JEL Classification: C14, C21, Z32
Keywords: quantile regression, travel and tourism, robust regression, least weighted squares

Least Weighted Squares Quantiles Reveal How Competitiveness Contributes to Tourism Performance

Jan KALINA – The Czech Academy of Sciences, Institute of Computer Science, & Charles University, Faculty of Mathematics and Physics (kalina@cs.cas.cz)
corresponding author

Petra VIDNEROVÁ – The Czech Academy of Sciences, Institute of Computer Science

Abstract

Standard regression quantiles, which are commonly used in heteroscedastic regression models, are highly vulnerable with respect to the presence of leverage points in the data. The aim of this paper is to propose a novel robust version of regression quantiles, which are based on the idea to assign weights to individual observations. The novel method denoted as least weighted squares quantiles (LWSQ) is applied to a world tourism dataset, where the number of international arrivals is modeled for 140 countries of the world as a response of 14 pillars (indicators) of the Travel and Tourism Competitiveness Index (TTCI). Here, the economic motivation is to investigate whether tourism competitiveness promotes tourism performance. The data analysis reveals the advantages of LWSQ. Particularly, LWSQ is able to clearly outperform standard regression quantiles in several artificially contaminated versions of the tourism dataset. From the economic point of view, the study determines countries which are not effective in transforming their competitiveness to higher levels of tourist arrivals.

1. Introduction

In the standard linear regression model, regression quantiles represent a popular statistical methodology for obtaining more complex information compared to fitting a single regression hyperplane (Koenker, 2005; Koenker, 2017). They allow to capture the distribution of errors in the regression model, which makes them especially appealing for (not only econometric) models with heteroscedasticity and/or asymmetric random errors. Because regression quantiles do not have appealing properties in terms of their statistical robustness, i.e. they are vulnerable with respect to the presence of leverage points (outlying values of the independent variables) in the data, it is desirable to search for their more robust alternatives. This motivates us to propose robust regression quantiles based on the highly robust least weighted squares regression estimator; a need for more sophisticated regression quantiles was expressed in Hallin and Šiman (2018). The performance of the novel

https://doi.org/10.32065/CJEF.2022.02.03

The work was supported by the projects GA21-19311S and GA22-02067S of the Czech Science Foundation. The authors would like to thank the editor and two anonymous referees for valuable help; Patrik Janáček provided technical help.

quantiles will be illustrated in a study interested in investigating whether tourism competitiveness promotes tourism performance.

Tourism represents an important sector of the economy with an indisputable contribution to the economic growth or employment opportunities for basically all countries around the world. The direct contribution of travel and tourism to GDP worldwide was approximately 4.5 trillion U.S. dollars in 2019, which corresponds to 10.4 % of global GDP; on the other hand, the total government spending on tourism is estimated to reach 413 billion U.S. dollars (Manzo, 2020).

To recall a few regression quantiles applications to tourism research (mainly related to heteroscedastic financial models), regression quantiles were used to model and predict tourist expenditures in Madeira (Almeida and Garrod, 2017), to study the effect of inbound tourism on foreign direct investments in Arain et al. (2020), or to investigate the stock performance of Korean tourism companies in Jeon (2020). Pérez-Rodríguez and Ledesma-Rodríguez (2021) used regression quantiles to find the main predictors of tourism expenditures in the Canary Islands. An analogous study of the United Kingdom data in Rudkin and Sharma (2017) used the regression quantile methodology for selecting the main predictors as well, but with the aim to recommend optimal decisions for policy makers. Other empirical studies used methods of time series analysis or spatial econometrics (cf. Ouchen and Montargot, 2021) for modeling and predicting the tourism performance evaluated as tourism GDP or the number of international arrivals.

Competitiveness of individual countries in terms of their potential for travel and tourism is measured by the Travel and Tourism Competitiveness Index (TTCI), which is reported biennially since 2007 by the World Economic Forum. As stated in the report of Calderwood and Soshkin (2019), TTCI represents a strategic tool for dialogue and cooperation of relevant stakeholders (politicians, investors, destination managers) in the travel and tourism industry within the participating countries worldwide. The 14 individual pillars of the TTCI characterize a potential or preparedness of countries for the development of travel and tourism, in other words the material and technical basis or precondition evaluated without reflecting the real tourism performance (Das and Dirienzo, 2012). Some of the TTCI pillars were also shown to contribute to the competitiveness of the whole economy in Krstic et al. (2016), who performed cluster analysis of TTCI pillars of 31 sub-Saharan countries, or to be correlated with indexes related to the globalization of the economies worldwide (Ivanov and Webster, 2013). Still, TTCI without being confronted with the real tourism performance would remain limited for any practical applications.

It seems thus natural to model (explain, predict) tourism performance evaluated e.g. by the number of international arrivals to each individual country by the competitiveness evaluated by TTCI pillars. This motivates us to be interested in the hypothesis whether tourism competitiveness promotes tourism performance, in other words if TTCI may be interpreted as a reliable predictor of tourism performance. The question also is which particular TTCI pillars are the most relevant in this relationship. From the application point of view, such investigations allow to evaluate whether a given country practically fulfils its potential for tourism, i.e. whether the tourist performance corresponds to the expected values conditioning on its actual infrastructure. However, such connections of tourism performance with the potential (e.g. evaluated by TTCI) have been described as rare (Bazargani and Kiliç, 2021) and continually overlooked (Hanafiah et al., 2016).

Let us now recall some remarkable studies, where the connection between tourism performance and competitiveness was investigated. The number of international arrivals was shown to be related with some of the TTCI pillars more strongly than with other economic indicators in the study of Hanafiah et al. (2016) over 8 countries of Southeastern Asia; the study however used only correlation analysis and asked for a more advanced statistical treatment of the problem. The number of tourist arrivals in 10 Mediterranean countries was modeled as a response of the TTCI pillars in Marti and Puertas (2017); the regression model considered the numbers of arrivals from one country to another, however only using an artificial dichotomization of the countries to two groups (more competitive and less competitive) to avoid more complex statistical modeling. Bazargani and Kiliç (2021) considered a regression model of tourism performance explained by TTCI pillars for countries across the world. They concluded that the relationships found by ordinary least squares are too heterogeneous, multidimensional, and difficult to interpret, as they used logarithm of all variables including the number of international arrivals and values of TTCI pillars. We are however not aware of any study of the relationship between tourism performance and TTCI based on regression quantile methodology. The importance of regression quantiles for such task was emphasized by Lyócsa et al. (2020), who used advanced time series methods to analyze tourism activity data for 4 countries of Southern Europe; their cross-quantilograms revealed tourism activities to behave similarly across countries.

The methodological Section 2 recalls important robust regression estimators and proposes a robust version of regression quantiles. In Section 3, the novel least weighted squares quantiles are applied to reveal how competitiveness contributes to tourism performance in the analysis of a world tourism dataset, where the number of international arrivals is explained by 14 TTCI pillars across 140 countries. The robustness of the novel quantiles is revealed over artificially contaminated versions of the dataset. Section 4 brings conclusions.

2. Methods

Throughout the paper, the linear regression model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + e_i, \quad i = 1, \ldots, n, \tag{1}$$

is considered, which can be rewritten in the matrix notation as $Y = X\beta + e$. We denote $X_i = \left(X_{i1}, \ldots, X_{ip}\right)^T \in \mathbb{R}^p$ for $i = 1, \ldots, n$.

2.1 Least Squares and Weighted Least Squares

The most common estimation technique for estimating $\beta = \left(\beta_0, \ldots, \beta_p\right)^T \in \mathbb{R}^{p+1}$, i.e. the least squares estimator, is vulnerable to outlying values (outliers) in the data and therefore various robust regression estimators, which are resistant to the presence of outliers in the data, have been proposed and investigated as more reliable alternatives (Maronna et al., 2018; Jurečková et al., 2019).

To avoid further confusions, let us recall the weighted least squares (WLS) with specified non-negative weights $w_1, \ldots, w_n$ fulfilling the natural requirement $\sum_{i=1}^{n} w_i = 1$. Let $u_i(b)$ denote the residual of the i-th measurement based on a given estimate $b = (b_0, \ldots, b_p)^T$ of β, i.e.

$$u_i(b) = Y_i - b_0 - b_1 X_{i1} - \cdots - b_p X_{ip}, \;\; i = 1, \ldots, n. \tag{2}$$

The WLS estimator, which is also known as Aitken estimator or generalized least squares (Greene, 2018), is defined by minimizing a weighted estimate of σ^2 in the form

$$b_{WLS} = \arg \min_{b \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} w_i u_i^2(b).$$

The WLS estimator has an explicit form; denoting by $W = diag(w_1, \ldots, w_n)$ the diagonal matrix with the weights on the main diagonal, the WLS estimator is obtained as

$$b_{WLS} = (X^T W X)^{-1} X^T W Y.$$

2.2 Robust Regression

An important focus of robust statistics is to combine high robustness and high efficiency of estimators (Maronna et al., 2018). Highly robust methods are understood as methods with a high breakdown point, i.e. a high resistance against outliers in the data. Formally, the finite-sample breakdown point is defined as the minimal fraction of data that can drive an estimator beyond all bounds when set to arbitrary values (Davies and Gather, 2005). Because robust statistics evolved as a methodology suitable for a contaminated normal distribution, efficiency is considered as a key characteristic as well. Efficiency of a linear regression estimator evaluates its performance in the model with normal errors without outliers; more formally, it evaluates the asymptotic variability of an estimator relatively to the optimal (smallest) variability, which is achieved by maximum likelihood estimates.

Established robust regression methods for estimating β in (1) include MM-estimators or the least trimmed squares. M-estimators, based on a generalization of the maximum likelihood method (Huber and Ronchetti, 2009), were the first robust regression estimators, which however turned out not to attain a high breakdown point. Subsequently, S-estimators were defined as minimizers of a selected robust measure scale evaluated for the residuals, which achieve a high breakdown point but only a low efficiency (Davies, 1990). Later, MM-estimators (not to be confused with the method of moments) were proposed as two-stage procedures with a computation starting with estimating the scale by an S-estimator, and proceeding with computing an M-estimator with such fixed scale. Such two-stage construction ensures MM-estimators to control both the breakdown point and the efficiency above the required level specified by the user (Yohai, 1987).

The least weighted squares (LWS) estimator (Víšek, 2011; Kalina and Tichavský, 2020) represents another (but much less known) robust alternative to the least squares, which is able to combine high e_ciency and high robustness in terms of the breakdown point, if a suitable weight function is chosen (Čížek, 2011). The LWS estimator, which is distinct from the WLS estimator, turned out to yield reliable results over simulated as well as real data. An extension of the LWS to a robust instrumental variables estimator was proposed in Kalina (2012) and a nonlinear version of the LWS in Kalina et al. (2021). We denote the ranks of (2) by $R_1(b), \dots, R_n(b)$, i.e. with $R_i(b)$ denoting the rank of $u_i^2(b)$ among $u_1^2(b), \dots, u_n^2(b)$, to stress the dependence on b. We will use the concept of weight functions (Víšek, 2011) denoted as $\psi: [0,1] \to [0,1]$ that are defined as non-increasing and continuous functions on $[0,1]$ with $\psi(0) = 1$ and $\psi(1) = 0$.

The LWS estimator formally defined as

$$b_{LWS} = \arg\min_{b \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} \psi\left(\frac{R_i(b) - 1/2}{n}\right)(u_i(b))^2 \tag{3}$$

performs down-weighting of individual measurements through the idea to assign small (or zero) weights to potential outliers and is thus robust also with respect to leverage points, i.e. observations outlying in the independent variables (regressors) (Jurečková et al., 2019). Equivalently, (3) may be expressed as

$$b_{LWS} = \arg\min_{b \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} \psi\left(\frac{i - 1/2}{n}\right) u_{(i)}^2(b),$$

where $u_{(1)}^2(b) \leq \cdots \leq u_{(n)}^2(b)$. In the computations here, we use the LWS with the particular choice

$$\psi_1(t) = \left(1 - \frac{t}{\alpha}\right) I[t < \alpha], \;\; t \in (0,1), \tag{4}$$

where I denotes indicator function and $\alpha \in (1/2\,,1)$ is specified. The choice $\alpha = 3/4$ is used in the computations of this paper. Such choice ensures the breakdown point to be 0.25, which is the most common choice in robust statistics as claimed e.g. by Hubert et al. (2008). The weight function (4) turned to outperform the prediction performance (cross-validation mean square error) of other weight functions in numerical experiments of Kalina and Tichavský (2020).

The least trimmed squares (LTS) estimator can be perceived as a special case of the LWS with weights equal only to 1 or 0. More formally, it is defined as (3) with

$$\psi_2(t) = I[t < \alpha], \;\; t \in (0,1), \tag{5}$$

where again $\alpha = 3/4$ is used in our computations. The LTS estimator corresponds (approximately) to the least squares computed only across _n data points with the smallest squares residuals.

Local robustness with respect to small changes of the data has received much less attention compared to robustness with respect to outliers (see p. 30 of Jurečková et al. (2019)). In (1), the least squares estimator is not sensitive to local replacement of values, i.e. is locally robust. The LTS estimator is known to be highly locally sensitive (see Section 4.9 of Jurečková et al. (2019)) and the LWS is much better in this respect (Víšek, 2011). However, local sensitivity has not been sufficiently investigated for other common robust estimators including MM-estimators. Still, robust regression methods as such are not appropriate for models with a skewed response, which represents a serious complication of analyzing the tourism performance models (Ivanov and Webster, 2013). This follows also from the analysis of the TTCI data by Kalina et al. (2019), who focused on methodological issues of regression modeling under heteroscedasticity. Therefore, we now take resort to considering regression quantiles for the model (1).

2.3 Regression Quantiles

Regression quantiles (Koenker, 2005) are able to capture the relationship of the response on the regressors in a more complex way compared to the least squares regression. In (1), the residuals of approximately $\tau \cdot 100$ % of the data points have to lie below the hyperplane of the regression τ-quantile with a given $\tau \in (0,1)$. Formally, the regression τ-quantile in (1) is defined as

$$b_{RQ(\tau)} = \arg \min_{b \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} \rho_\tau(u_i(b)), \tag{6}$$

where the function ρ_τ is defined as

$$\rho_\tau(x) = \begin{cases} (\tau - 1)x, & if\ x < 0; \\ \tau x, & if\ x \geq 0. \end{cases}$$

Equivalently, (6) can be expressed as

$$b_{RQ(\tau)} = \arg \min_{b \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} \big((1-\tau) I[u_i(b) < 0] + \tau I[u_i(b) > 0]\big)\, |u_i(b)|\,.$$

A set (family) of several regression quantiles jointly has been many times successfully applied to the analysis of heteroscedastic regression models. The L_1 estimator (i.e. the regression quantile with $\tau = 0.5$, regression median, or least absolute deviation estimator) has sometimes been described as a robust estimator, although it does not possess a high breakdown point.

Concerning robustness properties of regression quantiles, they are known to have a bounded influence function and therefore a small gross-error sensitivity (p. 43 of Koenker (2005)) but only a small breakdown point (p. 46 of Koenker (2005)), so they are not robust if the regressors contain leverage points. It is for this reason that regression quantiles are not suitable for outlier detection. In addition, they are not locally robust (Castillo et al., 2008).

2.4 Least Weighted Squares Quantiles

The aim of this section is to propose a novel robust version of regression quantiles for the model (1), stemming from the idea of the LWS or LTS estimators. The least weighted squares quantiles (shortly LWS-quantiles, LWSQ) are formally defined in the following way.

Definition 2.1. Let us consider the linear regression model (1) and let us assume a given weight function ψ. For a given $c > 0$, the LWSQ as an estimator of β in (1) is defined as

$$b_{LWSQ}(c) = \arg\min_{b \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} \psi\left(\frac{R_i(b) - 1/2}{n}\right)\begin{pmatrix} cI[u_i(b) > 0] \\ +I[u_i(b) < 0] \end{pmatrix}(u_i(b))^2. \qquad (7)$$

It will be convenient to consider an equivalent expression for (7) in the form

$$b_{LWS} = \arg\min_{b \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} \psi\left(\frac{i - 1/2}{n}\right)(cI[u_i(b) > 0] + I[u_i(b) < 0])u_{(i)}^2(b). \qquad (8)$$

The LWSQ is based on implicit weights assigned to individual observations, which is the very idea of the LWS (3). In fact, the choice $c = 1$ corresponds exactly to the LWS estimator, while estimates with $c < 1$ focus on upper quantile (ideally above the LWS estimator) and $c > 1$ on lower quantiles. Thus, it is recommendable to consider a family of LWS-quantiles for the model (1) for various values of *c* above 1 as well as below 1. In the computations, we use the weight function ψ_1 defined in (4) and justified in Section 2.2.

As a special case of LWSQ, let us now consider the implicit weights to be equal only to 1 or 0. Such weights corresponding to the weight function ψ_2 (5) are used in the LTS estimator, which is a special case of the LWS. The LWSQ with (5) will be denoted as the LTS-quantile (LTSQ). Using ψ_2 and $c = 1$ exactly yields the LTS estimator.

To study equivariance properties of LWSQ, let us now denote LWSQ (with an arbitrary *c*) computed for the matrix *X* of regressors and the vector of the responses *Y* as $b_{LWSQ}[X, Y]$. Let us recall that scale equivariant estimators in (1) are defined as

$$b_{LWSQ}[X, kY] = kb_{LWSQ}[X, Y] \quad \text{for all } k > 0, \qquad (9)$$

regression equivariant estimators as

$$b_{LWSQ}[X, Y + Xt] = b_{LWSQ}[X, Y] + t \quad \text{for all } t \in \mathbb{R}^p, \qquad (10)$$

and affine equivariant estimators as

$$b_{LWSQ}[XA, Y] = A^{-1}b_{LWSQ}[X, Y] \qquad (11)$$

for all positive definite matrices *A* of dimension $p \times$ p.

Theorem 2.1. In the model (1), the least weighted squares quantiles (LWSQ) are scale equivariant, regression equivariant, and affine equivariant.

The computation of the LWSQ and LTSQ may exploit a straightforward adaptation of the FAST-LTS algorithm of Rousseeuw and Van Driessen (2006). While there is no immediate interpretation of *c* within LWSQ, it remains possible to use LWSQ to divide the observations to 5 groups. This can be performed using Algorithm 1 for LWSQ and in an analogous way for LTSQ, using the notation $\lfloor x \rfloor$ for the integer part of $x > 0$. With such notation, $\tilde{c}_1$ corresponds to the bottom quintile and $\tilde{c}_4$ to the top quintile.

The step of formula (3) is conservative for situations with quantile crossing, which is a phenomenon common for standard regression quantiles. To explain this, let us consider an example of regression quintiles (i.e. the set of 4 quantiles dividing the errors to 5 groups of about equal sizes). It would be undesirable if e.g. the computed 3rd regression quintile of the response was smaller than the 2nd quintile. While our main focus is robustness and not quantile crossing, let us remark that available non-crossing regression quantiles are obtained as constrained versions of regression quantiles with a simple condition not allowing the crossing (Bondell et al., 2010) or with parametric models for the conditional quantile functions, which is the case of the NCQR approach of Sottile and Frumento (2021). All these non-crossing approaches however remain vulnerable with respect to leverage points.

The result of LWSQ (and particularly of Algorithm 1) can be visualized by means of a map, which will be denoted as LWS-ReQuieM (as abbreviation of LWS-Regression QUIntilE Map). The map is obtained as a map depicting the quantity (3) for each particular country. An analogous map may be created also for standard regression quintiles with given constants $\tau_1, \dots, \tau_4$ (and without a need to apply Algorithm 1). In such a case, the map will be denoted as RQ-ReQuieM, i.e. regression quintile map obtained with (standard) regresssion quantiles.

ALGORITHM 1 Dividing the Observations in (1) to 5 Quintiles According to LWSQ.

Data: Data for the regression model (1)
Result: A factor variable $(q_1, \dots, q_n)^T$ assigning each country to one 5 quintiles according to LWSQ
Result: Optimal values of $\tilde{c}_1,\dots,\tilde{c}_4$ yielding such assignment

1.
$$\tilde{c}_1 := \arg\min_{c>0} \sum_{i=1}^{n} I\left[Y_i \leq X_i^T b_{LWSQ}(c)\right] = \lfloor n/5 \rfloor$$

2. For $j \in \{2,3,4\}$, find $\tilde{c}_j$ as
$$\tilde{c}_j := \arg\min_{0\leq c\leq \tilde{c}_{j-1}} \sum_{i=1}^{n} I\left[X_i^T b_{LWSQ}(\tilde{c}_{j-1}) \leq Y_i \leq X_i^T b_{LWSQ}(c)\right] = \lfloor n/5 \rfloor$$

3.
$$q_i := \sum_{j=1}^{4} I\left[Y_i \leq X_i^T b_{LWSQ}(\tilde{c}_j)\right], \quad i = 1, \dots, n.$$

3. Analysis of the World Tourism Data

3.1 Data Description

The aim of the analysis is to model the number of international arrivals (in millions, denoted as Y) as a response of $p = 14$ pillars of TTCI overviewed in Table 1. We work with $n = 140$ countries, for which both the response and the TTCI pillars are available. The regressors are used from Calderwood and Soshkin (2019) and the values of the response (originally published by World Bank) from Roser

(2020). The response comes from the year 2016; because the response is not available for a few of developing countries for 2016, we had to take their response from the last available year. We perform all computations in R software using packages het.test, quantreg, qrcm, RobRSVD, robustHD, rrcov, and rworldmap.

Table 1 The 14 Pillars of the Travel and Tourism Competitivenes Index (TTCI), Which Are All Used as Predictors in the Main Model (12)

Index		*Pillar*
1	*BE*	Business Environment
2	*SS*	Safety and Security
3	*HH*	Health and Hygiene
4	*HRLM*	Human Resources and Labor Market
5	*ICT*	Information and Communication Technologies Readiness
6	*TT*	Prioritization of Travel and Tourism
7	*IO*	International Openness
8	*PC*	Price Competitiveness
9	*ES*	Environmental Sustainability
10	*ATI*	Air Transport Infrastructure
11	*GPI*	Ground and Port Infrastructure
12	*TSI*	Tourist Service Infrastructure
13	*NR*	Natural Resources
14	*CRBT*	Cultural Resources and Business Travel

3.2 Exploratory Data Analysis

The response and all the considered regressors are continuous and always positive and the regression errors are highly asymmetric. The median of the response is 2.55 million; 33 countries (i.e. 23 % of the total number of the countries) have the response above 10 million. All further computations consider LTS and LTSQ with the weight function (5), LWS and LWSQ with (4), and MM-estimators with breakdown point 0.5 and efficiency 0.95.

Our basic model is the linear model (1) with p = 14, i.e.

$$Y_i = \beta_0 + \beta_1 BE_i + \cdots + \beta_{14} CRBT_i + e_i, \qquad i = 1, \ldots, n, \tag{12}$$

Testing by means of Student's t-test reveals that pillars CRBT, GPI and TSI are significant in (12), while CRBT has the strongest association with the response. The plots of Y against single individual regressors reveal a strong heteroscedasticity in (12); some of such plots are shown in Figure 1. Values of the correlation coefficient r computed for important pairs of regressors are equal to $r(X_{14}, X_{11}) = 0.41, r(X_{14}, X_{12}) = 0.40,$ and $r(X_{11}, X_{12}) = 0.70.$ The quite high value of the coefficient of determination $R^2 = 0.70$ in (12) advocates using the linear model; a similar value 0.68 is obtained across the 75 % of observations used within the LTS estimator. The linearity of the trend in (12) was also found to be reasonable in plots of Y against every regressor separately. Also standard diagnostic tools including the plot of residuals against fitted values reveal the linear model to be suitable, although heteroscedasticity is apparent there (as discussed in Section 3.3).

Further, we performed outlier detection by means of the LTS and LWS estimators. This approach declares such observations to be outliers which fulfil

$$\left|u_i^{LWS}\right| > 2.5\,\hat{\sigma}, \qquad i = 1, \dots, n,$$

where $u_1^{LWS}, \dots, u_n^{LWS}$ are residuals of the LWS fit and $\hat{\sigma}_{LWS}$ is an LWS- or LTS estimate of σ (Kalina, 2012). The method finds 5 outliers (France, United States, Spain, China, and Italy), which are in fact the 5 countries with the largest response. An analogous outlier detection based on the LTS detects the same outliers as the LWS.

3.3 Heteroscedasticity Testing and Modeling

In (1), heteroscedasticity is confirmed by the White test (White, 1980) with *p*-value $p \approx 4 \cdot 10^{-6}$ or the Breusch-Pagan test with $p \approx 2 \cdot 10^{-6}$, where the latter explains the heteroscedasticity by fitting residuals of (1) against all regressors. Because the estimates in (1) are influenced by heteroscedasticity, it is recommended to estimate its parameters by means of Aitken estimator (Section 2.1). Here, let us consider the Aitken model

$$\frac{Y_i}{\sqrt{k_i}} = \gamma_0 + \gamma_1 \frac{X_{i1}}{\sqrt{k_i}} + \cdots + \gamma_p \frac{X_{ip}}{\sqrt{k_i}} + f_i, \qquad i = 1, \dots, n, \tag{13}$$

with parameters $\gamma_0, \dots, \gamma_p$ and random errors $f_1, \dots, f_n$. We take $\sqrt{k_i} = \hat{Y}_i$ for $i = 1, \dots, n,$ where $\hat{Y}_i$ is the estimated value of Y_i obtained by the least squares in the original model (12). The White test performed in the Aitken model yields *p=0.96*. Although it may seem that it is preferable to use the Aitken model, which is in fact popular in tourism modeling applications (Assaf and Tsionas, 2020), graphical visualizations not shown here reveal that there is basically no trend of the response against the regressors in the Aitken model (even after a possible trimming away of potential outliers). We thus conclude that the auxiliary regression model (13) is not precisely suitable; instead of a tedious tuning of this transformed model, we take resort to regression quantiles and LWSQ, where mainly the latter turns out to be suitable for the given heteroscedastic data (without any specific tuning).

Table 2 Results of Hypothesis Tests Applied to Regression Quantiles and LWS-Quantiles. Values of $\tilde{c}_1,\dots,\tilde{c}_4$ Were Found by Algorithm 1

Method	*Significant variables in (12)*
Regression quantile (τ = 0.2)	CRBT, PC, NR
Regression quantile (τ =0.4)	CRBT, SS, PC, TSI, ICT, HH
Regression quantile (τ =0.6)	CRBT, TSI, PC, ICT
Regression quantile (τ =0.8)	CRBT, ES, SS, ATI
LWSQ ($\tilde{c}_1$)	CRBT, PC
LWSQ ($\tilde{c}_2$)	CRBT, PC, TSI, SS
LWSQ ($\tilde{c}_3$)	CRBT, TSI, PC
LWSQ ($\tilde{c}_4$)	CRBT, TSI, ATI, SS

3.4 Regression Quantiles and LWSQ

We computed regression quantiles and their non-crossing alternative (NCQR) of Sottile and Frumento (2021) in the model (12), so that the countries are divided to

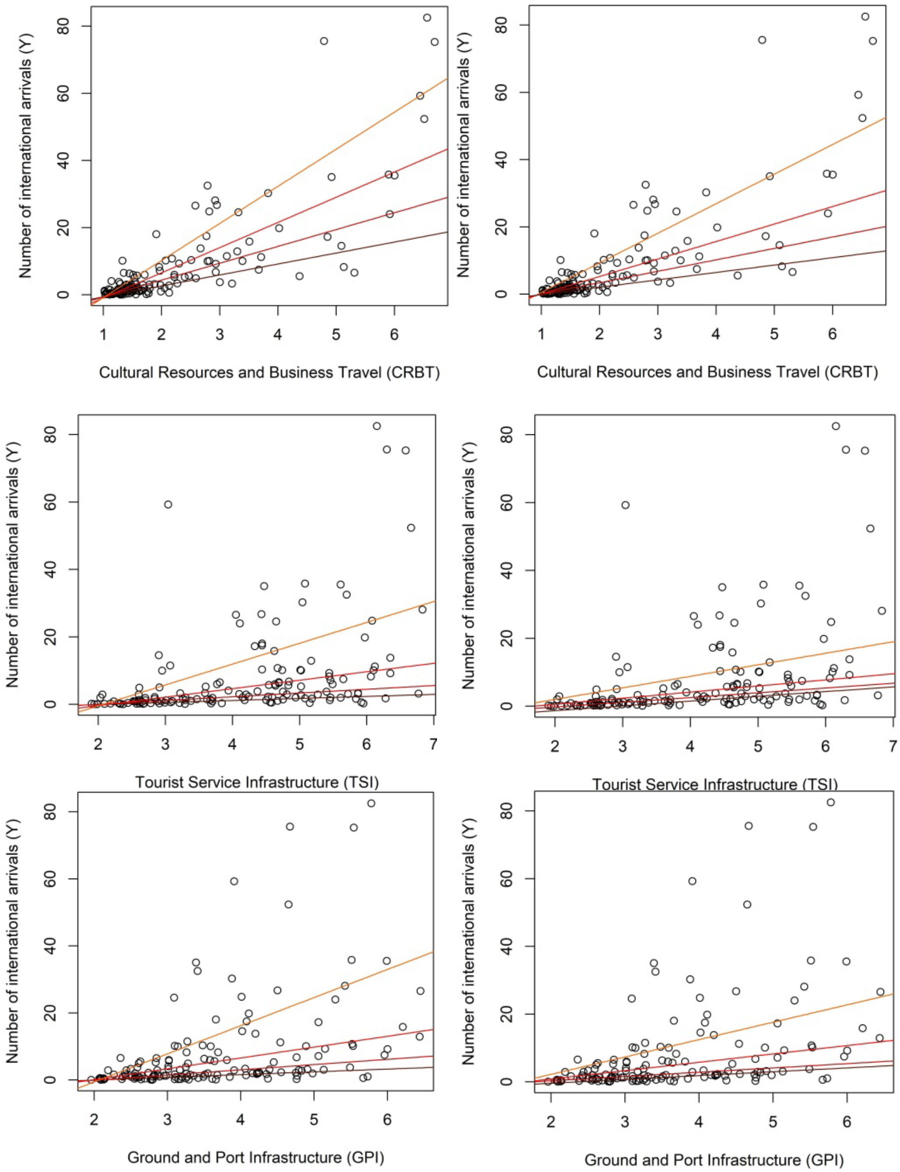
5 quintiles. Also LTS-quantiles and LWS-quantiles are computed, which divide the countries 5 quintiles; this computation used the automatic procedure of Algorithm 1 for finding suitable values of $\tilde{c}_1,\ldots,\tilde{c}_4$.

We performed a variable selection for regression quantiles and LWSQ by means of a standard backward stepwise selection procedure. Individual tests for regression quantiles are based on regression rank scores (RSS) as proposed by Gutenbrunner et al. (1993), and tests for LWSQ are based on nonparametric bootstrap. The significant variables are reported in Table 2. For regression quantiles, the significant variables are quite unstable in their dependence on τ, and some of the regressors (especially for $\tau = 0.4$) have their p-values close to 5 %, so their effect is not unambiguous. For LWSQ, it is clearer to see how the significant regressors depend (in a more stable way) on c: CRBT remains the most important predictor of Y for all quantiles, PC is the second strongest predictor for lower quantiles, and TSI the second strongest predictor for upper quantiles. Such finding has the following economic interpretation. CRBT is focused on cultural resources, i.e. potential for an interesting spending of leisure time, which is an important factor for basically all destinations. Specifically for visitors of countries in top quantiles, high quality of services and a higher level of tourist infrastructure (TSI) are important. For visitors of countries in bottom quantiles with less developed services, it is the lower price (i.e. an appealing level of PC) which makes the countries attractive and has thus effect on the tourism performance.

To perform a comparison of various tools used in (12), we used nonparametric bootstrap in a standard way to estimate variability of $b_0, \ldots, b_p$ for various estimators. Estimates of β_{14}, i.e. of the parameter corresponding to the CRBT pillar, are reported in the first column of Table 3. Nonparametric bootstrap estimates of standard deviations are presented there as well, which were obtained with 1000 bootstrap samples in each case. The variability of each type of quantiles depends on the actual quantile and is typically increasing with an increasing quintile. The variability of LWSQ is smaller compared to that of regression quantiles and especially compared to LTSQ; in an analogous way, the LWS estimator has a smaller variability compared to the LTS. An interpretation of the regression quantiles and LWSQ on the level of individual countries follows in Section 3.5.

A graphical visualization of regression quantiles may be presented only for the relationship of Y against a single regressor; such graphs for selected regressors reveal a clear heteroscedasticity, as shown in the left images of Figure 1, where the variability of Y typically increases with increasing values of individual pillars. Standard regression quintiles give a quite poor fit especially for countries with a low value of the response (also due to their leverage effect). For regression quantiles, quantile crossing does not represent a major problem of regression quantiles here, as revealed in our evaluation using the diagnostic tools (function diagnose.qc) of the qrcm package (Frumento, 2021). Graphical visualizations of NCQR are quite similar to those obtained with standard regression quantiles. The LWS-quantiles for the relationship of Y against a single regressor are presented in the right images of Figure 1. The LWS-quantiles do not suffer from quantile crossing and mainly are more parallel (and thus more suitable for countries with a low value of the response) compared to standard regression quantiles. An interpretation of regression quantiles and LWSQ on the level of individual states follows in Section 3.5.

Figure 1 Plots of the Response Y (Number of International Arrivals) Against Selected Individual Pillars of TTCI

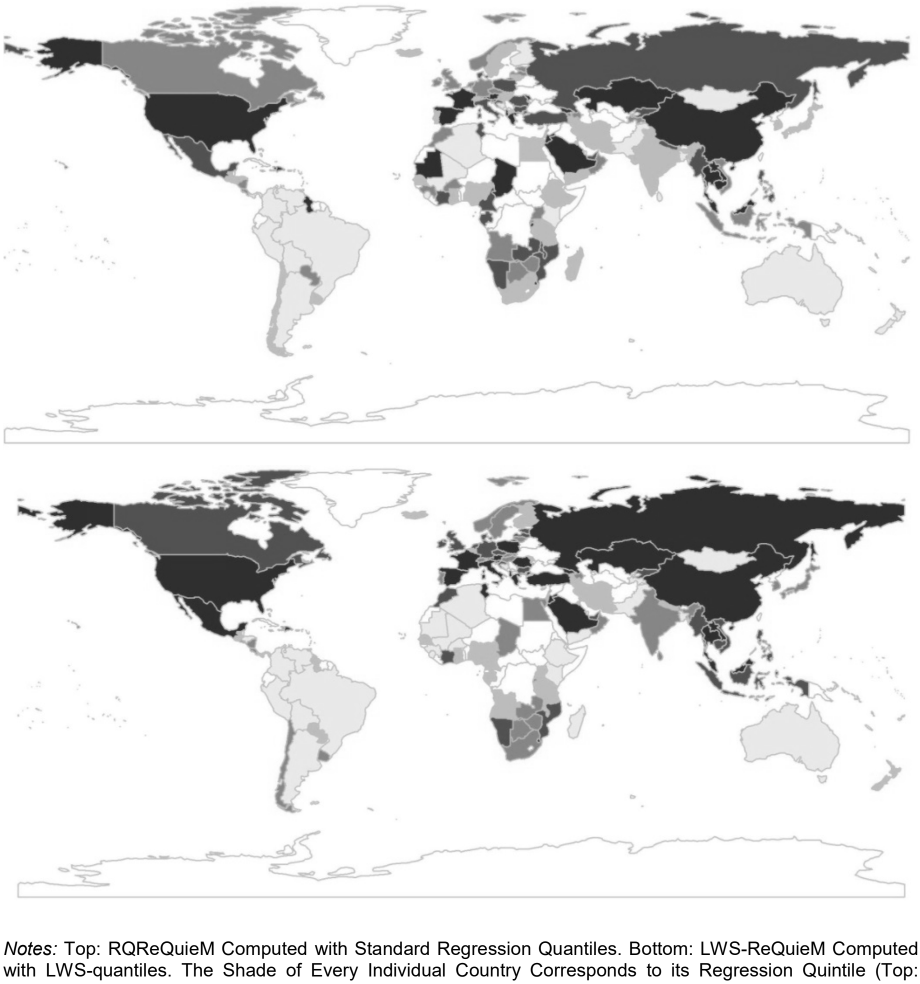


Notes: Left: Regression Quintiles with the Parameter τ Equal to 0.2 (Bottom Line), 0.4, 0.6, and 0.8 (Top Line). Right: LWS-quintiles with Values of the Parameter *c* Automatically Obtained by Algorithm 1. First Row: *Y* Against CRBT (Cultural Resources and Business Travel); Second Row: *Y* Against TSI (Tourist Service Infrastructure); Third Row: *Y* Against GPI (Ground and Port Infrastructure).

3.5 Regression Quantiles and LWSQ: Interpretation by Means of Quintile Maps

Returning to the full model (12), the multivariate information captured by regression quantiles or LWS-quantiles about the relationship between Y and the regressors on the nationwide level may be revealed by the RQ-ReQuieM map or LWS-ReQuieM map, which are shown in Figure 2; there, countries with unavailable data (i.e. not included in the $n = 140$ countries) remain white. These maps show 5 groups of countries around the world according to their real tourism performance conditioning on their theoretical potential (expressed by TTCI pillars); such knowledge cannot be acquired by means of standard tools of Sections 3.2 or 3.3.

Figure 2 Regression Quintile Maps of the World Computed for the Model (12)



Notes: Top: RQReQuieM Computed with Standard Regression Quantiles. Bottom: LWS-ReQuieM Computed with LWS-quantiles. The Shade of Every Individual Country Corresponds to its Regression Quintile (Top: Standard, Bottom: LWSQ). There are 5 Shades from Light (Bottom Quintile) to Dark (Top Quintile); for White Countries/Regions, the Data are not Available.

These maps allow a clear economic interpretation of the computed quantiles. In reality, some countries may have a much higher level of tourism performance than one could expect based on its values of the 14 pillars. This is e.g. the case of Saudi Arabia with numerous religious pilgrims who are not discouraged by a low tourist infrastructure, or Turkey with beautiful beaches and good (but unmeasurable) reputation among travellers. On the other hand, it is even more useful to identify countries which do not efficiently exploit their tourism capacity relatively to (or conditioning on) their highly developed infrastructure. Such countries with a low effectivity of exploiting their (actually high) potential for tourism include Belgium, Switzerland, or Finland; also Portugal remains to belong among these countries to some extent, in spite of its large investments to tourism in the last decade.

A comparison of regression quantiles and LWSQ, i.e. a comparison of the top and bottom maps of Figure 2, reveals however some remarkable differences for some countries. Italy belongs to the top fifth according to LWSQ, which is reasonable, as it belongs to the 5 countries with the largest Y (cf. Section 2.1); it surprisingly belongs only to the second top fifth according to standard regression quantiles. Further, let us consider Chad with a very small number of tourists (its Y is 0.12) and at the same time very low performance in the pillars. The fitted values of all 4 standard regression quintiles evaluated for Chad are negative and thus the country would appear in the top fifth even with a zero number of tourists. This problem disappears if using LWSQ, where Chad is found in the middle fifth among all the countries. More realistic results are obtained with LWSQ also for Mauritania or Guyana. On the whole, we believe that the results of LWSQ are more suitable compared to those of standard regression quantiles especially due to the leverage effect of the countries with a very small number of tourists, or due to local sensitivity of regression quantiles; this motivates us to perform further comparisons of the quantiles (and other estimators).

3.6 Artificially Contaminated Data

To show the main advantages of LWSQ, i.e. its robustness with respect to data contamination, we compared various methods for 4 contaminated versions of the dataset of Section 3.1. For each of the following contamination scenarios denoted as (A)–(D), the contaminated data are generated 1000 times and the presented results are averaged.

(A) *Asymmetric contamination by severe outliers for countries with a large response. For the 25 countries described for scenario B, the response Y_i is replaced by $Y_i - |Z_i|$, where $Z_1, \ldots, Z_n$ are i.i.d. generated from the normal N(0; 25) distribution.*

(B) *Symmetric contamination by severe outliers for countries with a large response. Particularly, if $CRBT_i > 4$ and $Y_i > 15$ for the i-th country, the response Y_i is replaced by $Y_i + Z_i$, where $Z_1, \ldots, Z_n$ are i.i.d. generated from N(0; 49). In this way, 25 countries are contaminated.*

(C) *Contamination for countries with a very small response. If $CRBT_i < 1.2$ and at the same time $Y_i < 1$ for the i-th country, the response Y_i is replaced by $Y_i + Z_i$, where $Z_1, \ldots, Z_n$ are i.i.d. following the uniform $U(-2,2)$ distribution. In this way, 14 countries are contaminated.*

(D) *Local contamination of the response for all the 140 countries; the response Y_i for every $i = 1, \dots, n$ is replaced by $Y_i + Z_i$, where $Z_1, \dots, Z_n$ are i.i.d. generated from N(0; 25).*

For each contamination scenario, the averaged values of b_{CRBT} obtained for contaminated data divided by b_{CRBT} for raw data are reported in Table 3. To assess the variability of these ratios, nonparametric bootstrap estimates of their standard deviations are presented as well; for the sake of computational feasibility, the contaminated data were generated only 100 times here and 100 bootstrap samples were considered for each contaminated dataset. Ideally, robust tools should have values of the ratio r close to 1 together with small values of the standard deviation.

Table 3 Results of the Computations of Various Regression Methods Evaluated in the Model (12)

		Contaminated data			
	Raw data	***Scenario A***	***Scenario B***	***Scenario C***	***Scenario D***
Estimator	b_{CRBT} (SD)	Ratio (SD)	Ratio (SD)	Ratio (SD)	Ratio (SD)
LS	7.94 (1.31)	0.85 (0.016)	1.00 (0.048)	1.03 (0.002)	1.00 (0.034)
MM	4.48 (1.15)	0.97 (0.008)	1.02 (0.122)	1.30 (0.013)	1.01 (0.152)
LTS	2.44 (2.06)	0.97 (0.022)	1.02 (0.157)	1.87 (0.300)	1.44 (0.621)
LWS	4.97 (1.13)	0.96 (0.011)	1.02 (0.061)	1.05 (0.003)	1.03 (0.063)
RQ(0.2)	4.10 (1.07)	0.89 (0.094)	1.02 (0.117)	1.42 (0.025)	1.14 (0.246)
RQ(0.4)	5.18 (1.24)	0.87 (0.069)	1.02 (0.093)	1.14 (0.017)	1.09 (0.133)
RQ(0.6)	7.93 (1.67)	0.91 (0.047)	0.99 (0.072)	1.10 (0.004)	1.02 (0.098)
RQ(0.8)	9.88 (1.91)	0.92 (0.038)	1.02 (0.077)	1.07 (0.004)	1.00 (0.080)
NCQR(0.2)	3.82 (1.23)	0.92 (0.103)	1.04 (0.126)	1.37 (0.058)	1.12 (0.267)
NCQR(0.4)	5.21 (1.27)	0.87 (0.070)	1.02 (0.096)	1.15 (0.018)	1.09 (0.139)
NCQR(0.6)	7.89 (1.73)	0.91 (0.047)	0.99 (0.074)	1.10 (0.004)	1.02 (0.102)
NCQR(0.8)	9.86 (2.02)	0.92 (0.038)	1.02 (0.078)	1.07 (0.004)	1.00 (0.082)
LTS($\tilde{c}_1$)	1.96 (0.87)	0.97 (0.020)	1.03 (0.176)	2.02 (0.308)	1.67 (1.357)
LTS($\tilde{c}_2$)	2.31 (1.22)	0.97 (0.022)	1.01 (0.163)	1.95 (0.263)	1.41 (0.942)
LTS($\tilde{c}_3$)	3.24 (2.19)	0.97 (0.019)	1.03 (0.159)	1.83 (0.255)	1.38 (0.578)
LTS($\tilde{c}_4$)	4.85 (2.32)	0.98 (0.018)	1.03 (0.153)	1.74 (0.249)	1.26 (0.429)
LWS($\tilde{c}_1$)	2.20 (0.64)	0.98 (0.017)	1.02 (0.109)	1.08 (0.004)	1.06 (0.124)
LWS($\tilde{c}_2$)	2.48 (0.91)	0.98 (0.013)	1.01 (0.094)	1.05 (0.004)	1.07 (0.085)
LWS($\tilde{c}_3$)	5.17 (1.15)	0.96 (0.012)	1.03 (0.061)	1.06 (0.003)	1.02 (0.061)
LWS($\tilde{c}_4$)	8.83 (1.37)	0.96 (0.012)	1.02 (0.075)	1.03 (0.004)	1.01 (0.053)

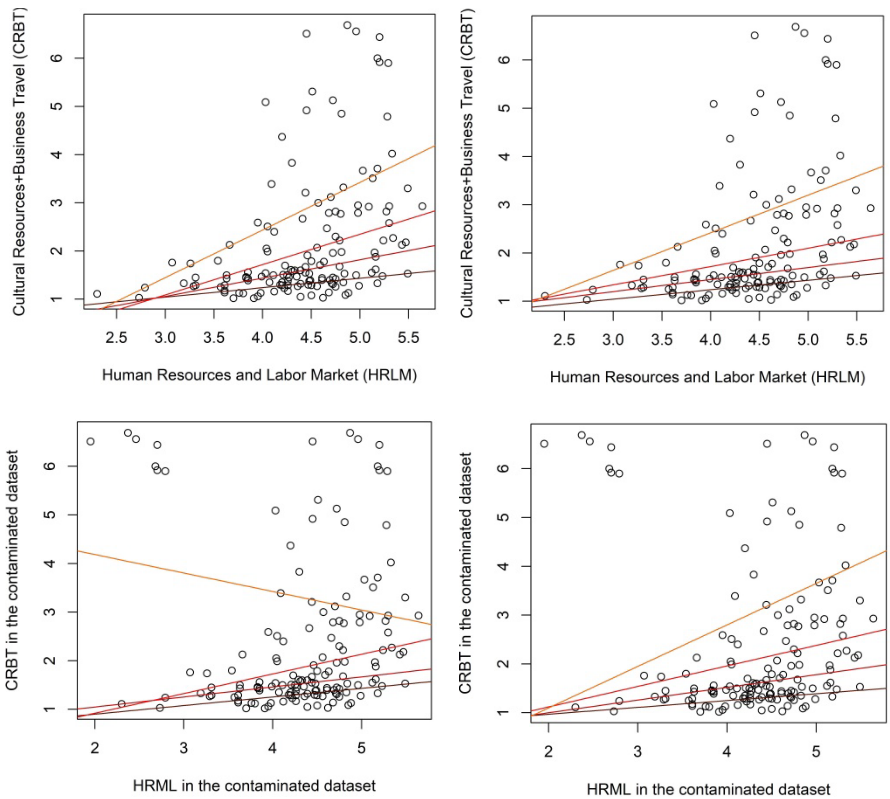
Notes: Values of $\tilde{c}_1$,…,$\tilde{c}_4$ Were Found by Algorithm 1. (I) For the Raw TTCI data, Estimates of β_{14} (Denoted Here as b_{CRBT}) Corresponding to the Effect of CRBT (Cultural Resources and Business Travel) are Reported, Together with Nonparametric Bootstrap Estimates of the Standard Deviation of each Estimate. (II) For 4 Contamination Scenarios, the Averaged Values of b_{CRBT} Obtained for Contaminated Data Divided by b_{CRBT} for Raw Data are Reported; These Ratios are Accompanied by Their Averaged Standard Deviations.

Contamination (A) by asymmetric outliers deteriorates the performance of regression quantiles, while the robust quantiles LWSQ and LTSQ remain resistant; the variability of LWSQ remains smaller compared to variability of all other quantiles. For all types of quantiles under (A), estimating the lowest quintile is more difficult (with a larger variability) compared to the top quintile. If the data are contaminated by outliers in a symmetric way (B), the estimation is not so difficult compared to the situation (A) and values of r do not change greatly. For scenario (C), standard regression quantiles are harmed by the artificially introduced influential points with leverage effect, while LWS-quantiles outperform all other quantiles in terms of both r and its variability. For local contamination (D), LWS-quantiles again outperform all other types of quantiles in terms of both *r* and its variability. For all 4 contamination scenarios jointly, we can say that the performance of NCQR is not much different from that of standard regression quantiles.

To give an additional argument in favor of LWSQ, we perform another contamination study denoted as scenario (E). Instead of (12), we consider the model with CRBT as the response of a single regressor HRLM.

(E) There are 6 artificial observations considered, which are obtained as shifted copies of the 6 countries with the largest value of CRBT. The 6 new observations shown in the top left corner of the bottom plots in Figure 3 are shifted by 2.5 horizontally with an unaltered value of CRBT.

Figure 3 Plots of CRBT against HRLM



Notes: Left: Regression Quintiles with the Parameter τ Equal to 0.2 (Bottom Line), 0.4, 0.6, and 0.8 (Top Line). Right: LWS-quintiles with Values of the Parameter c Automatically Obtained by Algorithm 1. Top Row: Raw Data. Bottom Row: Contaminated Data in Contamination Scenario (E).

The results of standard regression quantiles for contamination (E) shown in Figure 3 are very different from those obtained for raw data; the top regression quintile is heavily influenced by the leverage effect of the contamination, i.e. is focused on the minority trend with a negative slope, while the other quintiles correspond to the majority trend with a positive slope. LWS-quantiles work well for contamination (E) and remain to yield a robust solution, which is visually much similar to the figure obtained for raw data. To explain such good fit, LWS-quintiles are determined by the large majority of countries with a small value of CRBT here, and the outliers have weights exactly zero in the computation of the top LWS-quintile; naturally, also the points in the left bottom corner (such as Mauritania with the very smallest CRBT) have a small influence on LWS-quintiles. Thus, scenario (E) reveals the robustness of LWSQ and non-robustness of regression quantiles with respect to leverage points.

4. Conclusions

In this paper, novel robust regression quantiles (LWSQ) for the standard linear regression model (1) are proposed. The method is inspired by the LWS estimator, which has appealing properties (high efficiency, high robustness to outliers, and high local robustness to small changes of the data). While investigating statistical properties of LWSQ remains to be an open problem for future research, the performance of LWSQ is investigated here on a particular world tourism dataset. The results of LWSQ for the original data turn out to be meaningful, and remain to be resistant for data after an artificial contamination. From the point of view of small variability, they outperform the standard regression quantiles and the non-crossing (but also non-robust) regression quantiles of Sottile and Frumento (2021) on the presented data. LWS-quantiles also turned out empirically to outperform standard regression quantiles in robustness and stability of variable selection.

From the economic point of view, this paper brings a new insight into the tourism performance (at least based on the last available pre-COVID-19 numbers), which is modeled here by means of TTCI pillars on the level of individual countries across the whole world. The model turns out to have a large explanatory power and the presented unique analysis by regression quantile methodology allows an easy identification of the position of every individual country (Vašaničová et al., 2017) from the point of view of its exploiting the potential for tourism. Such differences between the current and potential tourist flow, which has been denoted as the latent tourism demand, have remained quite complicated to reveal until now (Eugenio-Martin and Cazorla-Artiles, 2020). Also the pillars of TTCI which are the main determinants of tourism performance, i.e. the key factors with the largest explanatory power, are identified here.

Particularly, the analysis allows to determine countries which are not effective in transforming their competitiveness to higher levels of tourist arrivals as those in the bottom quantiles. As a consequence, these countries should make investments to their infrastructure in order to enhance tourism growth (ideally if retaining principles of sustainable tourism). We can say that countries in higher quantiles can be used as reference points for the development of those lagging behind (Marti and Puertas, 2017). Comparing the quantiles of different countries and identifying groups (clusters) of countries in the same quantile may lead to international coordination of promoting tourism within such clusters. Such idea corresponds to the findings of Lyócsa et al. (2020), who revealed tourism activities of different countries to be interconnected and encouraged cooperative strategies for policymakers of different countries allowing to contribute to their joint growth of tourism performances.

The main limitation of the presented results consists in relying on the pillars of the TTCI index, which remains popular but has already been subject to criticism. Alternative versions of TTCI were proposed by Rodríguez-Díaz and Pulido-Fernández (2021), who proposed a sophisticated standardization of the weights (compared to using arbitrary weights of TTCI), or by Férnandez et al. (2022), who suggested to adapt TTCI to reflect the vulnerability of destinations to consequences of the COVID-19 pandemic. Future studies of tourist arrivals would profit from considering more detailed data on the level of time series or on the level of international arrivals from one country to another country, instead of the aggregated number of arrivals. Our study is also void of using any additional economic indexes, which also could contribute to explaining the number of international arrivals, such as the Global Competitiveness Index (GCI), the KOF Globalization Index, or the Global Knowledge Index (GKI).

APPENDIX

Proof of Theorem 2.1. It is convenient to consider the definition (8) of LWSQ. To prove scale equivariance, it will be useful to write

$$\arg\min_{b\in\mathbb{R}^{p+1}} \sum_{i=1}^{n} \psi\left(\frac{i-1/2}{n}\right)\left(cI\left[v_{(i)}(b)>0\right]+I\left[v_{(i)}(b)<0\right]\right)v_{(i)}^2(b).$$

with $v_i = kY_i - X_i^T b$ for $i = 1, \dots, n$, which may be further expressed as

$$\arg\min_{b\in\mathbb{R}^{p+1}} \sum_{i=1}^{n} \psi\left(\frac{i-1/2}{n}\right)\left(cI\left[u_{(i)}(b/k)>0\right]+I\left[u_{(i)}(b/k)<0\right]\right)\left(u_{(i)}(b/k)\right)^2.$$

Here, b/k plays the role of $b_{LWSQ}[X,Y]$, so that $b_{LWSQ}[X,kY] = kb_{LWSQ}[X,Y]$, in other words (9) is obtained for any $k > 0$. To prove regression equivariance, we may express $b_{LWSQ}[X, Y + Xt]$ again as (4) but this time with $v_i = Y_i + X_i^T t - X_i^T b$ for $i = 1, \dots, n$. Then, we obtain

$$\begin{aligned}\arg\min_{b\in\mathbb{R}^{p+1}} \sum_{i=1}^{n} \psi\left(\frac{R_i(b)-1/2}{n}\right)&(cI[Y_i - X_i^T(b-t)>0] \\ &+ I[Y_i - X_i^T(b-t)<0])\left(u_i(b)\right)^2,\end{aligned}$$

where $b - t$ corresponds to $b_{LWSQ}[X,Y]$, so that (10) is immediately obtained. Finally to prove affine equivariance, let us express $b_{LWSQ}[XA,Y]$ as (4) but here with $v_i = Y_i - X_i^T Ab$ for $i = 1, \dots, n.$ Now Ab may be interpreted as $b_{LWSQ}[X,Y]$, which immediately leads to obtaining (11).

REFERENCES

Alfons A, Croux C, Gelper S (2013): Sparse Least Trimmed Squares Regression for Analyzing High-Dimensional Large Data Sets. *Annals of Applied Statistics*, 7: pp. 226–248.

Almeida A, Garrod B (2017): Insights from Analysing Tourist Expenditure Using Quantile Regression. *Tourism Economics*, 23: pp. 1138–1145.

Arain H, Han L, Sharif A, Meo MS (2020): Investigating the Effect of Inbound Tourism on FDI: The Importance of Quantile Estimations. *Tourism Economics*, 26: pp. 682–703.

Assaf AG, Tsionas M (2020): Correcting for Endogeneity in Hospitality and Tourism Research. *International Journal of Contemporary Hospitality Management*, 32: pp. 2657–2675.

Bazargani RHZ, Kiliç H (2021): Tourism Competitiveness and Tourism Sector Performance: Empirical Insights from New Data. *Journal of Hospitality and Tourism Management*, 46: pp. 73–82.

Bondell HD, Reich BJ,Wang H (2010): Non-Crossing Quantile Regression Curve Estimation. *Biometrika*, 97: pp. 825–838.

Calderwood LU, Soshkin M (2019): The Travel & Tourism Competitiveness Report 2019: Travel and Tourism at a Tipping Point. World Economic Forum, Geneva [http://www3.weforum.org/docs/WEF_TTCR_2019.pdf].

Castillo E, Castillo C, Hadi AS, Mínguez R (2008): Duality and Local Sensitivity Analysis in Least Squares, Minimax, and Least Absolute Values Regressions. *Journal of Statistical Computation and Simulation*, 78: pp. 887–909.

Čížek P (2011): Semiparametrically Weighted Robust Estimation of Regression Models. *Computational Statistics & Data Analysis*, 55: pp. 774–788.

Das J, Dirienzo CE (2012): Tourism Competitiveness and the Role of Fractionalization. *International Journal of Tourism Research*, 14: pp. 285–297.

Davies L (1990): The Asymptotics of S-Estimators in the Linear Regression Model. *Annals of Statistics*, 18: pp. 1651–1675.

Davies PL, Gather U (2005): Breakdown and Groups. *Annals of Statistics*, 33: pp. 977–1035.

Eugenio-Martin JL, Cazorla-Artiles JM (2020): The Shares Method for Revealing Latent Tourism Demand. *Annals of Tourism Research*, 84: No. 102969.

Fernández JAS, Martínez JMG, Martín JMM (2022): An Analysis of the Competitiveness of the Tourism Industry in a Context of Economic Recovery Following the COVID19 Pandemic. *Technological Forecasting & Social Change*, 174: No. 121301.

Frumento P (2021): qrcm: Quantile Regression Coefficients Modelling. R package version 3.0 [https://CRAN.R-project.org/package=qrcm].

Greene WH (2018): *Econometric Analysis*. 8th edn. Pearson Education Limited, Harlow.

Gutenbrunner C, Jurečková J, Koenker R, Portnoy S (1993): Tests of Linear Hypotheses based on Regression Rank Scores. *Journal of Nonparametric Statistics*, 2: pp. 307–331.

Hallin M, Šiman M (2018): Multiple-output Quantile Regression. In Koenker R, Chernozhukov V, He X, Peng L (Eds): *Handbook of Quantile Regression*. Chapman & Hall/CRC Press, Boca Raton, pp. 185–207.

Hanafiah MH, Hemdi MA, Ahmad I (2016): Does Tourism Destination Competitiveness Lead to Performance? A Case of ASEAN Region. *Tourism: An International Interdisciplinary Journal*, 64: pp. 251–260.

Hastie T, Tibshirani R,Wainwright M (2015): *Statistical Learning with Sparsity: The Lasso and Generalizations*. CRC Press, Boca Raton.

Huber PJ, Ronchetti EM (2009): *Robust Statistics*. 2nd edn. Wiley, Hoboken.

Hubert M, Rousseeuw PJ, Van Aelst S (2008): High Breakdown Robust Multivariate Methods. *Statistical Science*, 23: pp. 92–119.

Ivanov S, Webster C (2013): Globalisation as a Driver of Destination Competitiveness. *Annals of Tourism Research*, 43: pp. 628–633.

Jeon JH (2020): Macro and Non-Macro Determinants of Korean Tourism Stock Performance: A Quantile Regression Approach. *Journal of Asian Finance, Economics and Business*, 7: pp. 149–156.

Jurečková J, Picek J, Schindler M (2019): *Robust Statistical Methods with R*. 2nd edn. CRC Press, Boca Raton.

Kalina J (2012): On Multivariate Methods in Robust Econometrics. *Prague Economic Papers*, 21: pp. 69–82.

Kalina J, Neoral A, Vidnerová P (2021): Effective Automatic Method Selection for Nonlinear Regression Modeling. *International Journal of Neural Systems*, 31: No. 2150020.

Kalina J, Tichavský J (2020): On Robust Estimation of Error Variance in (Highly) Robust Regression. *Measurement Science Review*, 20: pp. 6–14.

Kalina J, Vašaničová P, Litavcová E (2019): Regression Quantiles under Heteroscedasticity and Multicollinearity: Analysis of Travel and Tourism Competitiveness. *Ekonomický časopis/Journal of Economics*, 67: pp. 69–85.

Koenker R (2005): *Quantile Regression*. Cambridge University Press, Cambridge.

Koenker R (2017): Quantile Regression: 40 Years on. *Annual Review of Economics*, 9: pp. 155–176.

Krstic B, Jovanovic S, Jankovic-Milic V, Stanisic T (2016): Examination of Travel and Tourism Competitiveness Contribution to National Economy Competitiveness of Sub-Saharan Africa Countries. *Development Southern Africa*, 33: pp. 470–485.

Lyócsa Š, Vašaničová P, Litavcová E (2021): Quantile Dependence of Tourism Activity Between Southern European Countries. *Applied Economics Letters*, 27: pp. 206–212.

Manzo GG (2020): *Travel & Tourism Economic Impact*. World Travel & Tourism Council, London.

Maronna RA, Martin RD, Yohai VJ, Salibián-Barrera M (2018): *Robust Statistics: Theory and Methods*. 2nd edn. Hoboken, Wiley.

Marti L, Puertas R (2017): Determinants of Tourist Arrivals in European Mediterranean Countries: Analysis of Competitiveness. *European Journal of Tourism Research*, 15: pp. 131–142.

Ouchen A, Montargot N (2021): Non-Spatial and Spatial Econometric Analysis of Tourism Demand in a Panel of Countries around the World. *Spatial Economic Analysis*. In press.

Pérez-Rodríguez JV, Ledesma-Rodríguez F (2021): Unconditional Quantile Regression and Tourism Expenditure: The Case of the Canary Islands. *Tourism Economics*, 27: pp. 626–648.

Porter ME (1990): *The Competitive Advantage of Nations*. Harvard Business Review, 68: pp. 73–93.

Rodríguez-Díaz B, Pulido-Fernández JI (2021): Analysis of the Worth of the Weights in a New Travel and Tourism Competitiveness Index. *Journal of Travel Research*, 60: pp. 267–280.

Roser M (2020): https://ourworldindata.org/tourism.

Rousseeuw PJ, Van Driessen K (2006): Computing LTS Regression for Large Data Sets. *Data Mining and Knowledge Discovery*, 12: pp. 29–45.

Rudkin S, Sharma A (2017): Enhancing Understanding of Tourist Spending Using Unconditional Quantile Regression. *Annals of Tourism Research*, 66: pp. 188–191.

Sottile G, Frumento P (2021): Parametric Estimation of Non-Crossing Quantile Functions. *Statistical Modelling*. In press, doi: 10.1177/1471082X211036517.

Vašaničová P, Litavcová E, Jenčová S, Košíková M (2017): Dependencies Between Travel and Tourism Competitiveness Subindexes: The Robust Quantile Regression Approach. *The 11th International Days of Statistics and Economics MSED 2017*, Melandrium, Slaný, pp. 1729–1739.

Víšek JÁ (2011): Consistency of the Least Weighted Squares Under Heteroscedasticity. *Kybernetika*, 47: pp. 179–206.

White H (1980): A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity. *Econometrica*, 48: pp. 817–838.

Yohai VJ (1987): High Breakdown-Point and High Efficiency Robust Estimates for Regression. *Annals of Statistics*, 15: pp. 642–656.