = 15.860$	$\lambda = 0.0020$	$\lambda = 1.37e\text{-}09$
3	NA	NA	NA	NA	NA	NA	$\theta = 1.189$	$\theta = 0.24761$

Descriptive Statistics

Minimum	149
Maximum	4431
Median	1059
Mean	1493
Standard Deviation	1062.73
Coefficient of Variation	0.71



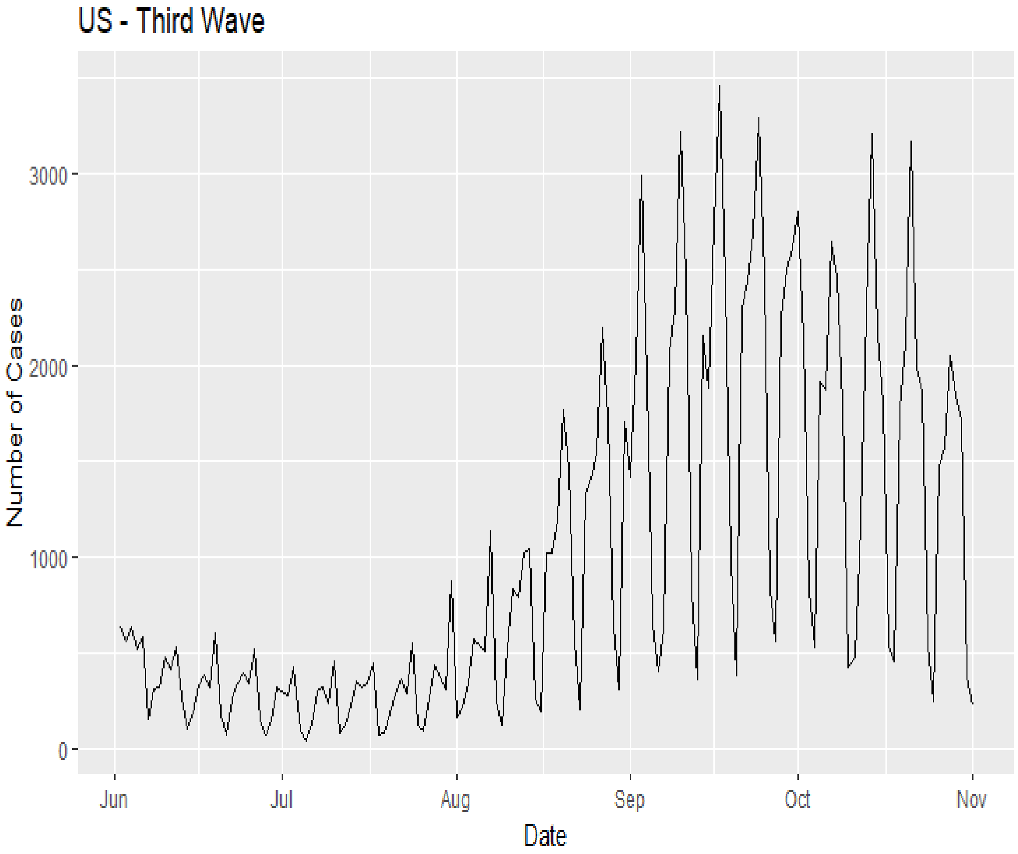


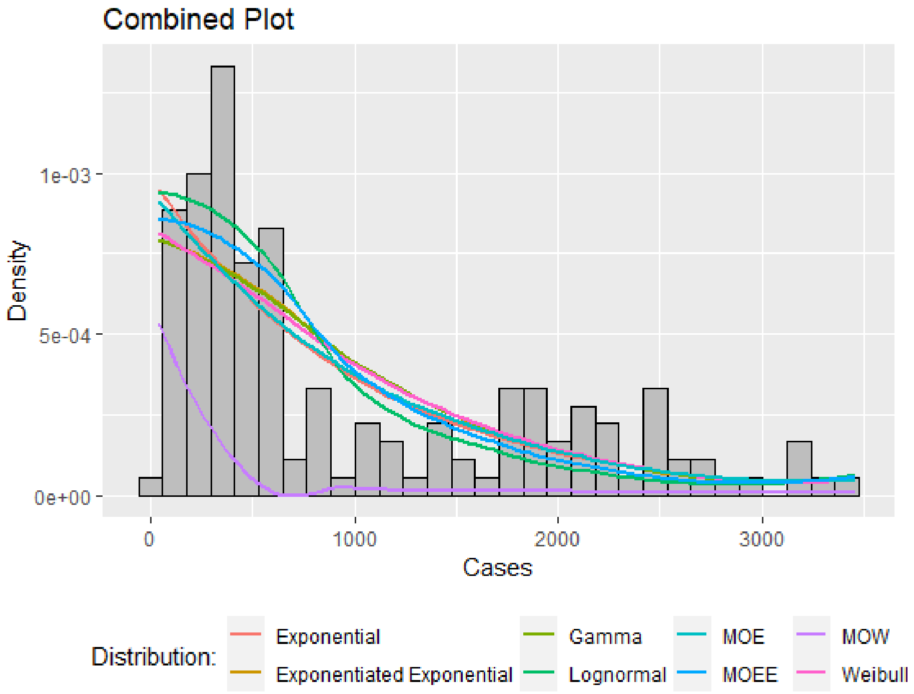
Distribution	Log Likelihood	P - Value
Exponential	-2268.21	7.75e-08
EE	-2228.60	0.01133
Gamma	-2230.07	0.00524
Weibull	-2236.04	0.00138
Lognormal	-2225.71	0.30485
MOE	-2246.45	0.00776
MOEE	**-2225.11**	**0.21682**
MOW	**-2225.62**	**0.12400**

Parameters	Exponential	EE	Gamma	Weibull	Lognormal	MOE	MOEE	MOW
1	$\lambda = 0.00066$	$\alpha = 2.43730$	$\alpha = 2.16526$	$k = 1.50291$	$\mu = 7.06010$	$\lambda = 0.00124$	$\alpha = 2.86470$	$\alpha = 2.15030$
2	NA	$\lambda = 0.00112$	$\beta = 0.00145$	$\lambda = 1665.74$	$\sigma = 0.72163$	$\theta = 3.82255$	$\lambda = 0.00085$	$\lambda = 3.039e\text{-}08$
3	NA	NA	NA	NA	NA	NA	$\theta = 0.36104$	$\theta = 0.13263$

Descriptive Statistics

Minimum	41
Maximum	3461
Median	546
Mean	1005
Standard Deviation	911.01
Coefficient of Variation	0.91



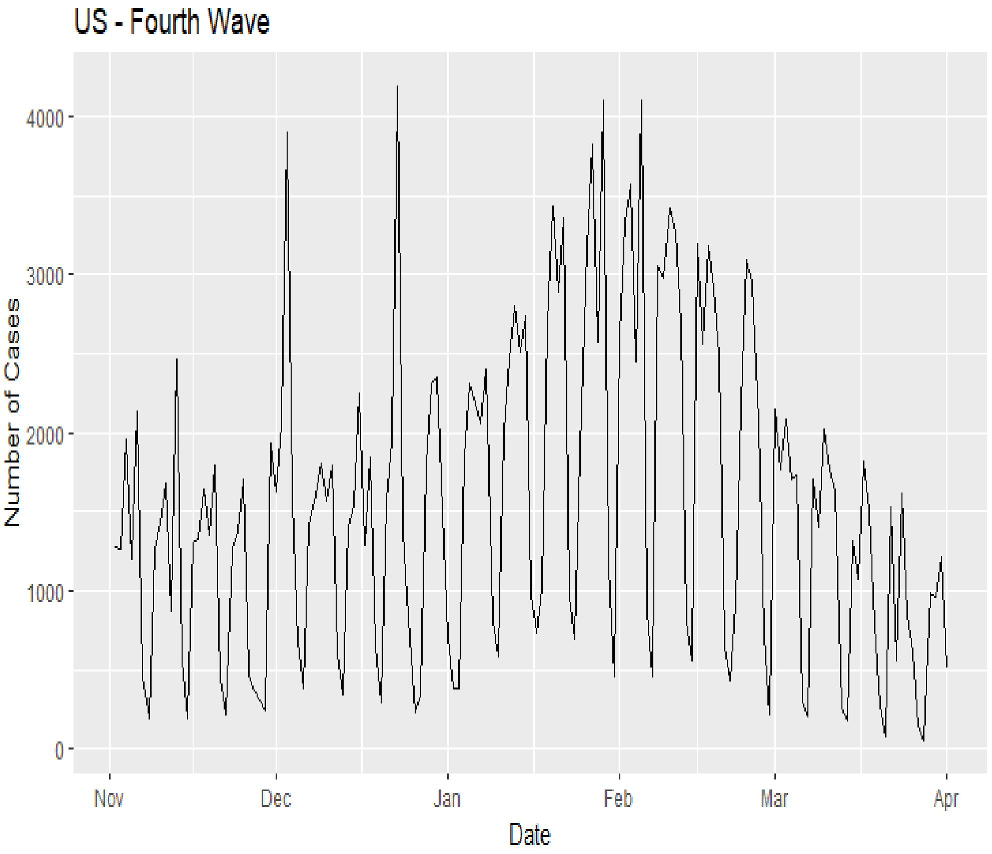


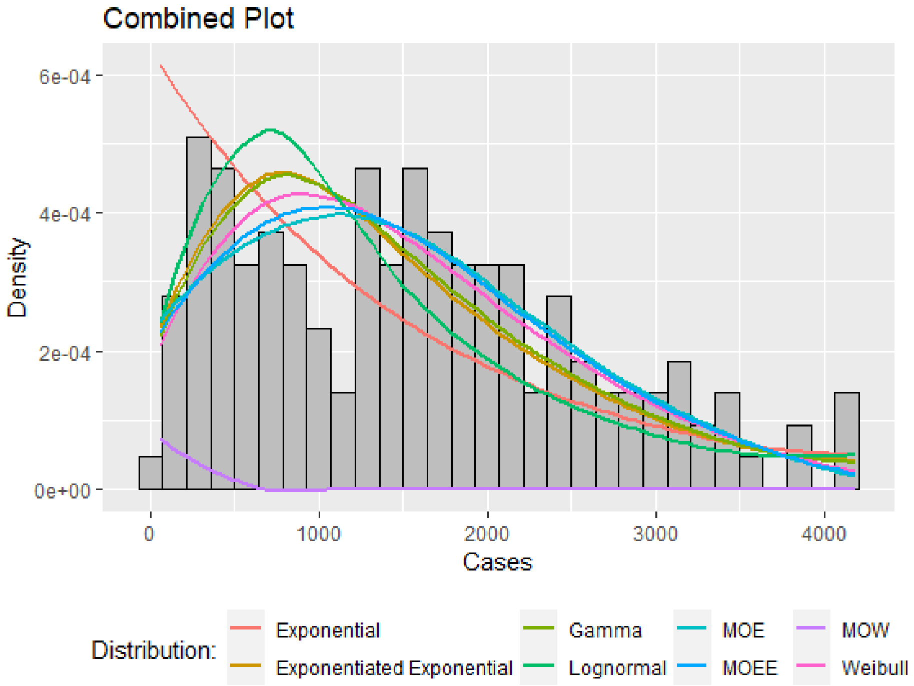
Distribution	Log Likelihood	P - Value
Exponential	-1210.58	0.0650
EE	-1209.10	0.0103
Gamma	-1209.13	0.0092
Weibull	-1209.57	0.0088
Lognormal	**-1207.99**	**0.0366**
MOE	-1210.53	0.0390
MOEE	**-1207.31**	**0.0262**
MOW	-1505.84	0

Parameters	Exponential	EE	Gamma	Weibull	Lognormal	MOE	MOEE	MOW
1	$\lambda = 0.00099$	$\alpha = 1.21308$	$\alpha = 1.19584$	$k = 1.09634$	$\mu = 6.43915$	$\lambda = 0.00104$	$\alpha = 1.49314$	$\alpha = 0.29402$
2	NA	$\lambda = 0.00112$	$\beta = 0.00119$	$\lambda = 1041.68$	$\sigma = 1.03799$	$\theta = 1.09440$	$\lambda = 0.00088$	$\lambda = 2.220e-16$
3	NA	NA	NA	NA	NA	NA	$\theta = 0.42073$	$\theta = 2.220e-16$

Descriptive Statistics

Minimum	53
Maximum	4192
Median	1500
Mean	1568
Standard Deviation	1018.65
Coefficient of Variance	0.65





Distribution	Log Likelihood	P - Value
Exponential	-1261.97	0.00075
EE	-1248.81	0.08603
Gamma	-1247.76	0.12979
Weibull	**-1245.01**	**0.48836**
Lognormal	-1260.83	0.00394
MOE	-1245.70	0.80253
MOEE	**-1245.36**	**0.75218**
MOW	-2146.38	0

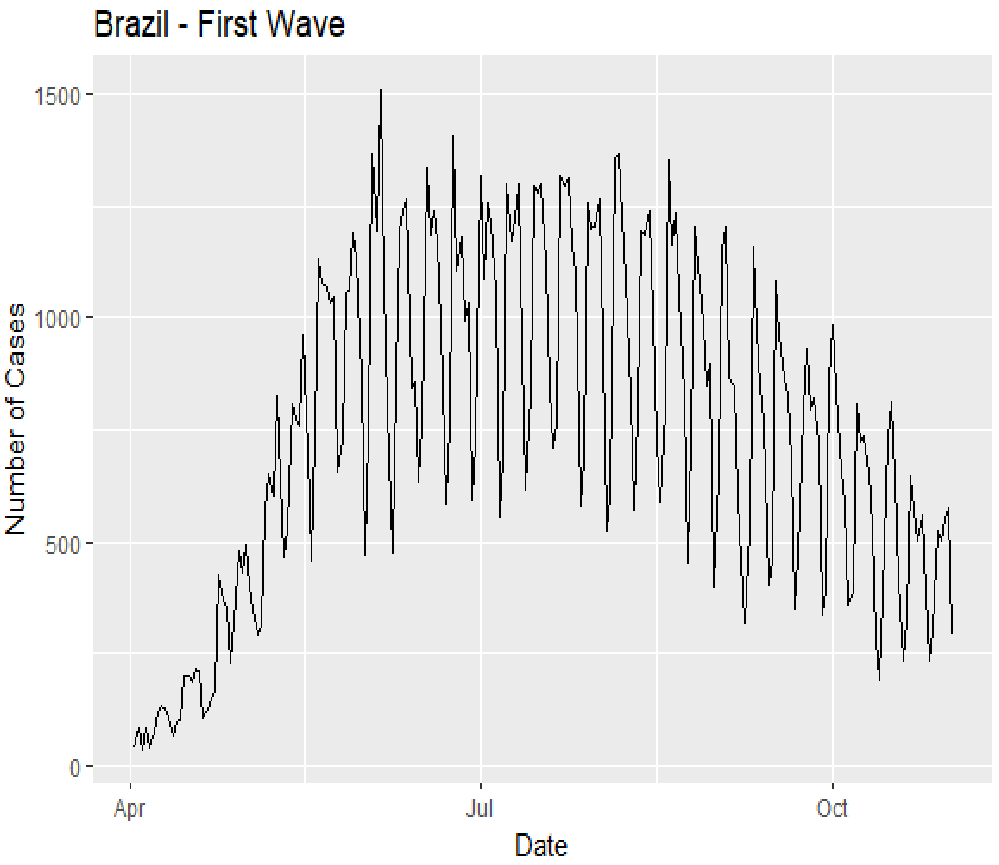
Parameters	Exponential	EE	Gamma	Weibull	Lognormal	MOE	MOEE	MOW
1	$\lambda = 0.00064$	$\alpha = 1.88065$	$\alpha = 1.84957$	$k = 1.52427$	$\mu = 7.06334$	$\lambda = 0.00129$	$\alpha = 1.28795$	$\alpha = 1.01405$
2	NA	$\lambda = 0.00092$	$\beta = 0.00118$	$\lambda = 1735.64$	$\sigma = 0.87596$	$\theta = 5.28424$	$\lambda = 0.00125$	$\lambda = 2.22e\text{-}16$
3	NA	NA	NA	NA	NA	NA	$\theta = 3.62557$	$\theta = 2.22e\text{-}16$

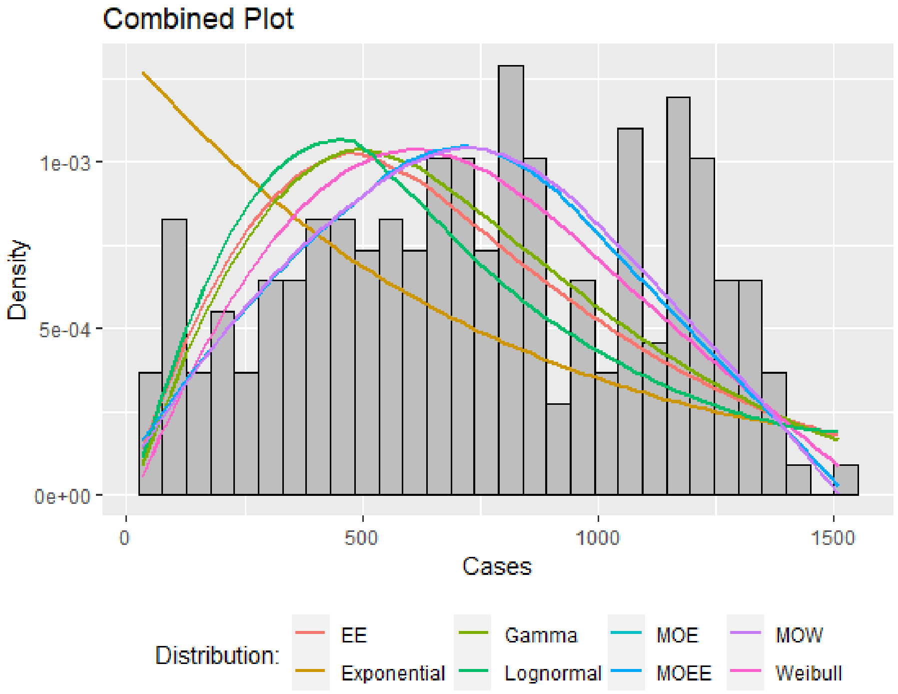
02

Brazil

Descriptive Statistics

Minimum	35
Maximum	1510
Median	750
Mean	746.6
Standard Deviation	372.37
Coefficient of Variation	0.49



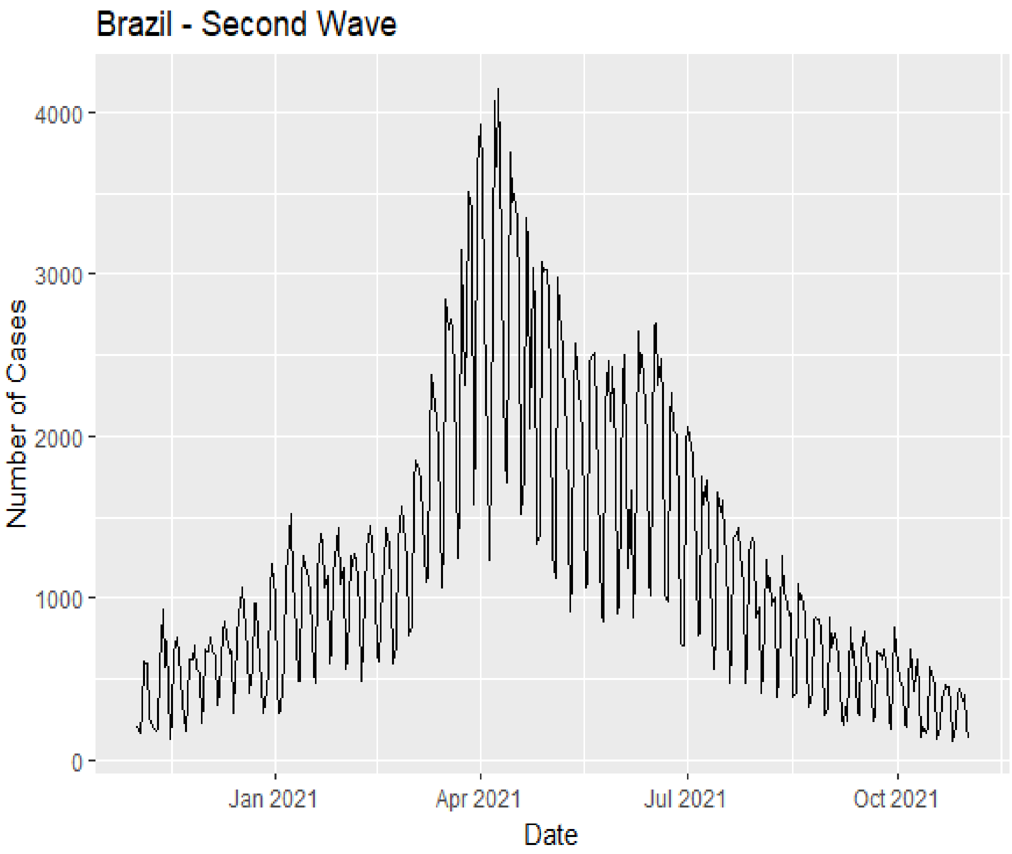


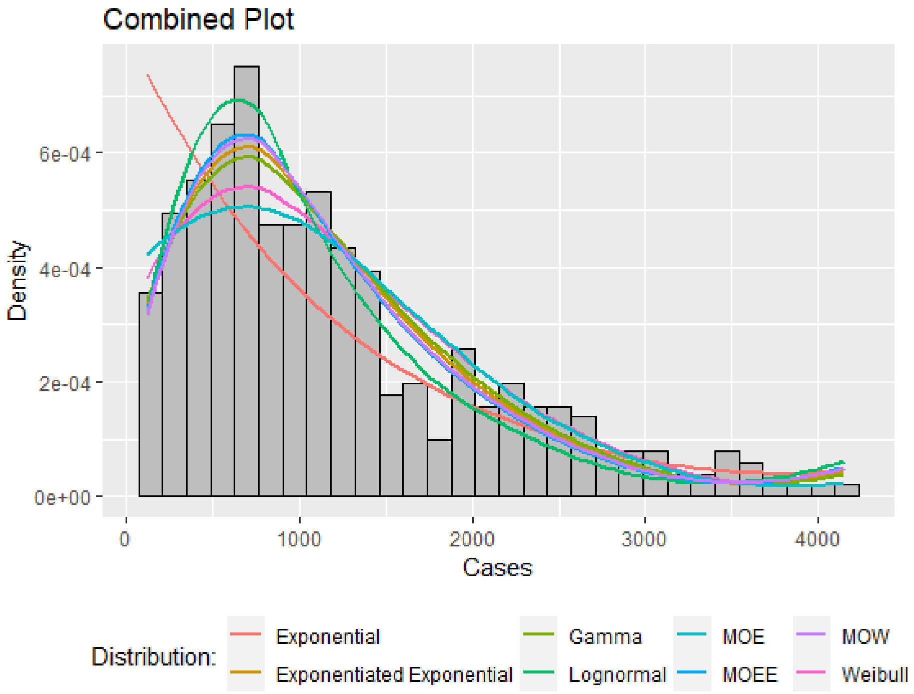
Distribution	Log Likelihood	P - Value
Exponential	-1629.72	1.21e-09
EE	-1589.68	0.0089
Gamma	-1584.98	0.0147
Weibull	-1571.19	0.1435
Lognormal	-1613.95	0.0009
MOE	**-1568.9583503**	**0.2948**
MOEE	**-1568.9583500**	**0.2947**
MOW	-1566.17	0.2553

Parameters	Exponential	EE	Gamma	Weibull	Lognormal	MOE	MOEE	MOW
1	λ = 0.00133	α = 2.695	α = 2.670	k = 2.0468	μ = 6.4167	λ = 0.00399	α = 1.0003	α = 1.0003
2	NA	λ = 0.0023	β = 0.0036	λ = 837.3	σ = 0.74540	θ = 17.894	λ = 0.004	λ = 8.6e-05
3	NA	NA	NA	NA	NA	NA	θ = 17.887	θ = 4.398

Descriptive Statistics

Minimum	119
Maximum	4148
Median	1014
Mean	1228
Standard Deviation	880.57
Coefficient of Variation	0.72



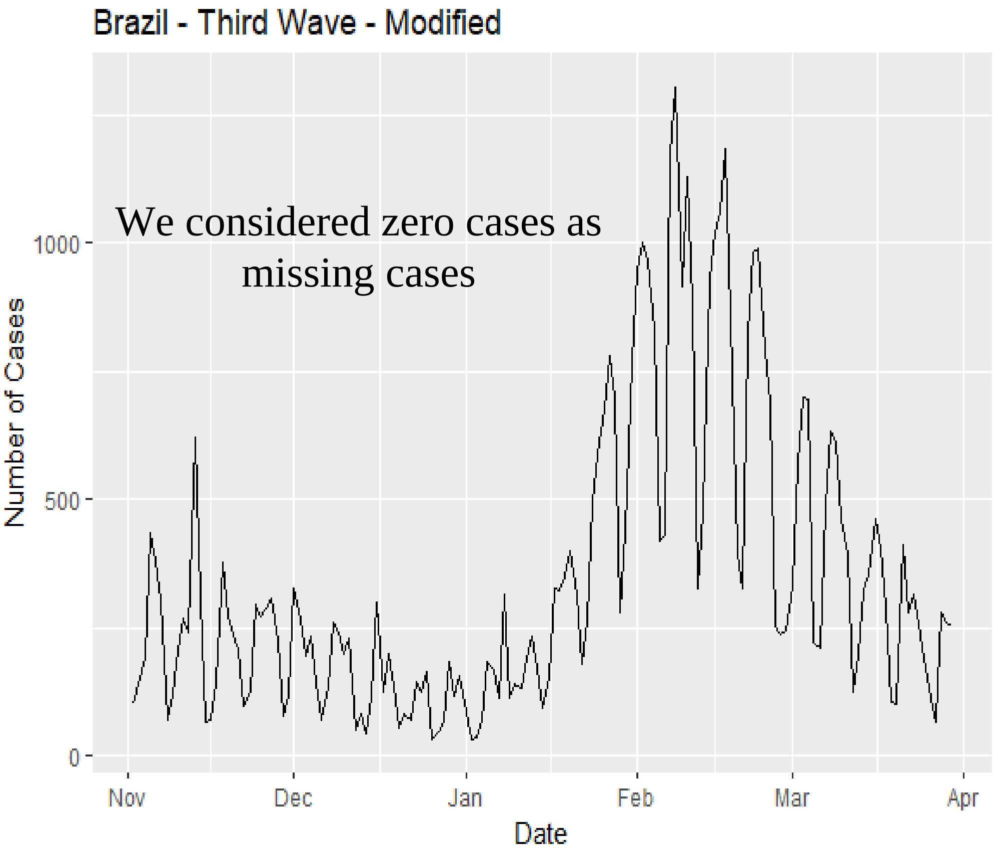


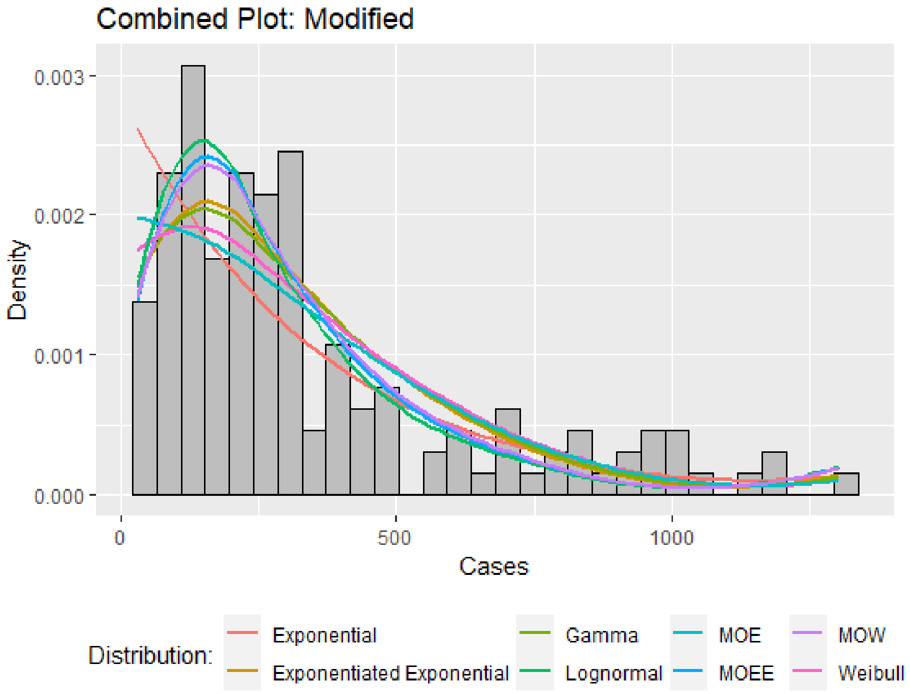
Distribution	Log Likelihood	P - Value
Exponential	-2961.21	1.24e-06
EE	-2918.92	0.4458
Gamma	-2919.35	0.3543
Weibull	-2923.53	0.3073
Lognormal	-2922.98	0.5165
MOE	-2933.50	0.1964
MOEE	**-2918.42**	**0.5101**
MOW	**-2917.77**	**0.5010**

Parameters	Exponential	EE	Gamma	Weibull	Lognormal	MOE	MOEE	MOW
1	$\lambda = 0.00081$	$\alpha = 2.1553$	$\alpha = 1.9914$	$k = 1.46658$	$\mu = 6.8413$	$\lambda = 0.00149$	$\alpha = 2.2981$	$\alpha = 1.89658$
2	NA	$\lambda = 0.0013$	$\beta = 0.0016$	$\lambda = 1362.68$	$\sigma = 0.7772$	$\theta = 3.79493$	$\lambda = 0.0012$	$\lambda = 4.6e\text{-}07$
3	NA	NA	NA	NA	NA	NA	$\theta = 0.7196$	$\theta = 0.24986$

Descriptive Statistics

Minimum	0
Maximum	1303
Median	255
Mean	344
Standard Deviation	292.47
Coefficient of Variation	0.85

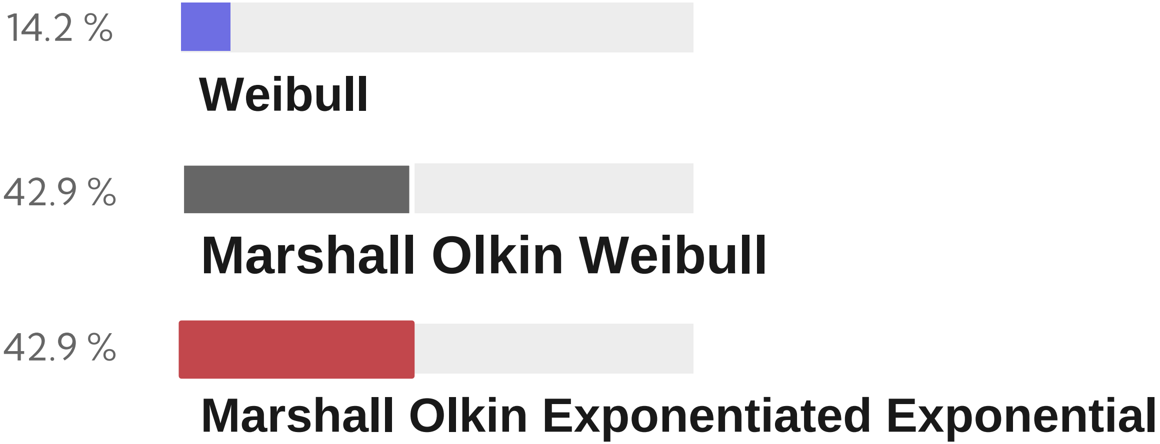




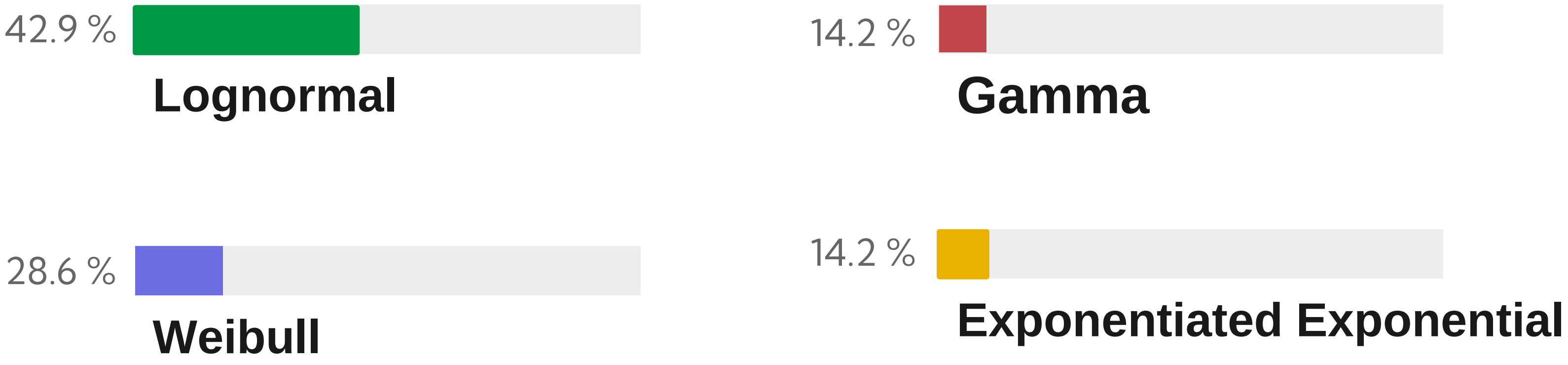
Distribution	Log Likelihood	P - Value
Exponential	-1021.23	0.0218
EE	-1011.26	0.1592
Gamma	-1011.84	0.1128
Weibull	-1014.14	0.0630
Lognormal	-1009.08	0.8958
MOE	-1017.61	0.1217
MOEE	**-1008.770**	**0.7872**
MOW	**-1008.768**	**0.7219**

Parameters	Exponential	EE	Gamma	Weibull	Lognormal	MOE	MOEE	MOW
1	$\lambda = 0.00287$	$\alpha = 1.75228$	$\alpha = 1.6339$	$k = 1.2816$	$\mu = 5.5177$	$\lambda = 0.0041$	$\alpha = 2.1643$	$\alpha = 1.83728$
2	NA	$\lambda = 0.00403$	$\beta = 0.0047$	$\lambda = 378.61$	$\sigma = 0.8485$	$\theta = 2.0859$	$\lambda = 0.0029$	$\lambda = 4.5e\text{-}06$
3	NA	NA	NA	NA	NA	NA	$\theta = 0.3191$	$\theta = 0.13030$

Does Covid-19 mortality data follow Exponential Distribution?



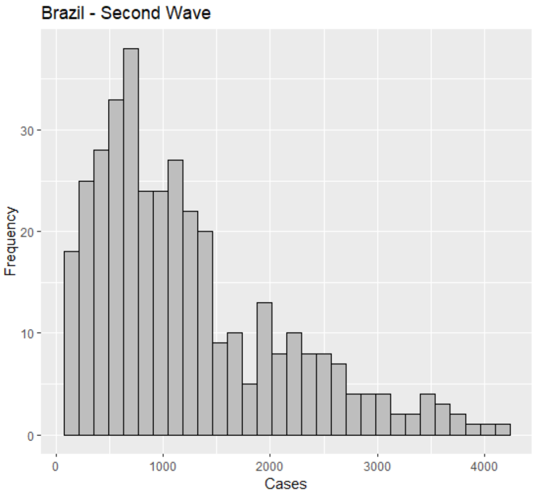
Does Covid-19 mortality data follow Exponential Distribution?

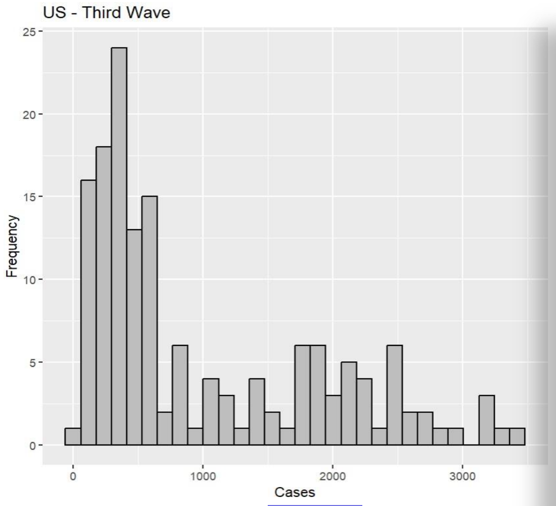


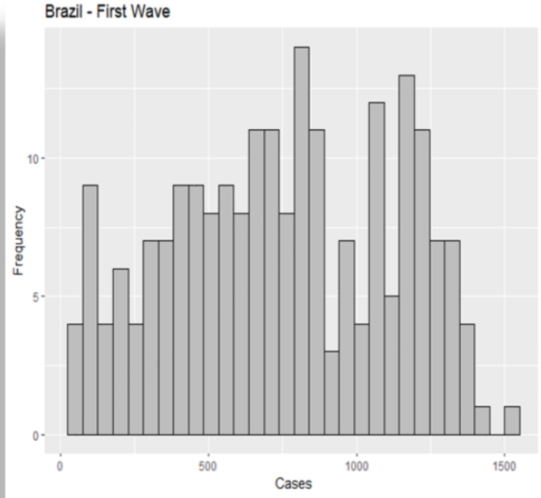
Conclusion:

Similarities:

- Flexibility: capture a wide range of shapes of density function
- Right-skewed: used to model positively skewed data (6 out of 7 waves are right skewed)







Thank You!