Journal of Data Science, Statistics, and Visualisation

May 2025, Volume V, Issue IV. *doi: 10.52933/jdssv.v5i4.126*

Supplementary Materials for "Robust Maximum L_q-Likelihood Covariance Estimation for Replicated Spatial Data"

Sihan Chen
King Abdullah University
of Science and Technology

Joydeep Chowdhury
King Abdullah University
of Science and Technology

Marc G. Genton
King Abdullah University
of Science and Technology

Abstract

Parameter estimation with the maximum L_q-likelihood estimator (MLqE) is an alternative to the maximum likelihood estimator (MLE) that considers the q-th power of the likelihood values for some $0 < q < 1$. In this method, extreme values are down-weighted because of their lower likelihood values, which yields robust estimates. In this work, we study the properties of the MLqE for spatial data with replicates. We investigate the asymptotic properties of the MLqE for Gaussian random fields with a Matérn covariance function, and carry out simulation studies to investigate the numerical performance of the MLqE. We show that it can provide more robust and stable estimation results when some of the replicates in the spatial data contain outliers. In addition, we develop a mechanism to find the optimal choice of the hyper-parameter q for the MLqE. The robustness of our approach is further verified on a United States precipitation dataset. Compared with other robust methods for spatial data, our proposal is more intuitive and easier to understand, yet it performs well when dealing with datasets containing outliers.

Keywords: Gaussian random fields, large-scale computation, maximum L_q-likelihood estimator, robust statistics, spatial statistics.

S1. Sensitivity Curves

In Section 2.4 of the main article, we derived the influence function of the MLqE. When $q = 1$, which is the case of the MLE, the influence function is unbounded, while it is bounded when $q < 1$. Here in this section, we present some sensitivity curves using simulated datasets, in order to further verify the boundedness of the influence function of the MLqE.

In Figures S1 and S2, we show the sensitivity curves of the parameters σ^2, β and ν. We first generate a spatial dataset using Matérn kernel with true parameters $\sigma^2 = 1, \beta = 0.1, \nu = 0.5$, with 50 replicates, then pick one of the 50 replicates and multiply it by a value c. Afterwards, we estimated the parameters σ^2, β and ν using the MLqE with several different choices of q, and see how the estimation results with different q would vary with c. This process is repeated 20 times, and the median of the estimation of each of the three parameters for each c and q is plotted in Figures S1 and S2. We can see from the curves that, when c becomes larger, which means that the distortion of the data is more significant, the MLE results of all the three parameters are deviating from the true values and we cannot observe bounds of the curves representing the MLE. However, all MLqE curves seem to be bounded, which verifies our finding in Lemma 2 of the main article, that the influence function for $q < 1$ is bounded, while it is unbounded for $q = 1$.

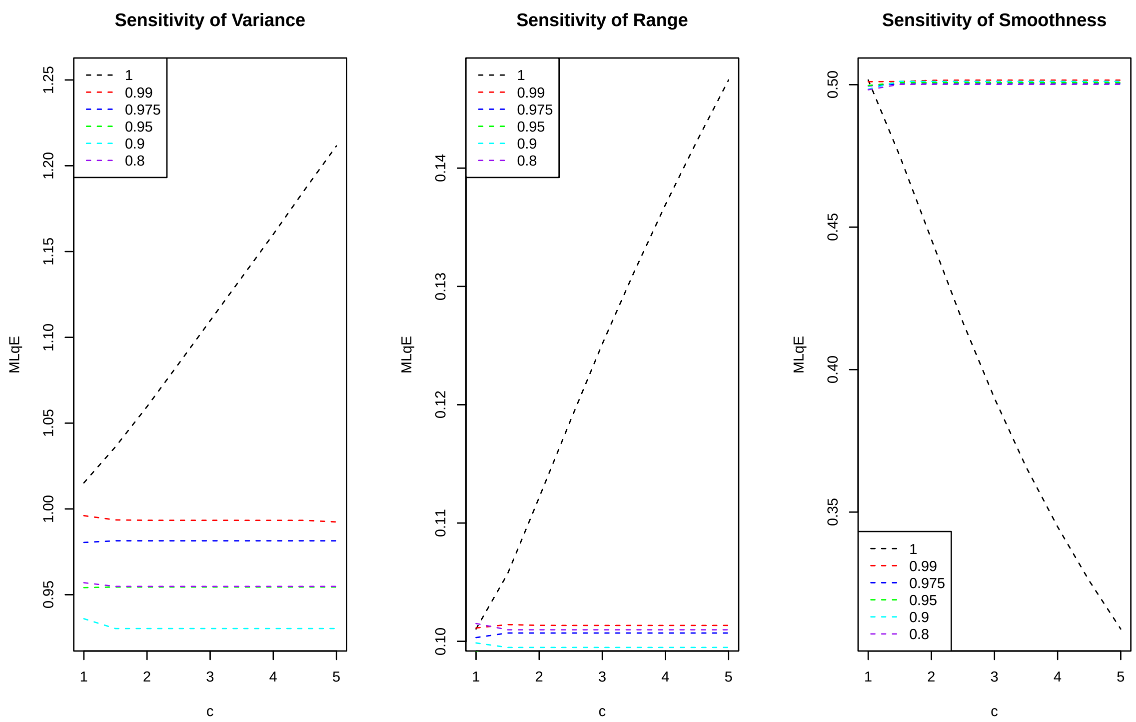

Figure S1: The sensitivity curves of the parameters σ^2, β and ν, with $q = 1, 0.99, 0.975, 0.95, 0.9, 0.8$ and $c = 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5$.

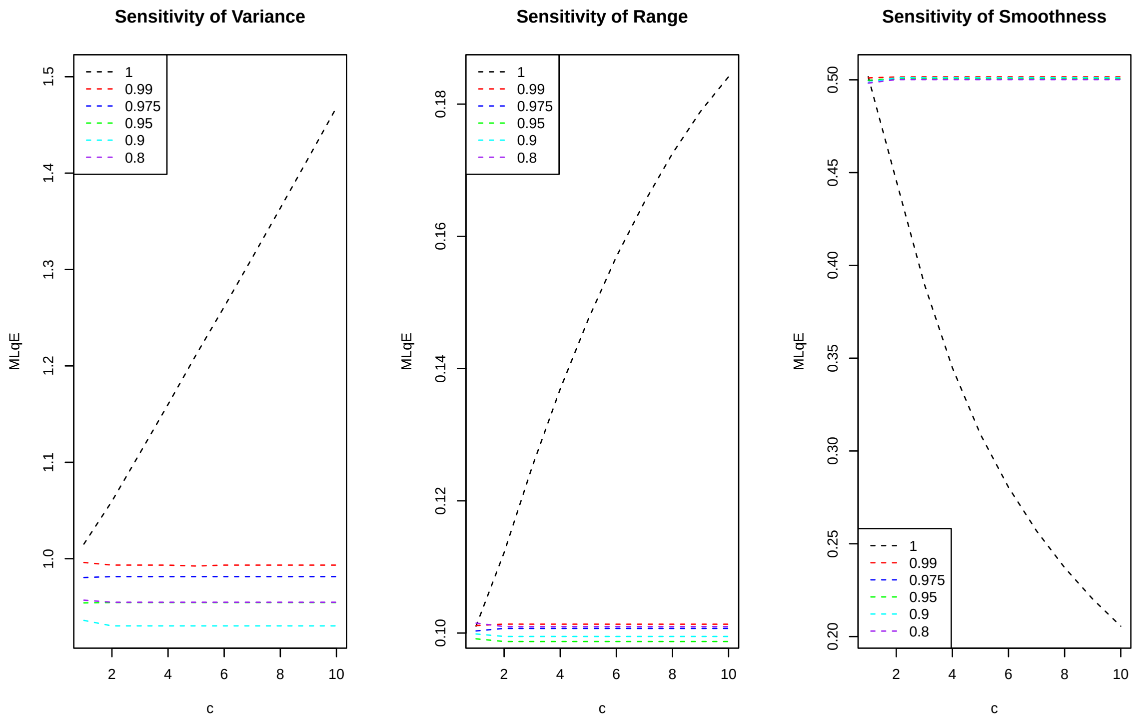

Figure S2: The sensitivity curves of the parameters σ^2, β and ν, with $q = 1, 0.99, 0.975, 0.95, 0.9, 0.8$ and $c = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$.

S2. Computation of Confidence Intervals

In this section, we demonstrate the corresponding asymptotic confidence intervals of the MLqE results for the precipitation dataset considered in Section 4 of the main text. In Theorem 1 of the main text, we showed the asymptotic property of the MLqE: if $\sqrt{m}(q_m - 1) \to 0$ as $m \to \infty$, then we have

$$\sqrt{m}\left(\mathbf{J}_m^{-1}\mathbf{K}_m\mathbf{J}_m^{-1}\right)^{-1/2}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{d} N_p(\mathbf{0}_p, \mathbf{I}_p) \text{ as } m \to \infty.$$

Using this theorem, we may find the asymptotic confidence interval of level α for the MLqE $\hat{\boldsymbol{\theta}}$:

$$\left(\hat{\boldsymbol{\theta}} - z_\alpha^* \frac{(\mathbf{J}_m^{-1}\mathbf{K}_m\mathbf{J}_m^{-1})^{1/2}}{\sqrt{m}}, \hat{\boldsymbol{\theta}} + z_\alpha^* \frac{(\mathbf{J}_m^{-1}\mathbf{K}_m\mathbf{J}_m^{-1})^{1/2}}{\sqrt{m}}\right),$$

where z_α^* represents the the upper $(1 - \alpha)/2$ critical value for the standard normal distribution.

In Table S1, we show the corresponding confidence intervals of the MLqE results in Table 1 of the main text. However, since the asymptotic theorem is under the condition that $\sqrt{m}(q_m - 1) \to 0$, and the size of the dataset is not large enough, the accuracy of the confidence intervals shown here are therefore not guaranteed.

Table S1: The MLqE results of the three parameters of Matérn covariance function, σ^2, β and ν, for the US precipitation data of all the 12 months, followed by the lower bound (lb) and upper bound (ub) of the corresponding asymptotic 95% confidence intervals.

	q selected	$\hat{\sigma}^2$	lb	ub	$\hat{\beta}$	lb	ub	$\hat{\nu}$	lb	ub
Jan	0.95	15.940	13.373	18.507	1.6142	1.6034	1.6250	0.1186	0.1178	0.1194
Feb	0.95	13.543	12.011	15.075	8.7970	8.783	8.811	0.0944	0.0937	0.0951
Mar	0.925	8.9084	−3.3432	21.160	16.919	16.555	17.283	0.0421	0.0276	0.0566
Apr	0.925	8.3288	8.3250	8.3326	1.8346	1.7711	1.8981	0.0375	0.0311	0.0439
May	0.95	13.592	−32.839	60.023	0.4672	−0.9009	1.8353	0.0587	0.0356	0.0818
Jun	0.95	18.490	12.954	24.026	3.9060	3.894	3.918	0.0598	0.0533	0.0663
Jul	0.925	18.613	18.611	18.615	1.2302	1.1430	1.3174	0.0695	0.0612	0.0778
Aug	0.95	15.419	15.417	15.421	3.9982	3.9521	4.0443	0.0335	0.0334	0.0336
Sep	0.95	12.164	8.0266	16.301	0.3535	0.0725	0.6345	0.0393	0.0313	0.0473
Oct	1	23.823	17.227	30.419	0.8104	0.7688	0.8520	0.0772	0.0696	0.0848
Nov	0.925	8.7646	3.5189	14.010	2.7119	2.6279	2.7959	0.0693	0.0632	0.0754
Dec	0.925	8.7253	−13.559	31.010	4.7304	3.9772	5.4836	0.0537	0.0321	0.0753

S3. Supplementary Experimental Results

In Section 3 of the main article, we presented the simulation results using spatial data generated using Matérn kernel with true parameters $\sigma^2 = 1, \beta = 0.1, \nu = 0.5$. Here, we present some simulation results with some different values of β and ν, with other settings the same as in the main article. Generally, we consider the cases with $\beta = 0.03, 0.1$, which we refer to as weak and medium correlation respectively, and $\nu = 0.5, 1$, which we refer to as rough and smooth field respectively. For all the experimental results shown here, we set the number of locations $n = 1{,}600$, number of replicates $m = 100$, and we repeat the experiment 100 times to make the boxplots. All experiments are conducted using the software **ExaGeoStat**.

Medium-smooth Case

Here we present the results for $\sigma^2 = 1, \beta = 0.1, \nu = 1$ in Figures S3 to S6.

Weak-rough Case

Here we present the results for $\sigma^2 = 1, \beta = 0.03, \nu = 0.5$ in Figures S7 to S10.

Weak-smooth Case

Here we present the results for $\sigma^2 = 1, \beta = 0.03, \nu = 1$ in Figures S11 to S14.

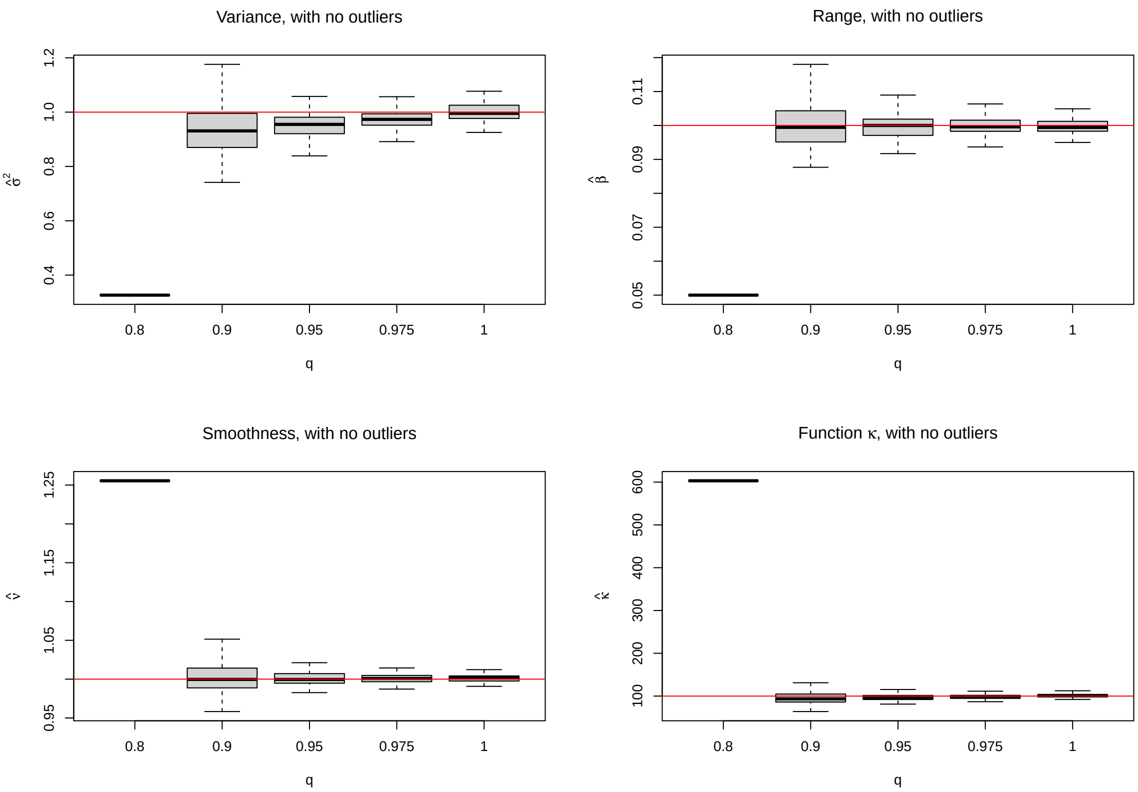

Figure S3: The MLE and MLqE estimation results, where the replicates in the data are not contaminated. The true values of the parameters are $\sigma^2 = 1, \beta = 0.1, \nu = 1$. The rightmost boxplot represents the MLE, while the others represent the MLqE and the values of the x-axis correspond to the values of q. The values of q that we use here are 0.8, 0.9, 0.95, 0.975, 1. The red horizontal lines correspond to the true values of the parameters or the function κ in Equation (7) of the main article.

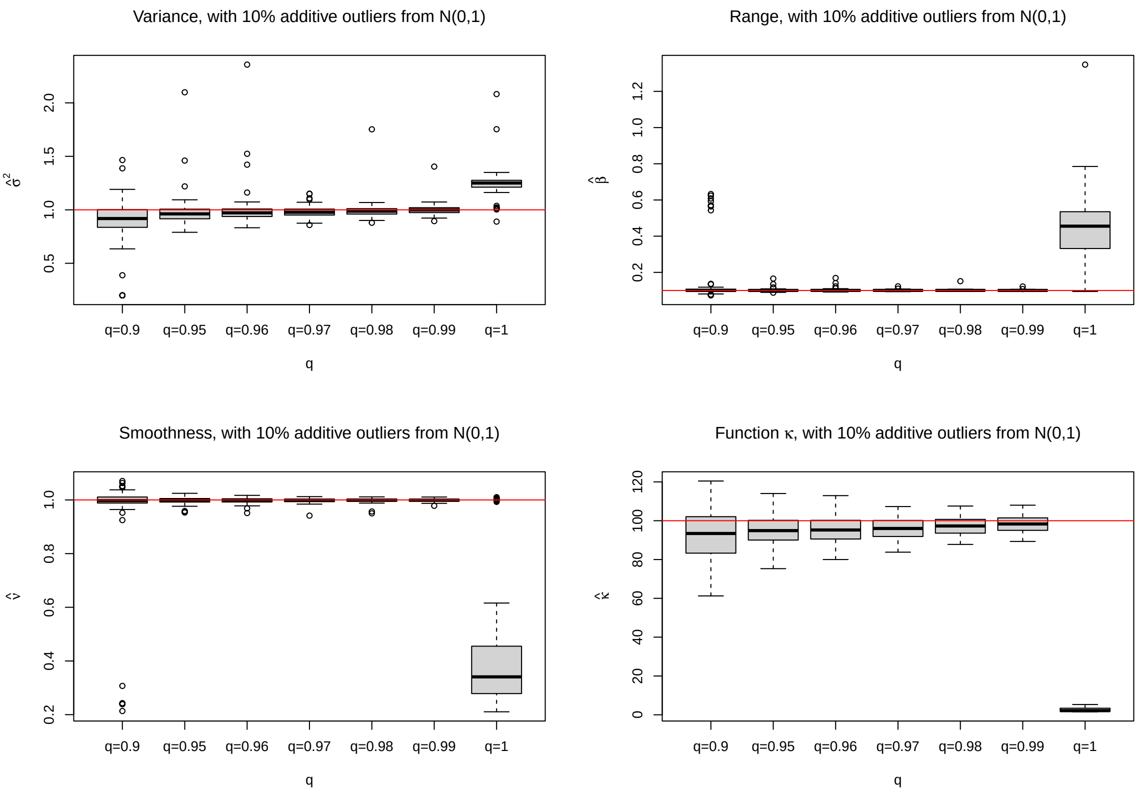

Figure S4: The MLE and MLqE estimation results, where the replicates in the data are contaminated by noises generated from $N(0, 1)$ with probability 10%. The true values of the parameters are $\sigma^2 = 1, \beta = 0.1, \nu = 1$. The rightmost boxplot represents the MLE, while the others represent the MLqE and the values of the x-axis correspond to the values of q. The values of q that we use here are 0.9, 0.95, 0.96, 0.97, 0.98, 0.99, 1. The red horizontal lines correspond to the true values of the parameters or the function κ in Equation (7) of the main article.

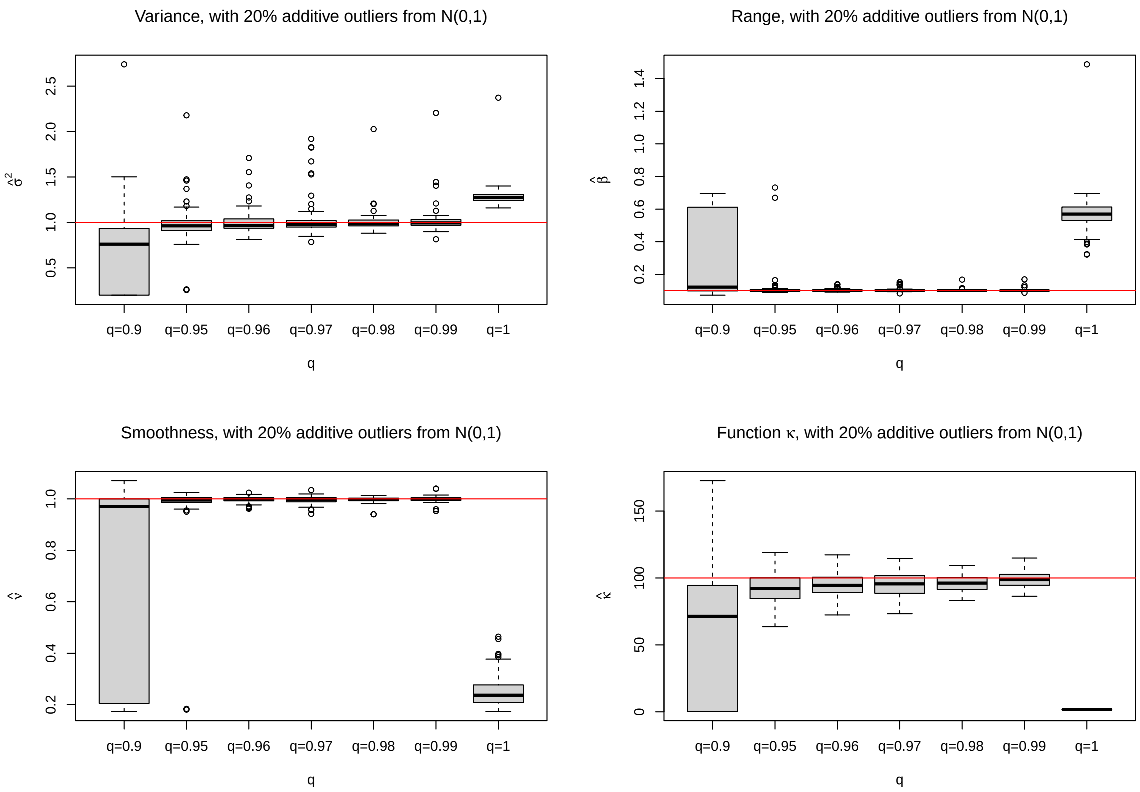

Figure S5: The MLE and MLqE estimation results, where the replicates in the data are contaminated by noises generated from $N(0,1)$ with probability 20%. The true values of the parameters are $\sigma^2 = 1, \beta = 0.1, \nu = 1$. The rightmost boxplot represents the MLE, while the others represent the MLqE and the values of the x-axis correspond to the values of q. The values of q that we use here are 0.9, 0.95, 0.96, 0.97, 0.98, 0.99, 1. The red horizontal lines correspond to the true values of the parameters or the function κ in Equation (7) of the main article.

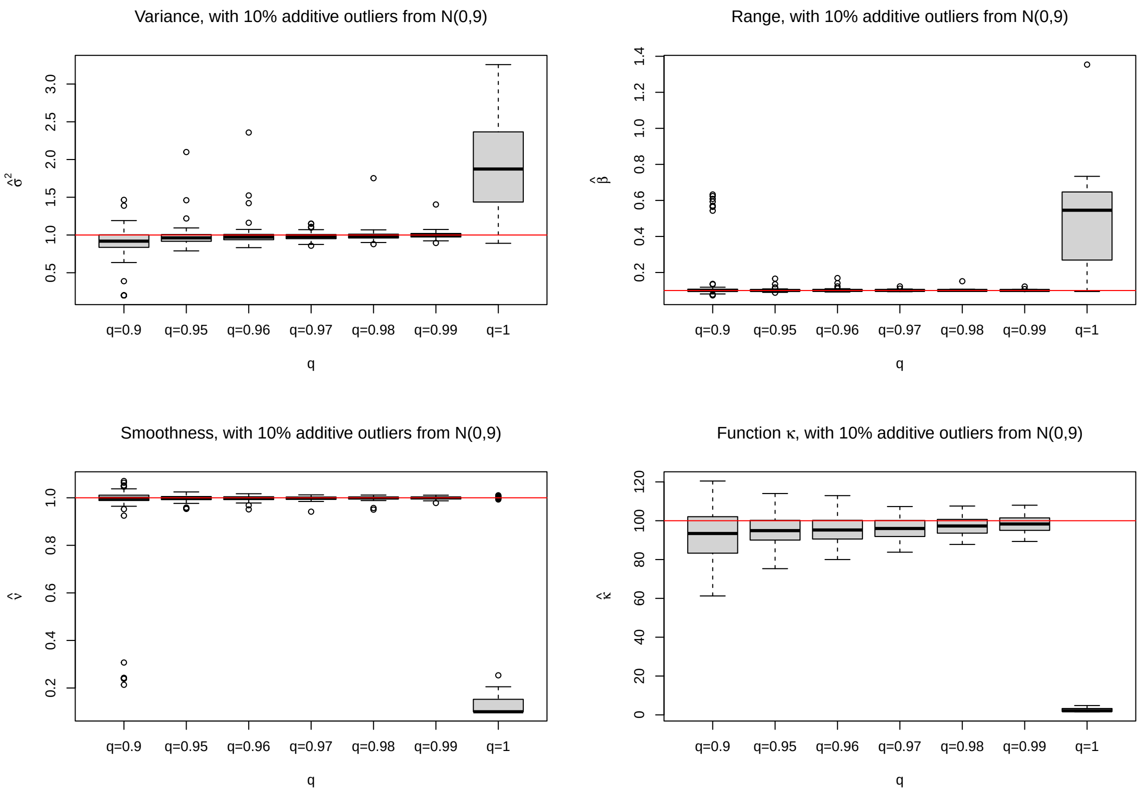

Figure S6: The MLE and MLqE estimation results, where the replicates in the data are contaminated by noises generated from $N(0, 9)$ with probability 10%. The true values of the parameters are $\sigma^2 = 1, \beta = 0.1, \nu = 1$. The rightmost boxplot represents the MLE, while the others represent the MLqE and the values of the x-axis correspond to the values of q. The values of q that we use here are 0.9, 0.95, 0.96, 0.97, 0.98, 0.99, 1. The red horizontal lines correspond to the true values of the parameters or the function κ in Equation (7) of the main article.

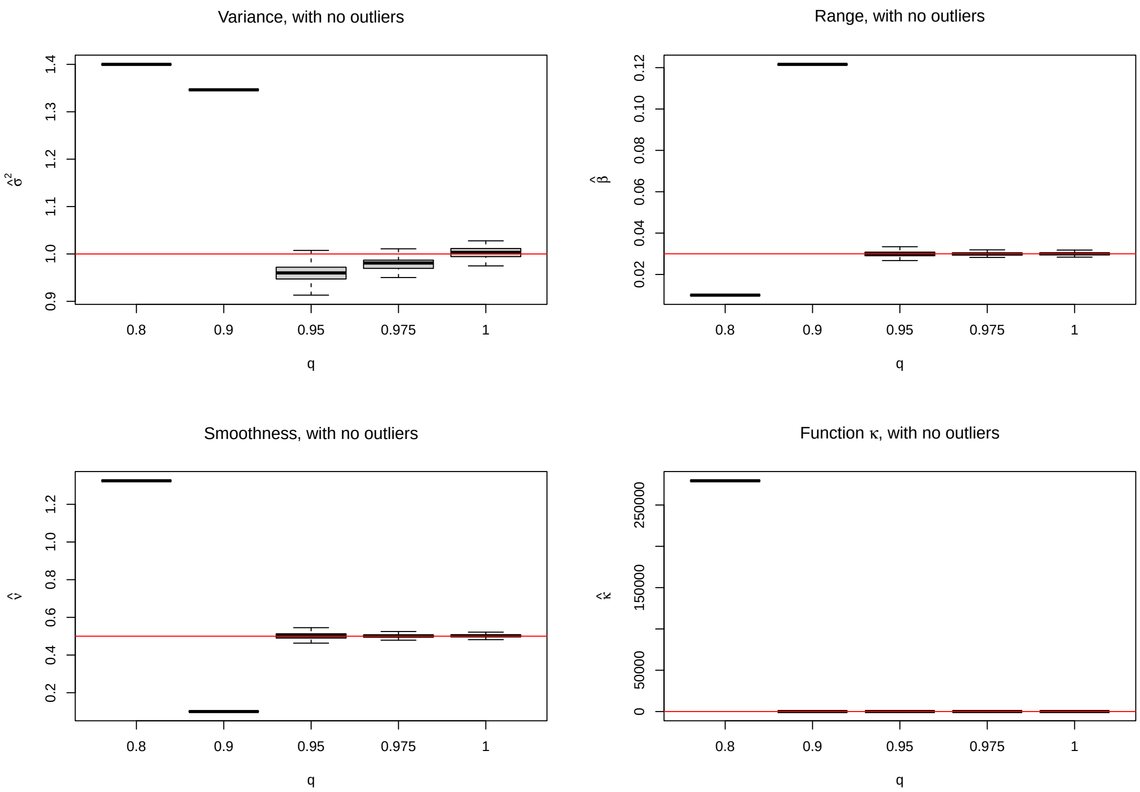

Figure S7: The MLE and MLqE estimation results, where the replicates in the data are not contaminated. The true values of the parameters are $\sigma^2 = 1, \beta = 0.1, \nu = 1$. The rightmost boxplot represents the MLE, while the others represent the MLqE and the values of the x-axis correspond to the values of q. The values of q that we use here are 0.9, 0.95, 0.975, 1. The red horizontal lines correspond to the true values of the parameters or the function κ in Equation (7) of the main article.

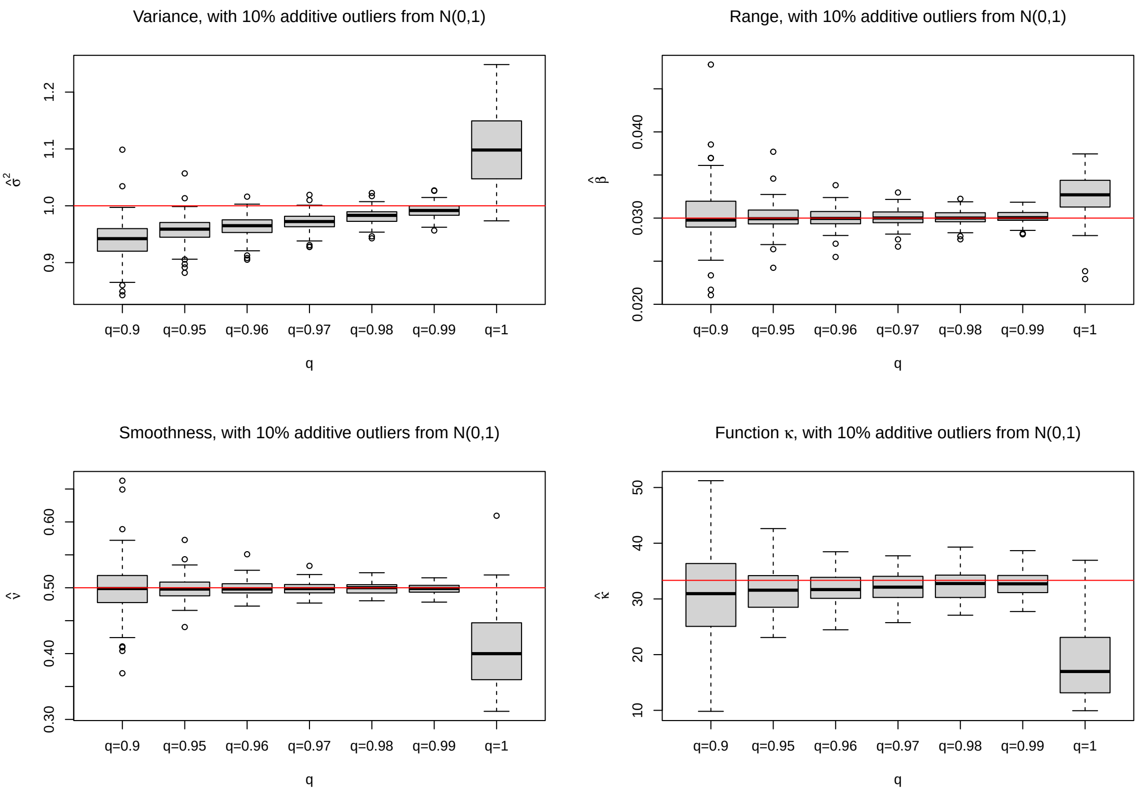

Figure S8: The MLE and MLqE estimation results, where the replicates in the data are contaminated by noises generated from $N(0,1)$ with probability 10%. The true values of the parameters are $\sigma^2 = 1, \beta = 0.03, \nu = 0.5$. The rightmost boxplot represents the MLE, while the others represent the MLqE and the values of the x-axis correspond to the values of q. The values of q that we use here are 0.9, 0.95, 0.96, 0.97, 0.98, 0.99, 1. The red horizontal lines correspond to the true values of the parameters or the function κ in Equation (7) of the main article.

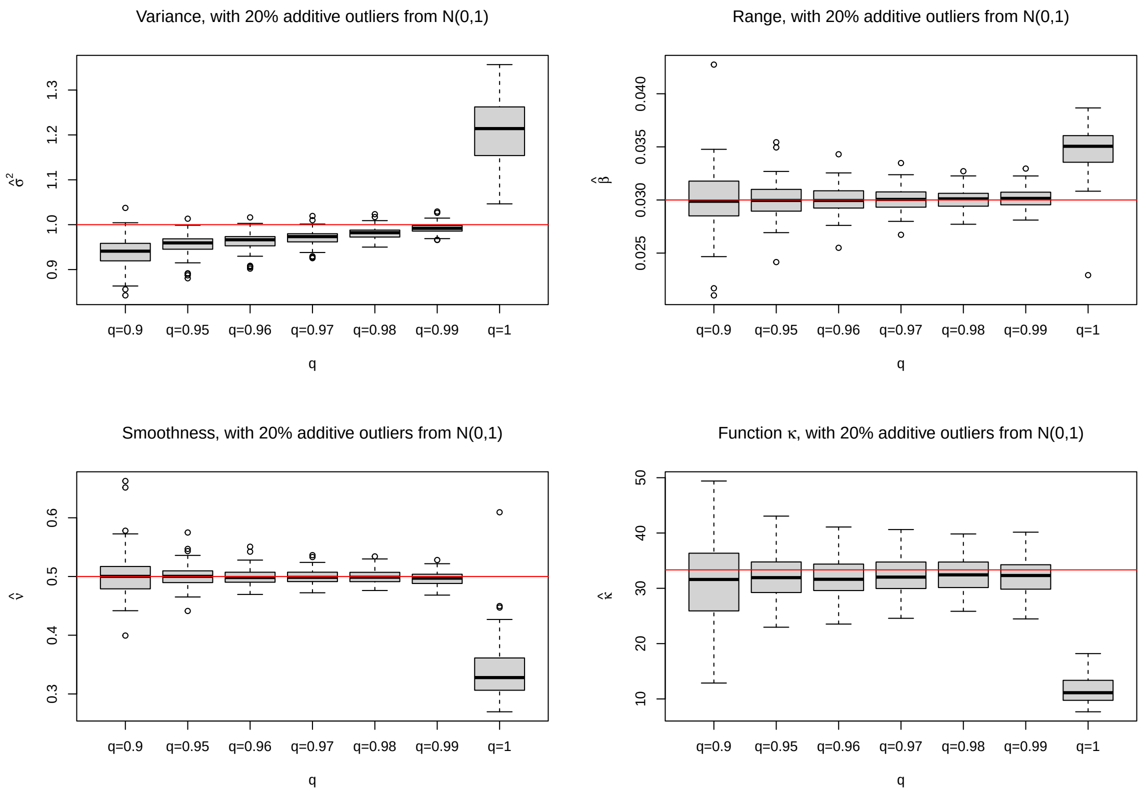

Figure S9: The MLE and MLqE estimation results, where the replicates in the data are contaminated by noises generated from $N(0,1)$ with probability 20%. The true values of the parameters are $\sigma^2 = 1, \beta = 0.03, \nu = 0.5$. The rightmost boxplot represents the MLE, while the others represent the MLqE and the values of the x-axis correspond to the values of q. The values of q that we use here are 0.9, 0.95, 0.96, 0.97, 0.98, 0.99, 1. The red horizontal lines correspond to the true values of the parameters or the function κ in Equation (7) of the main article.

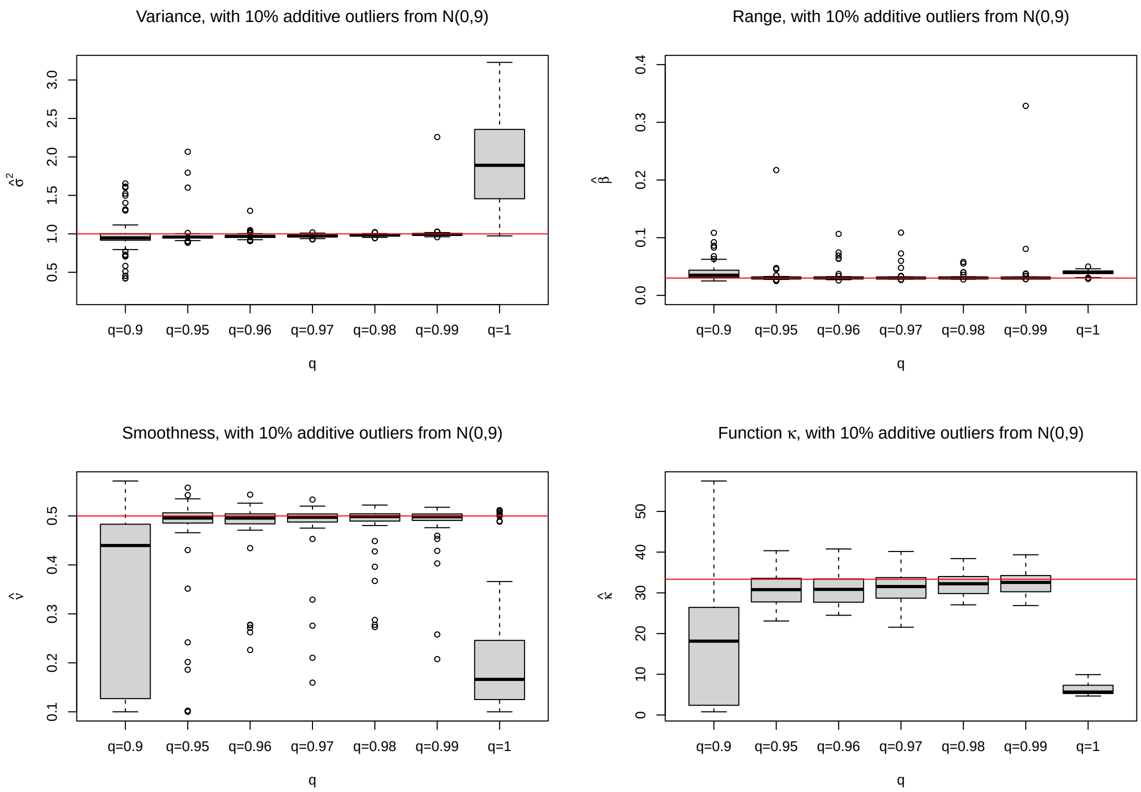

Figure S10: The MLE and MLqE estimation results, where the replicates in the data are contaminated by noises generated from $N(0, 9)$ with probability 10%. The true values of the parameters are $\sigma^2 = 1, \beta = 0.03, \nu = 0.5$. The rightmost boxplot represents the MLE, while the others represent the MLqE and the values of the x-axis correspond to the values of q. The values of q that we use here are 0.9, 0.95, 0.96, 0.97, 0.98, 0.99, 1. The red horizontal lines correspond to the true values of the parameters or the function κ in Equation (7) of the main article.

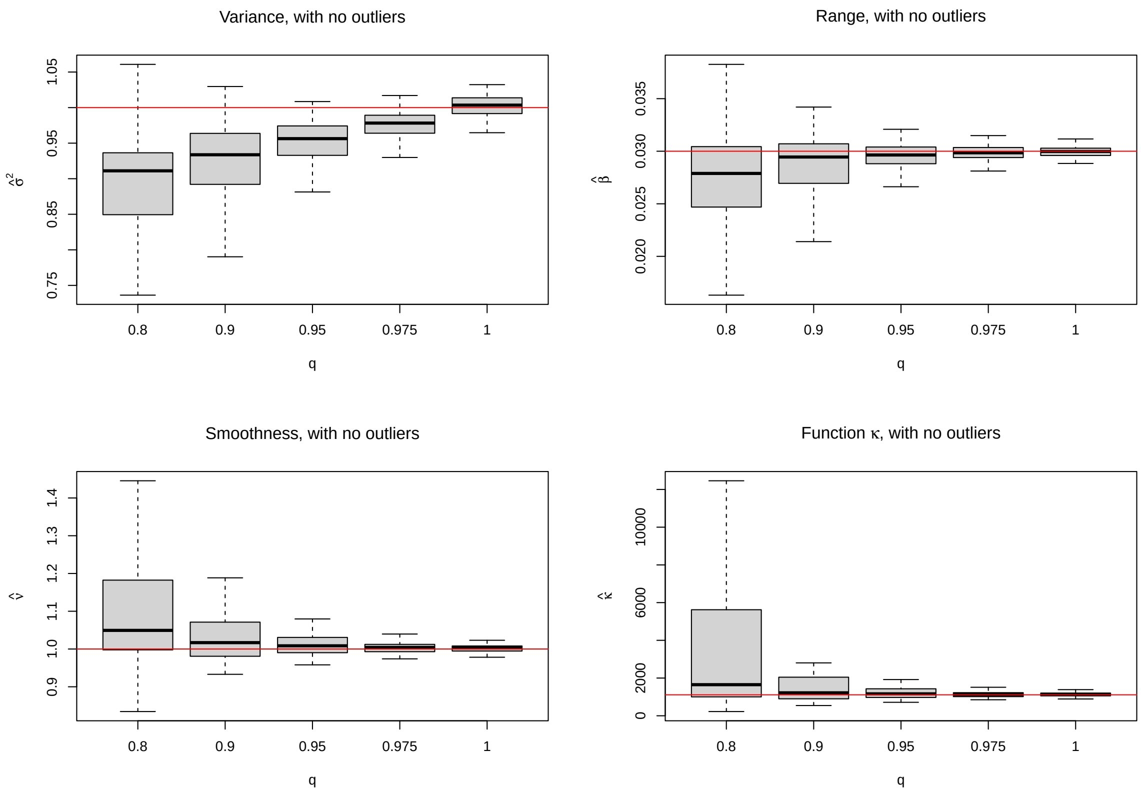

Figure S11: The MLE and MLqE estimation results, where the replicates in the data are not contaminated. The true values of the parameters are $\sigma^2 = 1, \beta = 0.1, \nu = 1$. The rightmost boxplot represents the MLE, while the others represent the MLqE and the values of the x-axis correspond to the values of q. The values of q that we use here are 0.8, 0.9, 0.95, 0.975, 1. The red horizontal lines correspond to the true values of the parameters or the function κ in Equation (7) of the main article.

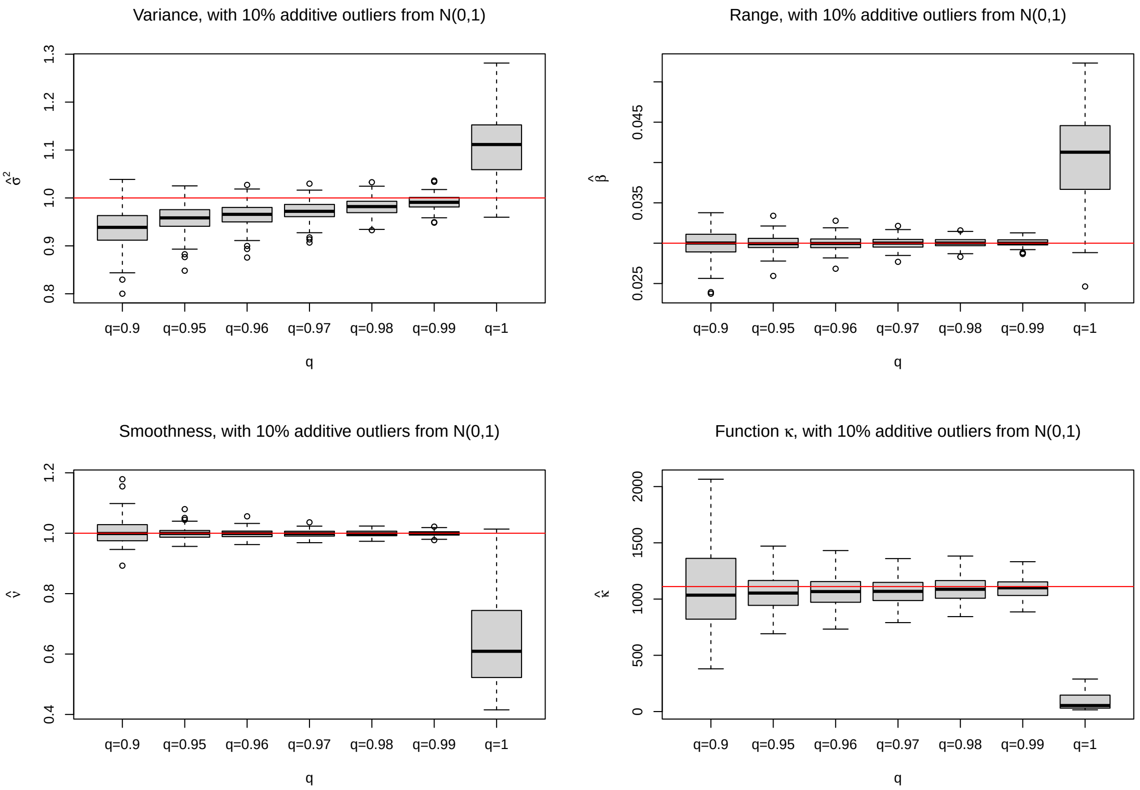

Figure S12: The MLE and MLqE estimation results, where the replicates in the data are contaminated by noises generated from $N(0,1)$ with probability 10%. The true values of the parameters are $\sigma^2 = 1, \beta = 0.03, \nu = 1$. The rightmost boxplot represents the MLE, while the others represent the MLqE and the values of the x-axis correspond to the values of q. The values of q that we use here are 0.9, 0.95, 0.96, 0.97, 0.98, 0.99, 1. The red horizontal lines correspond to the true values of the parameters or the function κ in Equation (7) of the main article.

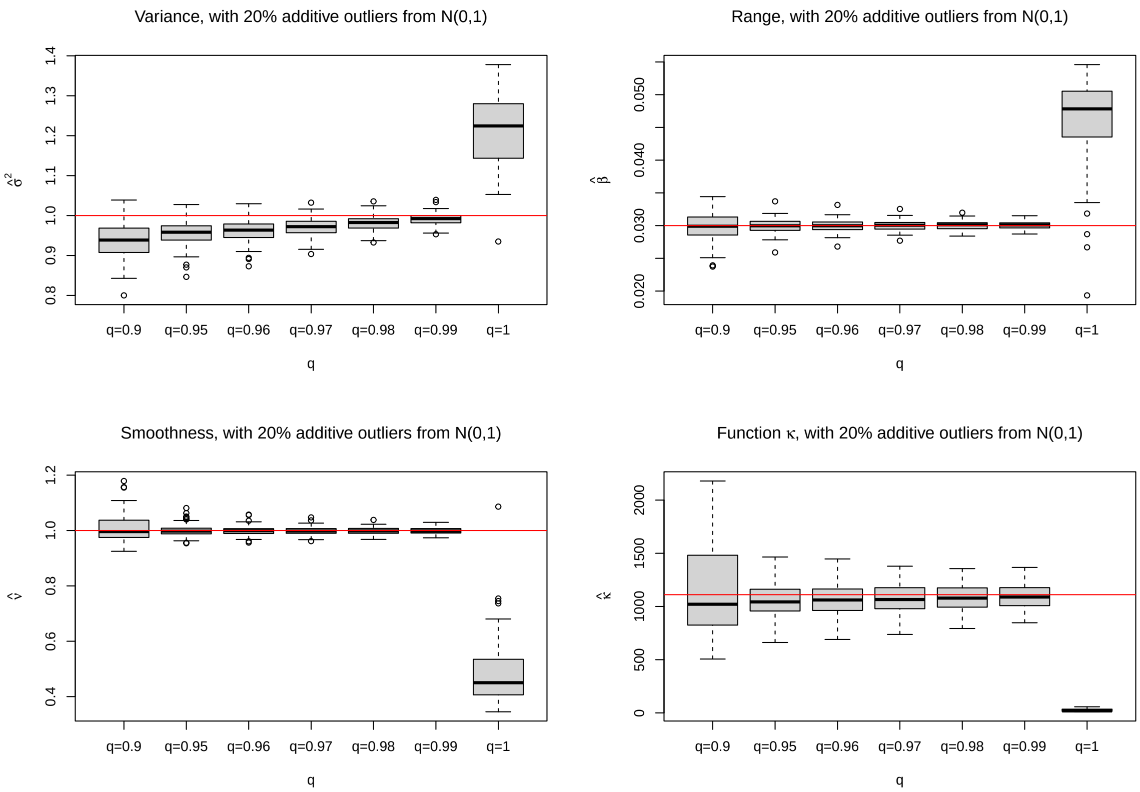

Figure S13: The MLE and MLqE estimation results, where the replicates in the data are contaminated by noises generated from $N(0,1)$ with probability 20%. The true values of the parameters are $\sigma^2 = 1, \beta = 0.03, \nu = 1$. The rightmost boxplot represents the MLE, while the others represent the MLqE and the values of the x-axis correspond to the values of q. The values of q that we use here are 0.9, 0.95, 0.96, 0.97, 0.98, 0.99, 1. The red horizontal lines correspond to the true values of the parameters or the function κ in Equation (7) of the main article.

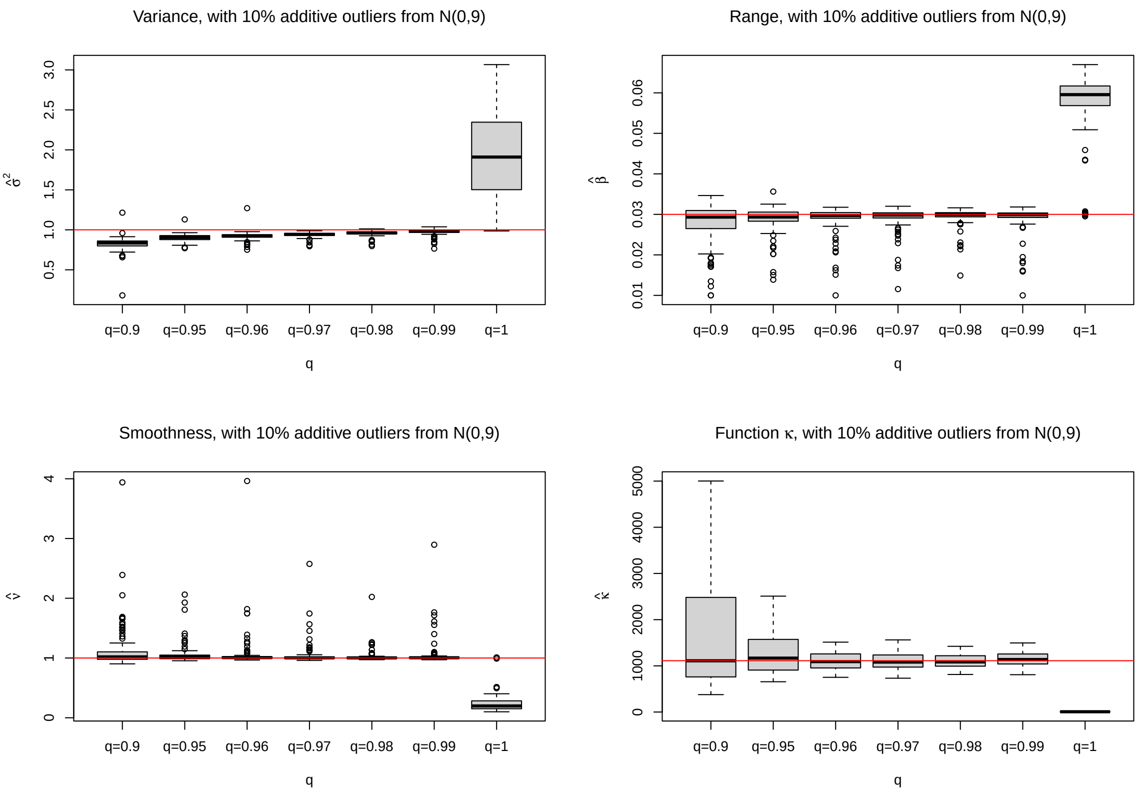

Figure S14: The MLE and MLqE estimation results, where the replicates in the data are contaminated by noises generated from $N(0, 9)$ with probability 10%. The true values of the parameters are $\sigma^2 = 1, \beta = 0.03, \nu = 1$. The rightmost boxplot represents the MLE, while the others represent the MLqE and the values of the x-axis correspond to the values of q. The values of q that we use here are 0.9, 0.95, 0.96, 0.97, 0.98, 0.99, 1. The red horizontal lines correspond to the true values of the parameters or the function κ in Equation (7) of the main article.

Affiliation:

Sihan Chen
Statistics Program
King Abdullah University of Science and Technology
Thuwal, Saudi Arabia, 23955
E-mail: sihan.chen@kaust.edu.sa
URL: https://cemse.kaust.edu.sa/stsds/people/person/sihan-chen

Joydeep Chowdhury
Statistics Program
King Abdullah University of Science and Technology
Thuwal, Saudi Arabia, 23955
E-mail: joydeep.chowdhury@kaust.edu.sa
URL: https://cemse.kaust.edu.sa/stsds/people/person/joydeep-chowdhury

Marc G. Genton
Statistics Program
King Abdullah University of Science and Technology
Thuwal, Saudi Arabia, 23955
E-mail: marc.genton@kaust.edu.sa
URL: https://cemse.kaust.edu.sa/stsds/people/person/marc-g-genton

Journal of Data Science, Statistics, and Visualisation https://jdssv.org/
published by the International Association for Statistical Computing
http://iasc-isi.org/

ISSN 2773-0689
May 2025, Volume V, Issue IV *Submitted:* 2024-11-12
doi:10.52933/jdssv.v5i4.126 *Accepted:* 2025-04-20