

Article

Gaussian and Lerch Models for Unimodal Time Series Forecasting

Azzouz Dermoune ¹, Daoud Ounaissi ²  and Yousri Slaoui ^{3,*} 

¹ CNRS, Laboratoire Paul Painlevé, UMR 8524, Université de Lille, 59653 Villeneuve d'Ascq, France; azzouz.dermoune@univ-lille.fr

² Ecole Supérieure des Agricultures, 49000 Angers, France; d.ounaissi@groupe-esa.com

³ Laboratoire de Mathématiques et Applications, Université de Poitiers, 11 Boulevard Marie et Pierre Curie, 86073 Poitiers, France

* Correspondence: yousri.slaoui@math.univ-poitiers.fr

Abstract: We consider unimodal time series forecasting. We propose Gaussian and Lerch models for this forecasting problem. The Gaussian model depends on three parameters and the Lerch model depends on four parameters. We estimate the unknown parameters by minimizing the sum of the absolute values of the residuals. We solve these minimizations with and without a weighted median and we compare both approaches. As a numerical application, we consider the daily infections of COVID-19 in China using the Gaussian and Lerch models. We derive a confident interval for the daily infections from each local minima.

Keywords: Gaussian model; Lerch model; least absolute deviation; daily infection; simplex algorithm; Nelder–Mead; optim function

1. Introduction

The least absolute deviations (LAD) method of curve-fitting proposed [1] consists of fitting the data (x_i, y_i) to a function $f(x_i, \theta)$, with $i = 1, \dots, n$. The parameter $\theta \in \mathbb{R}^p$ minimizes the sum of absolute deviations

$$\sum_{i=1}^n |y_i - f(x_i, \theta)|.$$

According to [2], in the linear regression case $f(x_i, \theta) = \sum_{j=1}^p x_{ij}\theta_j$, the minimization of the quantity

$$\sum_{i=1}^n |y_i - \sum_{j=1}^p x_{ij}\theta_j|$$

was suggested by Boscovitch (1757) (some asymptotic results are given in [3]), see also [4–6]. The latter objective function is convex with respect to the parameter θ . Hence, it has only one minima but may have many minimizers.

The linear regression case of LAD optimization is inherently more complex than the minimization of the sum of squares. The interest in the LAD method is associated with the development of robust methods. The LAD method is more resistant to the outliers in the data (see [7,8]).

The aim of our work is to analyze LAD minimization using a nonlinear regression motivated by the daily infections of COVID-19 in China during the first wave. We denote $I(t)$ the observed number of infected persons at time $t \in [1, T]$ with $T \leq 60$ (see Figure 1). The variable $t = 1, \dots, T$ represents day $1, \dots, T$.



Citation: Dermoune, A.; Ounaissi, D.; Slaoui, Y. Gaussian and Lerch Models for Unimodal Time Series Forecasting. *Entropy* **2023**, *25*, 1474. <https://doi.org/10.3390/e25101474>

Academic Editors: Geert Verdoolaege and Udo Von Toussaint

Received: 15 April 2023

Revised: 24 September 2023

Accepted: 20 October 2023

Published: 22 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

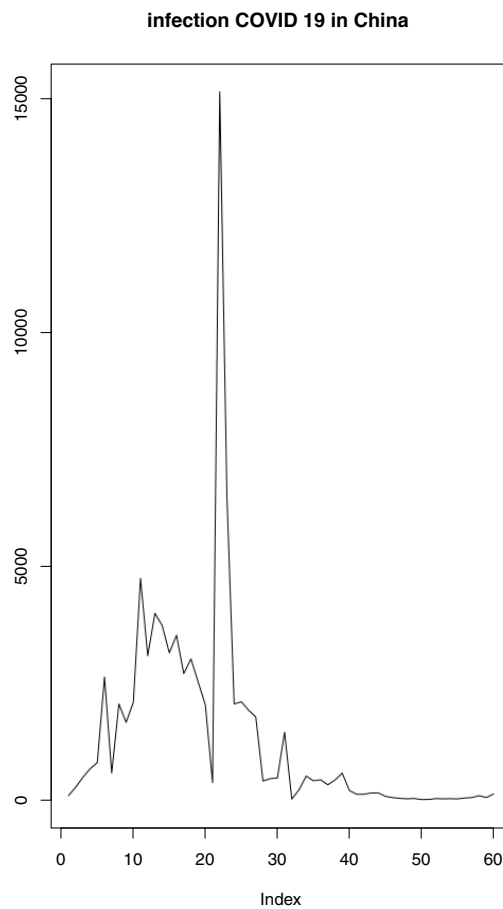


Figure 1. Daily infections of COVID-19 in China.

Justified by the sigmoidal nature of a pandemic, we propose two models: the Gaussian model (see [9])

$$I_{\text{Gauss}}(t) = a \exp\left(-\frac{(t-l)^2}{s^2}\right)$$

and the Lerch model

$$I_{\text{Lerch}}(t) = a \frac{z^t}{(v+t)^s},$$

as a prediction of the observed number of infected persons $I(t)$ at time t .

The Gaussian parameter $\theta = (a, l, s)$, with a, l, s denote, respectively, the peak, the location of the peak, and the width of the first wave of COVID-19.

The Lerch probability distribution on the non-negative integers is proportional to the function $\frac{z^t}{(v+t)^s}$ with $t = 0, 1, \dots$, the parameters $z \in (0, 1)$ and $v > 0$. The Lerch probability distribution is strongly unimodal when $s < -1$ and $v \geq 1$. In this case, its mode is at $\left[\frac{1}{z^{\frac{1}{s}-1}} - v\right] + 1$, where $[\cdot]$ signifies taking the integer part, see [10].

To estimate the three parameters $\theta = (a, l, s)$ (respectively, $\theta = (a, z, v, s)$) based on the T observations, we consider LAD nonlinear regression

$$f(T, \theta) = \frac{\sum_{t=1}^T |I(t) - I_m(t)|}{T},$$

with the subscript $m = \text{Gauss}$ (respectively, $m = \text{Lerch}$). The Gaussian model was studied in our previous work [11].

As in our previous work [11], we propose to solve our proposed LAD regression using the simplex Nelder–Mead algorithm implemented by the `optim` function in R software. The Nelder–Mead algorithm [7,12–14] is able to optimize functions without derivatives. It is a simplex method for finding a local minima of a function, the most widely used direct search method for solving optimization problems, and is considered one of the most popular derivative-free nonlinear optimization algorithms.

The output of the `optim` function depends on the initialization and is in general not a minimizer of the objective function. Restarting the Nelder–Mead algorithm from the last solution obtained (and continuing to restart it until there is no further improvement) can only improve the final solution and the latter is in general a local minimizer. Here is the iteration of the `optim` function until the convergence:

2. Probabilistic Interpretation of LAD Regression

Let us assume that

$$I(t) = I_m(t) + e(t),$$

where the errors $(e(t))$ are i.i.d. with the common probability distribution

$$\frac{1}{2\lambda} \exp\left(-\frac{|e|}{\lambda}\right),$$

where $e \in (-\infty, \infty)$ and $\lambda > 0$ are location and scale parameters, respectively (see, e.g., [15,16]). It was named after Pierre-Simon Laplace (1749–1827), as the distribution whose likelihood is maximized when the location parameter is set to the median. Based on the data $(I(1), \dots, I(T))$ the likelihood is equal to

$$\prod_{t=1}^T \frac{1}{2\lambda} \exp\left(-\frac{|I(t) - I_m(t)|}{\lambda}\right).$$

It comes that the maximum likelihood estimator of the parameters θ and λ are

$$\begin{cases} \hat{\theta} &= \arg \min\{f_m(T, \theta) : \theta\} \\ \hat{\lambda} &= f_m(T, \hat{\theta}). \end{cases}$$

In practice, $\hat{\theta}$ are given by an algorithm of optimization, and usually, they are only local minimizers. Having $\hat{\theta}$ and the scale $\hat{\lambda}$, we derive a confidence interval for $I(t)$ with $t > T$ as a solution of the equation

$$\int_{-q}^q \frac{1}{2\lambda} \exp\left(-\frac{|e|}{\lambda}\right) de = 0.95$$

given by $q = -\hat{\lambda} \ln(0.05) = 2.995732\hat{\lambda}$. We derive the confidence interval

$$IC_{0.95}[I(t)] = [\hat{I}_m(t) - 2.995732\hat{\lambda}; \hat{I}_m(t) + 2.995732\hat{\lambda}]$$

of $I(t)$ with the confidence level 0.95. Here, $\hat{I}_m(t) = \hat{a} \exp\left(-\frac{(t-\hat{l})^2}{\hat{s}^2}\right)$ in the Gaussian case, and $\hat{I}_m(t) = \hat{a}\hat{z}^t / (\hat{v} + t)^{\hat{s}}$ in the Lerch case.

3. LAD Regression Analysis Using Weighted Median

Before going forward, we recall the weighted median definition.

3.1. Weighted Median

We recall in the following proposition, the definition and the calculation of the weighted median. For more details, we advise the reader to see the work of [17].

Proposition 1. Let us consider a sequence $(x(t), w(t))$ of real numbers with positive weighted $w(t) > 0$ and $t = 1, \dots, T$. The minimizer of the function $a \rightarrow \sum_{t=1}^T w(t)|a - x(t)|$ (called the weighted median) is given as follows. We calculate the permutation $p(1), \dots, p(T)$, which rearranges the sequence $(x(t) : t = 1, \dots, T)$ into ascending order. We form the sequence $(w(p(t)) : t = 1, \dots, T)$, then we find the largest integer k which satisfies

$$\sum_{t=1}^k w(p(t)) \leq \frac{\sum_{t=1}^T w(t)}{2}.$$

If

$$\sum_{t=1}^k w(p(t)) < \frac{\sum_{t=1}^T w(t)}{2},$$

then the weighted median $a = x(p(k + 1))$.

If $\sum_{t=1}^k w(p(t)) = \frac{\sum_{t=1}^T w(t)}{2}$, then the weighted median $[x(p(k)), x(p(k + 1))]$ is equal to the interval $[x(p(k)), x(p(k + 1))]$.

3.1.1. Back to Our Proposed LAD Regression

Both our LAD regressions have the form

$$f(T, a, b) = \frac{1}{T} \sum_{t=1}^T |ag(t, b) - I(t)|,$$

with $a > 0$, and $g : (0, +\infty) \times D \rightarrow (0, +\infty)$ is a continuous positive map with D is a Euclidean domain. Now, we can announce the following corollary.

Corollary 1. For each fixed b , the minima of the function $a \rightarrow f(T, a, b)$ is attained at the weighted median $a(b)$ of the sequence $(x(t) = I(t)/g(t, b) : t = 1, \dots, T)$ endowed with the weights $(w(t) = g(t, b) : t = 1, \dots, T)$. Moreover, if (a^*, b^*) is a local minimizer of the function $(a, b) \rightarrow f(T, a, b)$, then a^* is the weighted median of $(x^*(t) = I(t)/g(t, b^*) : t = 1, \dots, T)$ endowed with the weights $(w^*(t) = g(t, b^*) : t = 1, \dots, T)$.

Proof. We observe that for fixed b , the map $a \rightarrow f(T, a, b)$ is a convex function. Now, let us assume that (a^*, b^*) is a local minimizer of the function $(a, b) \rightarrow f(T, a, b)$. Then, a^* is the global minimizer of the convex function $a \rightarrow f(T, a, b^*)$. Hence, a^* is the weighted median of $(x^*(t) = I(t)/g(t, b^*) : t = 1, \dots, T)$ endowed with the weights $(w^*(t) = g(t, b^*) : t = 1, \dots, T)$. \square

3.1.2. Comparison of the Minimizers of the Map $b \rightarrow f(T, a(b), b)$ and the Minimizers of the Map $(a, b) \rightarrow f(T, a, b)$

The following proposition is obvious.

Proposition 2. (1) For each fixed a , the map $b \rightarrow f(T, a, b)$ is above the map $b \rightarrow f(T, a(b), b)$ and they intersect at the curve $a = a(b)$.

(2) If b^* is a local minimizer of the map $b \rightarrow f(T, a(b), b)$, then $(a(b^*), b^*)$ is also a local minimizer of the map $(a, b) \rightarrow f(T, a, b)$.

(3) The local minimizers of the map $(a, b) \rightarrow f(T, a, b)$ belong to the set $\{(a(b), b) : b\}$. If $(a(b^*), b^*)$ is a local minimizer of the map $(a, b) \rightarrow f(T, a, b)$, then, in general, b^* is not a local minimizer of the map $b \rightarrow f(T, a(b), b)$. However, if $(a(b^*), b^*)$ is a global minimizer of $(a, b) \rightarrow f(T, a, b)$, then b^* is also a global minimizer of $b \rightarrow f(T, a(b), b)$.

Proof. For each fixed b , the minima of the map $a \rightarrow f(T, a, b)$ is attained at the point $a(b)$, which implies that $f(T, a, b) \geq f(T, a(b), b)$ and achieves the proof of (1). Let (a^*, b^*) be a local minimizer of the map $(a, b) \rightarrow f(T, a, b)$. It follows that $f(T, a, b) \geq f(T, a^*, b^*)$ when $a \in W, b \in V$ with W, V are some neighborhoods, respectively, of a^* and b^* . We derive that the minima of $a \in W \rightarrow f(T, a, b^*)$ is attained at b^* . As $a \in V \rightarrow f(T, a, b^*)$ is convex, it has the unique minima $b^* = a(b^*)$. Which achieves the proof of (3). The proof of (2) works as follows. There exists a neighborhood V of b^* such that

$$f(T, a(b), b) \geq f(T, a(b^*), b^*)$$

for each $b \in V$. By definition of $a(b)$, we have $f(T, a, b) \geq f(T, a(b), b)$ for all a . It follows that $(a(b^*), b^*)$ is a local minimizer of $(a, b) \rightarrow f(T, a, b)$. \square

Proposition 3. Assume that $b \rightarrow f(T, a(b), b)$ has only a global minimizer. Then, the map $(a, b) \rightarrow f(T, a, b)$ may have many local minimizers, and $b \rightarrow a(b)$ is discontinuous at any point b^* such that $(a(b^*), b^*)$ is a local minimizer of the map $(a, b) \rightarrow f(T, a, b)$.

Proof. By definition of the local minimizer, there exists a neighborhood V of $(a(b^*), b^*)$ such that $f(T, a, b) \geq f(T, a(b^*), b^*)$ for each point $(a, b) \in V$. Necessarily, $(a(b), b)$ is not in V for at least one point b near b^* , if not $f(T, a(b), b) \geq f(T, a(b^*), b^*)$ for all point b near b^* , and then b^* is a local minimizer of the map $b \rightarrow f(T, a(b), b)$. This is absurd because $b \rightarrow f(T, a(b), b)$ has only a global minimizer. \square

4. Numerical Results

In China, COVID-19 appeared on 23 December 2019 in the Wuhan region and after its fast-initial spreading, strict rules of social distancing were imposed almost one month later. Three months after the initially reported cases, the spreading in China subsided. The China data in Figure 1 were extracted from owid/COVID-19-data, available on the web.

Figure 1 shows that the peak and location, which are equal, respectively, to $a = 15,136$, and $l = 22$.

4.1. LAD Regression Using Gaussian Model with $T = 10$

By varying the initial condition and using the Algorithm 1 with the objectif function equals $(l, s) \rightarrow f_{\text{Gauss}}(T, a(T, l, s), l, s)$, we obtain only one minima: $a(10, l^*, s^*) = 2088.911, l^* = 10.11930, s^* = 5.712179$ and $f_{\text{Gauss}}(T, a(10, l^*, s^*), l^*, s^*) = 319.5446$. We recall that $a(T, l, s)$ is the weighted median of the sequence $(x(t) = I(t) \exp(-\frac{(t-l)^2}{s^2}) : t = 1, \dots, T)$ with the weights $(w(t) = \exp(-\frac{(t-l)^2}{s^2}) : t = 1, \dots, T)$.

However, by varying the initial condition and using the Algorithm 1 with the objectif function equals $(a, l, s) \rightarrow f_{\text{Gauss}}(T, a, l, s)$, we found several lists of minima. Table 1 shows some of them. The minima 331.487 corresponds to the minimizer $a^* = 5526.386, l^* = 22.018, s^* = 12.182$. We recall that the observed location is $l = 22$. The minima 337.0095 corresponds to the minimizer $a^* = 15,185.797, l^* = 31.994, s = 15.614$. We recall that the observed peak is $a = 15,136$.

Algorithm 1 The output of the iteration of optim function until convergence

Input: Fonction F

1: initialization θ_0 ,

2: **while** $F(\theta_0) \neq F(T, \text{optim}(\theta_0, F, \text{method} = \text{"Nelder - Mead"}))$ **do**

3: $\theta(\theta_0) = \text{optim}(\theta_0, F, \text{method} = \text{"Nelder - Mead"})$ // optim function applied to F with the initialization $\theta(\theta_0)$.

4: **end while**

output: $\theta(\theta_0)$ and $F(\theta(\theta_0))$.

Table 1. A list of minimizers of $(a, l, s) \rightarrow f_{\text{Gauss}}(T, a, l, s)$ with $T = 10$.

a^*	l^*	s^*	minima	% Predictions
2088.911	10.119	5.712	319.545	0.66
2137.277	10.98621	6.457	320.9898	0.66
2301.818	12.186	7	322.358	0.68
3229.843	16.346	9.609	325.995	0.88
5526.387	22.018	12.183	331.487	0.60
15,185.797	31.994	15.614	337.009	0.12

4.2. Confidence Intervals Using Six minimas with $T = 10$ in the Gaussian Model

We recall that the confidence interval of $I(t)$ from the minimizer $(\hat{a}, \hat{l}, \hat{s})$ is given by

$$IC_{0.95}[I(t)] = \left[\hat{a} \exp\left(-\frac{(t - \hat{l})^2}{(\hat{s})^2}\right) - 2.995732\hat{\lambda}; \hat{a} \exp\left(-\frac{(t - \hat{l})^2}{(\hat{s})^2}\right) + 2.995732\hat{\lambda} \right] \text{ with } t > 10.$$

In Figure 2, we present the confidence intervals for the global minima and five local minimas among the list $T = 10$. The predictions using the minimas 335.047 and 336.92 are clearly bad. However, their predictions at the location $l = 22$ are close to the real peak among the six minimas. The percentage of predictions is given in Table 1. An R source code is given in Appendix A, which can be used to determine the confidence intervals for the other values of T once the list of minimas is determined by using one of the three considered methods.

In Figure 2, we present the confidence intervals for the global minima and six local minimas among the list $T = 10$ in the Gaussian model.

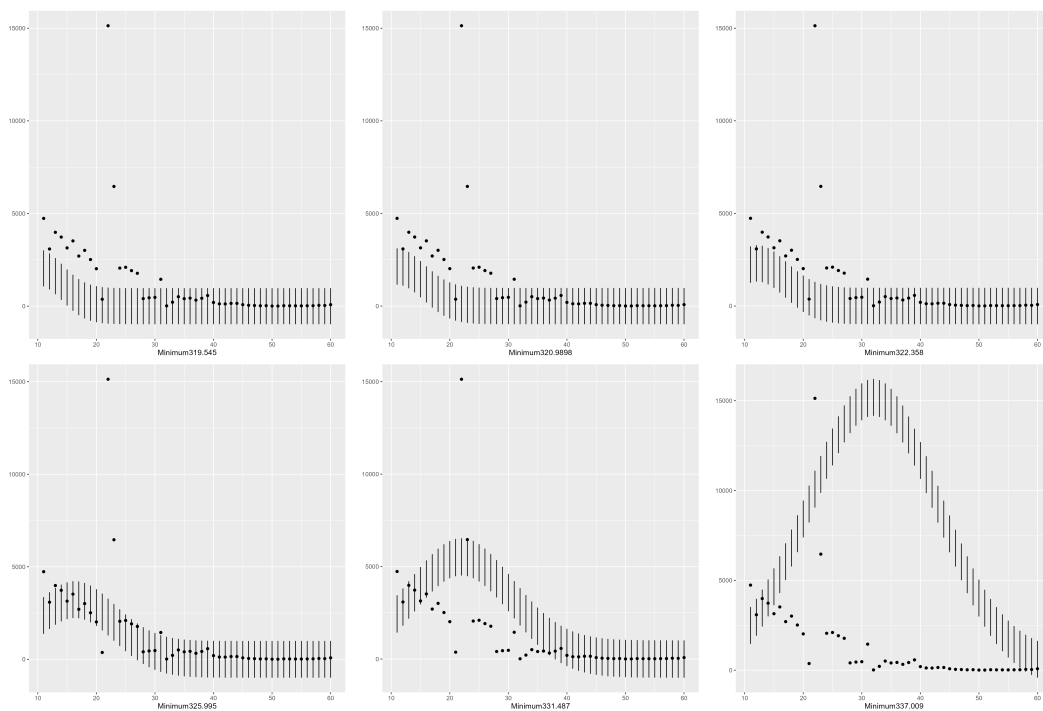


Figure 2. Confidence intervals using six minimas with $T = 10$ in the Gaussian model.

Remark 1. Let us conclude this paragraph with a comparison between the output of the `optim` function and the output of the iteration of the `optim` function until the convergence using the same initialization with $T = 10$. As initialization, we use $(a = 25, 325.01, l = 41.78141, s = 19.85630)$ for the two approaches; the use of the `optim` function leads to the following minima $(23, 034.33, 36.03957, 16.80575, 338.2104)$, while the iteration of the `optim` function until

convergence leads to (2137.277, 10.98621, 6.457341, 320.9898). Figure 3 illustrates the plots of daily infections $t \in [1, 60] \rightarrow I(t)$ of COVID-19 in China, the LAD regression $t \in [1, 60] \rightarrow I_{without}^*(t)$ using the minima 338.2104 of the *opt im* function and the LAD regression $t \in [1, 60] \rightarrow I_{iteration}^*(t)$ using the minima 320.9898 of the iteration of the *opt im* function until the convergence.

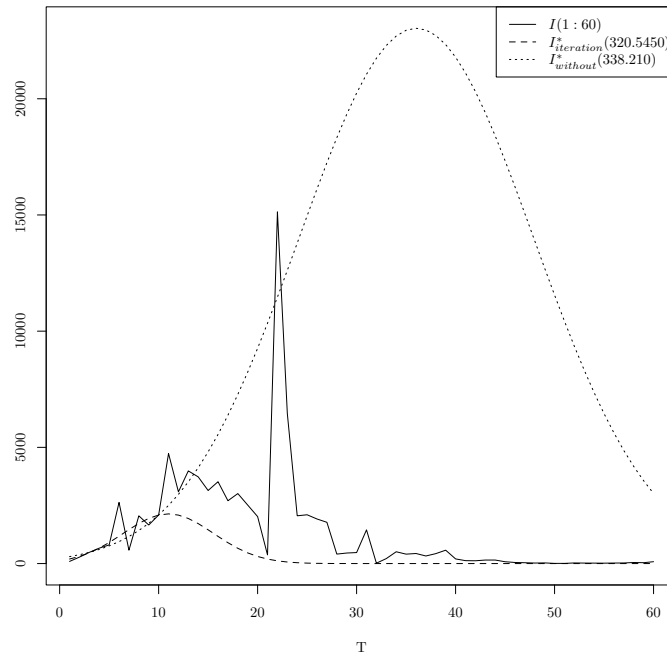


Figure 3. Optimal time series for $T = 10$, for the Gaussian model with and without iterations.

4.3. LAD Regression Using the Lerch Model with $T = 10$

By varying the initial condition and using the Algorithm 1 with the function $(z, v, s) \rightarrow f_{Lerch}(T, a(T, z, v, s), z, v, s)$ we found a huge number of minimas. Table 2 shows some of them. We recall that $a(T, z, v, s)$ is the weighted median of the sequence $(x(t) = I(t)(v + t)^s / z^t : t = 1, \dots, T)$ with the weights $(w(t) = z^t / (v + t)^s : t = 1, \dots, T)$. We also recall that in the Gaussian case, the surface $(l, s) \rightarrow f_{Gauss}(T, a(T, l, s), l, s)$ has only one minima equal to the global minima of the map $(a, l, s) \rightarrow f_{Gauss}(T, a, l, s)$.

Figures 4 and 5 show, respectively, the optimal time series for the Lerch and Gauss models for $T = 10$, using our list of local minimas. We can observe that the Lerch model fits better with the prediction of $I(t)$ for $t > 10$. But the Gauss model fits better with the prediction of the location and the size of the peak.

Table 2. A list of minimizers of $(a, z, v, s) \rightarrow f_{Lerch}(T, a, z, v, s)$ with $T = 10$.

a^*	z^*	v^*	s^*	Minima	% Predictions
5.990886×10^{-01}	0.81023163	3.201870	-3.976447	311.9606	0.84
1.354215×10^{-08}	0.59318684	8.174956	-10.683932	314.7595	0.70
1.555525×10^{-14}	0.50711308	10.701817	-15.255833	315.4869	0.68
8.235320×10^{-34}	0.35420137	16.512305	-28.741171	316.5241	0.66
$2.185815 \times 10^{-103}$	0.1647315	28.956740	-71.553876	317.5824	0.66
7.23785×10^{-266}	0.05804243	45.948388	-160.673224	318.2142	0.66

In Table 3, we report the columns of mode positions and the corresponding amplitudes for each minimizer.

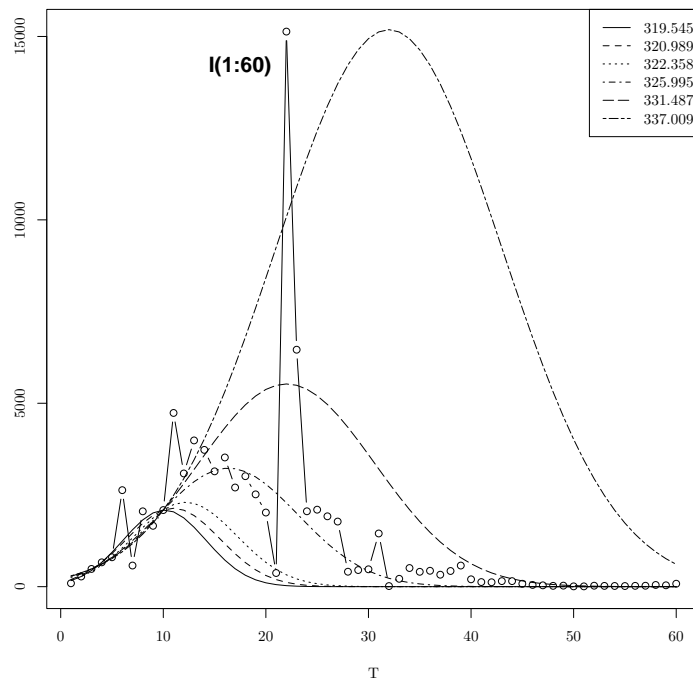


Figure 4. Optimal time series for $T = 10$ for Gauss.

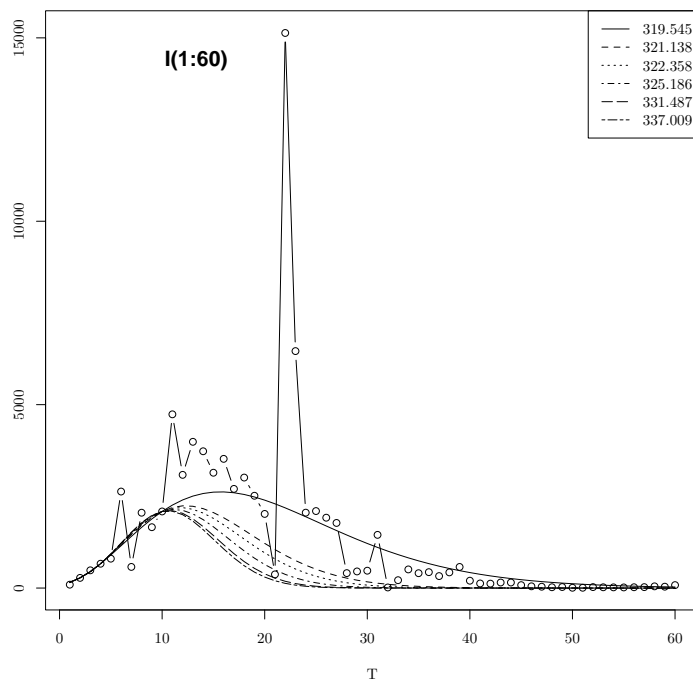


Figure 5. Optimal time series for $T = 10$ for Lerch.

Table 3. Mode position and amplitude using Lerch.

a^*	z^*	v^*	s^*	Mode Position	Amplitude
5.990886×10^{-01}	0.81023163	3.201870	-3.976447	16.19885	2618.210
1.354215×10^{-08}	0.59318684	8.174956	-10.683932	12.78678	2236.537
1.555525×10^{-14}	0.50711308	10.701817	-15.255833	12.26928	2186.184
8.235320×10^{-34}	0.35420137	16.512305	-28.741171	11.68264	2134.780
$2.185815 \times 10^{-103}$	0.1647315	28.956740	-71.553876	11.22172	2101.053
7.23785×10^{-266}	0.05804243	45.948388	-160.673224	11.05461	2090.932

4.4. Confidence Intervals Using Six Minimas with $T = 10$ in the Lerch Model

The percentage of predictions is given in Table 2. Observe that the best percentage of predictions occurs at the global minima of the Lerch model, but the best percentage of predictions occurs at the local minimizer 325.186 of the Gauss model.

In Figure 6, we present the confidence intervals for the global minima and six local minimas among the list $T = 10$ of the Lerch model.

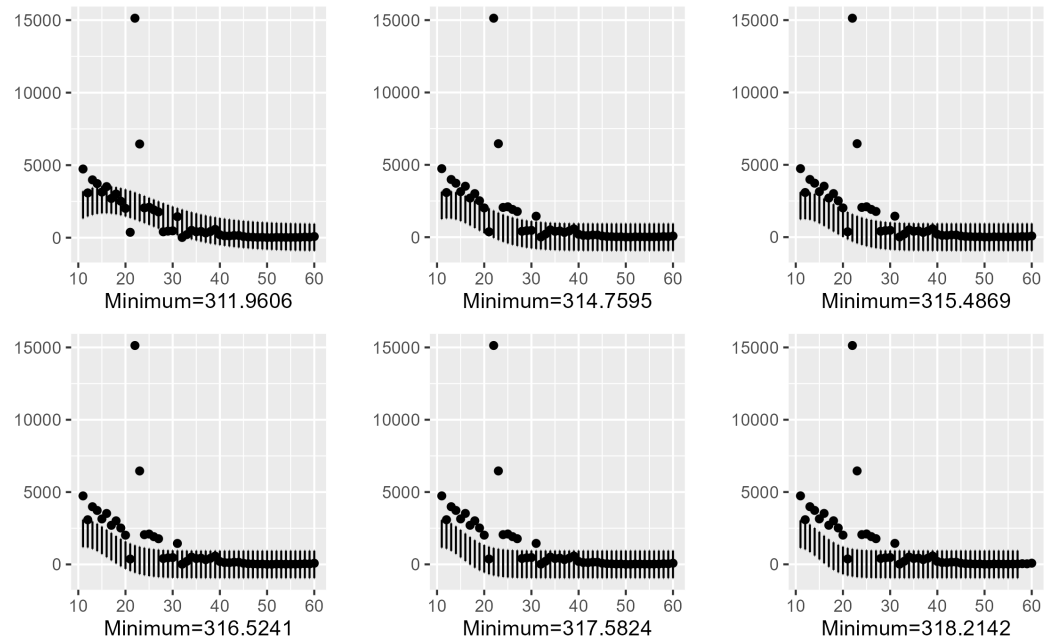


Figure 6. Confidence intervals using six minimas with $T = 10$ of the Lerch model.

5. The Case $T > 10$

In this section, we consider the cases $T = 20$ and $T = 60$. In the Lerch model case, LAD regression still has many minimizers for $T = 20$. See Table 4. But in the Gauss model, LAD regression has only one minimizer for $T = 20$. See Table 5.

Table 4. A list of minimizers of $(a, z, v, s) \rightarrow f_{Lerch}(T, a, z, v, s)$ with $T = 20$. The last column contains the percentage of predictions.

a^*	z^*	l^*	s^*	Minima	% Predictions
$4.366541 \times 10^{-268}$	0.1073967	55.31169	-154.51447	392.5692	0.825
$3.422419 \times 10^{-262}$	0.1387492	47.81770	-121.91023	393.5286	0.825
$2.572282 \times 10^{-256}$	0.1708497	41.39999	-97.69044	394.7033	0.825
$1.155039 \times 10^{-253}$	0.1860852	38.76658	-88.52397	395.2740	0.825
$3.035334 \times 10^{-234}$	0.3069971	23.34634	-43.86187	400.4082	0.825
$4.970115 \times 10^{-230}$	0.3344249	20.71433	-37.77693	401.7952	0.825
$1.667726 \times 10^{-233}$	0.3634773	22.30914	-36.60867	403.6049	0.825

Table 5. The unique minimizer of $(a, l, s) \rightarrow f_{Gauss}(T, a, l, s)$ with $T = 20$.

a^*	l^*	s^*	Minima	% Predictions
3712.1297	14.2556	7.60744	386.0318	0.80

From Tables 4 and 5, we can observe that the percentage of prediction of the Lerch model is better than the percentage of prediction of the Gauss model.

Finally, the minima for $T = 60$ for both models (Gauss and Lerch) are unique and respective to $(3318.433, 16.94084, 10.19735, 617.2386)$ and $(0.05691053, 0.7445283, 3.3413322, -5.2886448, 570.7861098)$.

Figure 7 shows the optimal time series for the Lerch and Gauss models for $T = 60$, using our local minima.

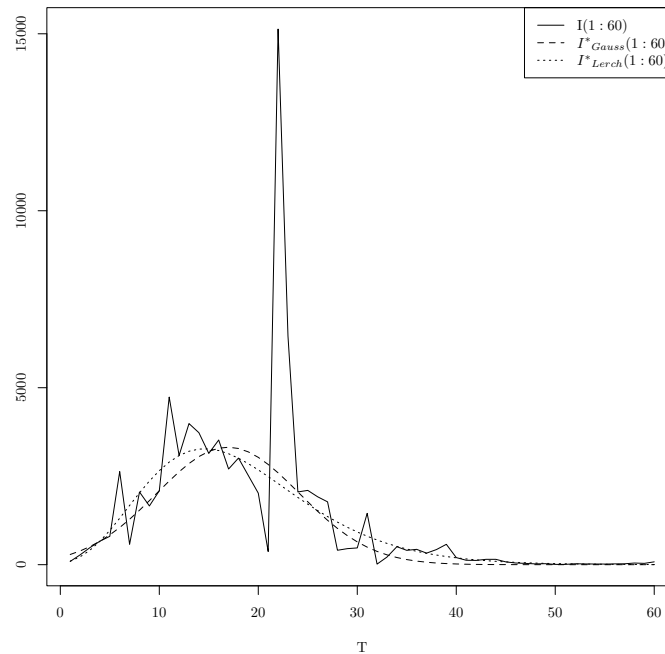


Figure 7. Optimal time series for $T = 60$ for Lerch and Gauss.

We observe that the Lerch approximation has a heavier tail than the Gaussian approximation.

6. Conclusions

In this work, we considered LAD nonlinear regression

$$f(T, \theta) = \frac{\sum_{t=1}^T |I(t) - I_m(t)|}{T},$$

with the subscript $m = Gauss$ (respectively, $m = Lerch$).

The two models have the form $(a, b) \rightarrow f(T, a, b) = \sum_{t=1}^T |ag(t, b) - I(t)|$. We have associated $f(T, \cdot)$ with the map $S(T, \cdot) : b \rightarrow f(T, med(x(b), w(b)), b)$ with the sequence $x(b) = (I(t)/f(t, b), t = 1, \dots, T)$, and the weights $w(b) = (g(t, b), t = 1, \dots, T)$, and $med(x(b), w(b))$ denotes the weighted median of $x(b)$ endowed with the weights $w(b)$. We showed that if b^* is a local minimizer of S , then $(med(x(b^*), w(b^*)), b^*)$ is also a local minimizer of $f(T, \cdot)$. The converse is, in general, false, i.e., if (a^*, b^*) is a local minimizer of $f(T, \cdot)$, then b^* is not, in general, a local minimizer of S . However, if (a^*, b^*) is the global minima of $f(T, \cdot)$, then b^* is also the global minima of S . We showed that if S has only a global minima, then $b \rightarrow f(T, med(x(b), w(b)), b)$ is discontinuous at each local minimizer. Using the data of the daily infections of COVID-19 in China during the first wave, we showed numerically in the Gaussian case $g(t, l, s) = \exp(-\frac{(t-l)^2}{s^2})$ that the map $f(T, a, l, s)$ has a huge number of local minimas, but the surface S has only a global minima, which is also the global minima of the map $f(T, \cdot)$. However, in the Lerch case $g(t, z, v, s) = \frac{z^t}{(t+v)^s}$, contrary to the Gaussian case, we showed that the maps $(a, z, v, s) \rightarrow f(T, a, z, v, s)$ and $(z, v, s) \rightarrow S(T, z, v, s)$ have each a huge number of local minimas. We derived confident intervals for the daily infections from each local minima. Our message is that each local minima contains a part of the information and can be used for the prediction of a part of the parameters.

In future work, we will deal with the log-transformation $\ln(I_{Gauss}(t)) = \beta_0 + \beta_1 t - \beta_2 t^2$, where $(\beta_0, \beta_1, \beta_2)$ are in bijection with the parameters a, l, s of the Gaussian model. This allows us to consider the linear model $\ln(I(t)) = \beta_0 + \beta_1 t - \beta_2 t^2 + \varepsilon(t)$, with the errors $(\varepsilon(t))$, which are i.i.d. We will deal with this problem using other robust regression estimators as the least trimmed median estimator (LTM) proposed in [18].

Author Contributions: Conceptualization, A.D. and Y.S.; methodology, A.D., D.O. and Y.S.; software, D.O. and Y.S.; validation, A.D., D.O. and Y.S.; formal analysis, D.O.; investigation, A.D. and Y.S.; resources, A.D.; data curation, A.D.; writing—original draft preparation, A.D., D.O. and Y.S.; writing—review and editing, Y.S.; visualization, A.D. and Y.S.; supervision, A.D. and Y.S.; project administration, A.D., D.O. and Y.S.; funding acquisition, Y.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data are unavailable due to privacy.

Acknowledgments: We would like to thank the Editor-in-Chief of Entropy, the associate Editor, and two anonymous reviewers for their helpful comments, which helped us to focus on improving the original version of the paper. The authors would like to thank the Editor, the Associate Editor, and the three referees for their very helpful comments, which led to a considerable improvement of the original version of the paper and a more sharply focused presentation.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. R Source Code

Here, we provide a source code for a better understanding of the proposed method.

Appendix A.1. *opt()* Function

```
library(optim)
OPT = function(T, init) {
  y =function(x) f(T, x)
  repeat {
    optim.result=optim(init, y)$par
    init= optim.result$par
    min.opt=optim.result$value
    if (f(T,init) == f(T, optim(init, y)$par)) {
      break
    } else{init=init}
  }
  return(c(init, min.opt))
}
```

Appendix A.2. Confidence Intervals Plot

```
library(ggplot2)
library(openxlsx)
conf.int.min=function(T,Min.list,T.inf,T.sup){
  # given T
  # Min.list is the list of minima obtained for T
  ## T.inf = T+1 and T.sup=upper limit of days
  j=1:(60-T)
  Tj=T+j

  born_sup=NULL
  born_inf=NULL
  confid.int=NULL
```

```

int_T=NULL
for(i0 in 1:nrow(Min.list)){
  for(i00 in 1:length(Tj)){
    born_sup=Min.list[i0,1]*exp(-(Tj[i00]-Min.list[i0,2])^2/(Min.list[i0,3]^2))
      +2.995732*Min.list[i0,4]
    born_inf=Min.list[i0,1]*exp(-(Tj[i00]-Min.list[i0,2])^2/(Min.list[i0,3]^2))
      -2.995732*Min.list[i0,4]
    int_T=rbind(int_T,c(born_inf,i[Tj[i00]],born_sup))
  }
  colnames(int_T)=c("lower","i","upper")
  confid.int=cbind(confid.int,int_T)
  int_T=NULL
}
confid.int=data.frame(confid.int)
row.names(confid.int)=as.character(Tj)
#export all confidence intervals
write.xlsx(confid.int,file=paste0("confid_int_",T,".xlsx"))
#creation of a matrix which is equal to the matrix of
#confidence intervals for use with ggplot below
ggplot.confid=confid.int
iter.plot=1
while(ncol(ggplot.confid)>=3){
  ggsave(ggplot(ggplot.confid[,1:3], aes(x=T.inf:T.sup, y=ggplot.confid[,2])) +
  xlab(paste0(paste0("minima",sep=Min.list[iter.plot,4]))) + ylab("") +
  geom_errorbar(aes(ymin=ggplot.confid[,1], ymax=ggplot.confid[,3]), width=.1) +
  geom_point(),
  file=paste0("int_conf_min_",T,"_", iter.plot,".png"))
  iter.plot=iter.plot+1
  ggplot.confid=ggplot.confid[,-c(1:3)]
}
}

```

References

- Bloomfield, P.; Steiger, W.L. Least Absolute Deviations Curve-Fitting. *SIAM J. Sci. Stat. Comput.* **1980**, *2*, 290–301. [\[CrossRef\]](#)
- Eisenhart, C. Boscovitch and the combination of observations. In *Roger Joseph Boscovitch*; Whyte, L.L., Ed.; Fordham University Press: New York, NY, USA, 1961.
- Koenker, R.; Bassett, G. On Boscovitch's estimator. *Ann. Stat.* **1985**, *13*, 1625–1628. [\[CrossRef\]](#)
- Rhodes, E.C. Reducing observations by the method of minima deviations. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1930**, *9*, 974–992. [\[CrossRef\]](#)
- Karst, O.J. Linear curve fitting using least deviations. *J. Am. Stat. Assoc.* **1958**, *53*, 118–132. [\[CrossRef\]](#)
- Sadovskii, A.N. Algorithm As 74. L1-norm fit of straight line. *J. R. Stat. Soc. Ser. C Appl. Stat.* **1974**, *23*, 244–248. [\[CrossRef\]](#)
- Dielman, T.E. Least absolute value regression: recent contributions. *J. Stat. Comput. Simul.* **2005**, *75*, 263–286. [\[CrossRef\]](#)
- Li, Y.; Arce, G.R. A maximum likelihood approach to least absolute deviation regression. *EURASIP J. Appl. Signal Process.* **2004**, *12*, 1762–1769. [\[CrossRef\]](#)
- Barmparis, G.D.; Tsironis, G.P. Estimating the infection horizon of COVID-19 in eight countries with a data-driven approach. *Chaos Solitons Fractals* **2020**, *135*, 109842. [\[CrossRef\]](#) [\[PubMed\]](#)
- Aksenov, S.V.; Savageau, M.A. Some properties of the Lerch family of discrete distributions. *arXiv* **2005**, arXiv:math/0504485.
- Dermoune, A.; Ounaissi, D.; Slaoui, Y. Confidence intervals from local minimas of objective function. *Stat. Optim. Inf. Comput.* **2023**, *11*, 798–810. [\[CrossRef\]](#)
- Gao, F.; Lixing, H. Implementing the Nelder-Mead simplex algorithm with adaptive parameters. *Comput. Optim. Appl.* **2010**, *51*, 259–277. [\[CrossRef\]](#)
- Lagarias, J.C.; Reeds, J.A.; Wright, M.W.; Wright, P.E. Convergence properties of the Nelder–Mead simplex method in low dimensions. *SIAM J. Optim.* **1998**, *9*, 112–147. [\[CrossRef\]](#)
- Nelder, J.A.; Mead, R. A simplex method for function minimization. *Comput. J.* **1965**, *7*, 308–313. [\[CrossRef\]](#)

15. Johnson, N.L.; Kotz, S.; Balakrishnan, N. *Continuous Univariate Distributions*; Wiley: Hoboken, NJ, USA, 1994.
16. Ord, J.K. *Laplace Distribution*, in *Encyclopedia of Statistical Science*; Kotz, S., Johnson, N.L., Read, C.B., Eds.; Wiley: New York, NY, USA, 1983; Volume 4, p. 473475.
17. Novoselac, V. The Component Weighted Median Absolute Deviations Problem. *Eur. J. Pure Appl. Math.* **2020**, *13*, 964–976. [[CrossRef](#)]
18. Croux, C.; Rousseeuw, P.J.; Van Bael, A. Positive-breakdown regression by minimizing nested scale estimators. *J. Stat. Plan. Inference* **1996**, *53*, 197–235. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.