Insurance: Mathematics and Economics 99 (2021) 1–8

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

Joint generalized quantile and conditional tail expectation regression for insurance risk analysis[☆]

Montserrat Guillen[a,*], Lluís Bermúdez[b], Albert Pitarque[a]

[a] Dept. Econometria, Estadística i Economia Aplicada, Riskcenter-IREA, Universitat de Barcelona, Av. Diagonal, 690, E08034 Barcelona, Spain
[b] Dept. Matemàtica Econòmica, Financera i Actuarial, Riskcenter-IREA, Universitat de Barcelona, Av. Diagonal, 690, E08034 Barcelona, Spain

ARTICLE INFO

Article history:
Received June 2020
Received in revised form January 2021
Accepted 12 March 2021
Available online 26 March 2021

Keywords:
Value at risk
Predictive models
Telematics
Motor insurance
Quantile regression

ABSTRACT

Based on recent developments in joint regression models for quantile and expected shortfall, this paper seeks to develop models to analyse the risk in the right tail of the distribution of non-negative dependent random variables. We propose an algorithm to estimate conditional tail expectation regressions, introducing generalized risk regression models with link functions that are similar to those in generalized linear models. To preserve the natural ordering of risk measures conditional on a set of covariates, we add extra non-negative terms to the quantile regression. A case using telematics data in motor insurance illustrates the practical implementation of predictive risk models and their potential usefulness in actuarial analysis.

1. Introduction

Our predictive modelling focuses on Value at Risk (*VaR*) and the Conditional Tail Expectation (*CTE*). While classical linear regression finds the effects of covariates on the mean of a response variable via a linear predictor, quantile regression focuses on the *VaR* of the response and *CTE* regression links covariates to the conditional average responses beyond a quantile. We will consider the usual case in insurance, where risk concentrates on positive losses. In finance, where risk focuses on negative returns, the usual risk measure is Expected Shortfall (ES) rather than *CTE*, because ES looks at the lack of resources needed to cope with unexpected negative outcomes. We will call these models *risk regressions*, in general.

Risk regressions have not been popular in insurance because of the technical difficulty of fitting the models. However, they might be extremely useful to identify factors that influence the worst case outcomes. There are many examples of loss random variables that are asymmetric and right skewed, where risk is located at higher quantiles. We can think of accident severity as a primary example.

The first complication in risk regressions, recently studied in the financial literature, lies in establishing a suitable score function, similar to the sum of squared residuals in the least squares method for linear regression. However, such a score function is not always possible to find. A risk measure is called *elicitable* (Gneiting, 2011) if there exists a scoring function such that the expected score under a distribution takes its unique minimum at the risk value of the distribution. Wang and Ziegel (2015) and Kou and Peng (2016) have shown that distortion risk measures are rarely elicitable.

Koenker and Bassett (1978) provided a score function for quantile regression (*VaR* regression) and initiated a methodology that has become increasingly popular over the years (see, Koenker, 2017). Gneiting (2011) showed that the ES is not a 1-elicitable risk measure (a risk measure is 1-elicitable when its corresponding loss function does not depend on other risk measures), which means that there is no score function that can be minimized to obtain ES (or *CTE*) alone. Therefore, it is not possible to estimate ES regression in the same way as quantile regression. However, Fissler and Ziegel (2016) have established the joint elicitability of *VaR* and ES risk measures for continuously distributed random variables that take negative values and capture left-side risks, and they have shown that the corresponding joint score function is not unique. One particular case is the score function proposed by Acerbi and Szekely (2014). Dimitriadis and Bayer (2019) have analysed possible choices for the family of score functions put forward by Fissler and Ziegel (2016), finding that they have the property of positive homogeneity such that linear rescaling of the input variable does not alter the ranking of losses.

All of the above works have dealt with left-side risks. However, in insurance and actuarial applications, economic losses

☆ We would like to thank Fundación BBVA projects in the area of Big Data, the ICREA, Spain Academia programme, AGAUR, Spain and the Spanish Ministry of Science and Innovation grant PID2019-105986GB-C21. We thank the comments received by two anonymous reviewers to an earlier version.

* Corresponding author.
E-mail address: mguillen@ub.edu (M. Guillen).

https://doi.org/10.1016/j.insmatheco.2021.03.006

are defined as positive. As a result, insurance risk analysis concentrates on large, positive values, which are naturally located on the right side of the distribution. We will therefore work with positive, right-side risks. This implies a change of sign that is sometimes confusing when drawing on sources that use the other convention. We will distinguish *CTE* regression, for positive right tails, from ES regression, for negative left tails, and we will convey all of our exposition in terms of the risk analysis of random variables defined on the positive real semi-axis. Many recent results for ES regression and forecasting can be easily, but cautiously rewritten for positive values by changing the sign of the response variable and establishing low quantile levels, such as 5% instead of 95%. This seems like a sign convention, however it turns out to be quite misleading to practitioners when confronted with the implementation.

The second complication is that the existing models (see, e.g. Acerbi and Szekely, 2014; Dimitriadis and Bayer, 2019) may produce negative predictions, or even a predicted *CTE* incompatible with the predicted *VaR* (i.e. *CTE* should be larger than the *VaR*, for positive responses). Our contribution solves these two issues by proposing a joint generalized additive quantile and *CTE* regression model. As far as we know, only Taylor (2019) addresses these issues applicable to time-series context. However, his contribution presents some limitations in our context (i.e. the dynamics of *VaR* may not be the same as the dynamics of *CTE*).

In sum, with the above references as our starting point, we provide some insights that are especially relevant for insurance applications. First, we adapt the existing literature on joint modelling to positive, right-side risks. Second, we propose a new joint modelling that guarantees non-negative predictions and the $CTE > VaR$ restriction, while keeping their individual dynamics, by using an exponential link function and an additive form, respectively. Third, in Section 3, we adapt the two-step estimation procedure used in the existing literature to our approach. In Section 4, a case study on a telematic data set is conducted in order to show the advantages of using our proposal and, simultaneously, the interest in empirical applications such as this one and its potential benefits to analyse insurance risks. Finally, Section 5 presents our concluding remarks.

2. Predictive models for *VaR* and *CTE*

Before delving into our regression modelling framework, let us formally define *VaR* and *CTE*. Value-at-risk at level τ, $\tau \in (0, 1)$, also known as the $(\tau \times 100)$th quantile or τ-quantile, is defined as follows:

$$VaR_\tau(Y) = inf\left\{y \in \mathbb{R}^+ : F_Y(y) > \tau\right\},$$

where Y is continuously distributed and has a finite mean, and $F_Y(y)$ corresponds to its cumulative distribution function. *VaR* does not consider observations beyond the quantile, but it is one of the most popular measures to analyse risk due to its simplicity and ease of understanding.

To account for observations in the tail, *CTE* averages the extreme values of the distribution function. This risk measure in continuous variables is also known as Tail Value at Risk (*TVaR*) or Conditional Value at Risk (*CVaR*) and it is the mean of the values that exceed the *VaR*. *CTE* is defined as follows:

$$CTE_\tau(Y) = \mathbb{E}\left[Y|Y > VaR_\tau\right].$$

Definition. A risk measure $\varphi(Y)$ of a random variable Y is elicitable when it minimizes the expected value of a scoring function, $S(\varphi, Y)$. So, an estimator of an elicitable $\varphi(Y)$ results from $\hat{\varphi} = \arg\min_\phi \mathbb{E}[S(\varphi,\ Y)]$.

In practice, for a sample $Y_1, \ldots, Y_n$ of size n, an estimator $\hat{\varphi}$ can be found minimizing $\sum_{i=1}^n S(\varphi,\ Y_i)$. While $VaR_\tau(Y)$ is elicitable, $CTE_\tau(Y)$ is not, essentially because $VaR_\tau(Y)$ is needed to define $CTE_\tau(Y)$.

2.1. Quantile regression

The starting point of our work is quantile regression (Koenker and Bassett, 1978). Even though quantile regression is a relatively new methodology, there is an increasing number of applications in a wide variety of fields (see, Uribe and Guillen, 2020, for an overview of recent methods and R implementation) .

Quantile regression is an extension of linear regression that is especially interesting when the response variable has asymmetry, for instance, when there is a substantial difference between the conditional mean and the conditional median. As is widely known, the median is robust to the presence of outliers, while the mean is not. Risk analysis actually focuses on quantile regression for large τ-quantiles.

Unlike linear regression, which estimates the effect of each explanatory variable on the mean of the response variable, quantile regression establishes the effect of explanatory variables on the quantile of the response variable. We can specify the τ-quantile regression model at level $\tau \in (0, 1)$ as follows:

$$Y_i = \beta_0^\tau + \beta_1^\tau X_{1i} + \beta_2^\tau X_{2i} + \cdots + \beta_k^\tau X_{ki} + \varepsilon_i^\tau,$$

where Y_i is the response variable for the ith individual ($i = 1, \ldots, n$), X_{ji} represents the value of the ith observation of explanatory variable j ($j = 1, \ldots, k$), β^τ is the vector of unknown parameters, and we assume that $VaR_\tau(\varepsilon_i^\tau) = 0$. There is no assumption made about the specific parametric distribution of Y_i, and this is the reason why quantile regression is sometimes called *semiparametric*.

Alternatively, we can write quantile regression as a link between the τ-quantile of Y_i and a linear combination of the regressors, i.e. the linear predictor:

$$VaR_\tau(Y_i|X_{1i}, \ldots, X_{ki}) = \beta_0^\tau + \beta_1^\tau X_{1i} + \beta_2^\tau X_{2i} + \cdots + \beta_k^\tau X_{ki}. \quad (1)$$

In short, $VaR_\tau(Y_i|X_i) = X_i'\beta^\tau$. Koenker and Bassett (1978) proposed an optimization framework to fit quantile regressions. Basically, the parameter estimates can be obtained as the solution of the following optimization problem (see, Koenker and Bassett, 1982; Koenker and Machado, 1999):

$$\hat{\beta}^\tau = \arg\min_{\beta^\tau} \sum_{i=1}^n \rho_q^\tau(Y_i - X_i'\beta^\tau), \quad (2)$$

where ρ_q^τ represents the score function of the τ-quantile, which is equal to $\tau(Y_i - X_i'\beta^\tau)$ when $(Y_i - X_i'\beta^\tau)$ is greater than or equal to 0, and $(\tau - 1)(Y_i - X_i'\beta^\tau)$ otherwise. The standard error of the estimated coefficients can be calculated following the bootstrap method (see, Chernick, 2011; Hesterberg, 2011).

With no loss of generality, we may introduce link function $F^\upsilon(\cdot)$ in (1). So, as opposed to quantile *linear* regression, or simply quantile regression, we can define the *generalized quantile regression* as:

$$VaR_\tau(Y_i|X_i) = F^\upsilon(X_i'\beta^\tau),$$

where $F^\upsilon(\cdot)$ is monotone and twice continuously differentiable to meet GLM assumptions (McCullagh and Nelder, 1989). For example, we can choose $F^\upsilon(z) = \exp(z)$ to guarantee that the predictions are positive. This is exactly the generalized quantile regression that is later implemented in our case study, using

$$VaR_\tau(Y_i|X_i) = \exp(X_i'\beta^\tau). \quad (3)$$

In a simultaneous and independent work, Dimitriadis and Schnaitmann (2019) also introduce link functions.

2.2. CTE regression specification

If it is possible to establish a relationship between the explanatory variables and *VaR*, it should also be possible to do so with other risk measures. The specification of a conditional tail expectation *linear* regression is:

$$CTE_\tau(Y_i|X_{1i},\ldots,X_{ki}) = \gamma_0^\tau + \gamma_1^\tau X_{1i} + \gamma_2^\tau X_{2i} + \cdots + \gamma_k^\tau X_{ki}, \quad (4)$$

where γ^τ corresponds to the parameters for the effects of the explanatory variables on the expectation above the conditional quantile $VaR_\tau(Y_i|X_i)$. Equivalently, we can use an error term whose CTE_τ equals zero. To ease notation, we write $CTE_\tau(Y_i|X_i) = X_i'\gamma^\tau$ and we assume that (1) and (4) have the same regressors. However, we could define a set X_i^q for (1) and another possibly overlapping set X_i^e (4) as in Dimitriadis and Bayer (2019).

As mentioned above, there is no parallel to expression (2) for *VaR* regression to estimate the parameters of *CTE* regression in (4). Dimitriadis and Bayer (2019) propose a two-step procedure to estimate jointly the parameters in (1) and (4). We will discuss estimation procedures in the next section.

With no loss of generality, we may also introduce a link function $F^e(\cdot)$ in (4). We only need the same monotonicity and regularity conditions as before. So, the generalized *CTE* regression is denoted as $CTE_\tau(Y_i|X_i) = F^e(X_i'\gamma^\tau)$. The generalized *CTE* regression that will be implemented subsequently in our case study is:

$$CTE_\tau(Y_i|X_i) = \exp(X_i'\gamma^\tau). \quad (5)$$

In Dimitriadis and Bayer (2019), link functions were not introduced.

2.3. A new proposal to specify joint generalized VaR and CTE regression models

We have now introduced the link function in the *VaR* and *CTE* models. This is the reason why we include the word “generalized" in the name of our models. The choice of the link function has to do with the domain of the response variable, which is non-negative for insurance risk analysis. Unlike in GLM, we do not have to specify a link between the canonical parameter of the exponential distribution of the dependent variable and the linear predictor. In risk regressions, our choice of a link function is guided by the need to provide predictions that stay in the domain of our variable of interest. In addition, in order to ensure that $CTE > VaR$, we will consider the following additive specification:

$$CTE_\tau(Y_i|X_i) = F^v(X_i'\hat{\beta}^\tau) + F^e(X_i'\eta^\tau), \quad (6)$$

where $\hat{\beta}^\tau$ is the corresponding term in the generalized quantile regression and η^τ is the vector of unknown parameters that guide the additive term in the *CTE* regression. Our proposal is to choose exponential links in order to guarantee that predictions are positive. If the variables are the same and we choose the identity link for F^v and F^e in (6), then there is an identification issue for the regression coefficients.

This specification, named as *joint generalized additive VaR and CTE regression*, guarantees that $CTE_\tau(Y_i|X_i) \leq VaR_\tau(Y_i|X_i)$ for all given $\tau \in (0,1)$ and that the predictions of *VaR* and *CTE*, conditional on X_i, are always positive. In addition, the additive specification proposed in (6) provides a nice interpretation of the parameters in the *CTE* part. These parameters explain a conditional mean expectation above the quantile.

One of the earliest attempts to introduce a connection between *VaR* and ES and the predictors appears in a recent study by Taylor (2019) in the context of time-series. His main proposal is to set:

$$CTE_\tau(Y_i|X_i) = VaR_\tau(Y_i|X_i)(1 + X_i'\gamma^\theta)$$

to force *CTE* to exceed *VaR* by creating a constant gap. He recognizes himself that this “expression is rather restrictive, as the dynamics of *VaR* may not be the same as the dynamics of ES”. On the contrary, our additive specification keeps the individual dynamics of *VaR* and *CTE*.

Taylor (2019) also presents an alternative formulation for ES using an autoregressive expression, which essentially smoothes the magnitude of exceedances beyond the quantile in order to avoid crossing. However, this autoregressive correction is only applicable to time-series context.

3. Estimation procedure

As mentioned before, the parameter estimates in quantile regression are obtained via the optimization problem as in (2). We will not reproduce the details on how to obtain the score function because this has already been developed extensively in Koenker and Bassett (1982).

Nadarajah et al. (2014) reviewed the estimation methods for *CTE* in the one response variable case, but they are not suitable for the inclusion of regressors. Unfortunately, there is no stand-alone score function to find the parameter estimates of a *CTE* regression. However, Acerbi and Szekely (2014) proposed a way to obtain *VaR* and *CTE* together using a score function that relates both risk measures.

Fissler and Ziegel (2016) showed that there are infinitely many score functions to achieve the joint elicitability of *VaR* and *CTE*, but they did not introduce regressors. In order to estimate the effects of the explanatory variables on *CTE*, we take the score function proposed by Acerbi and Szekely (2014) as a starting point. Dimitriadis and Bayer (2019), based on the work by Fissler and Ziegel (2016), conducted a simulation study in which they showed that some particular choices in the score function might have better small-sample properties than others. Dimitriadis and Bayer (2019) created an R package, `esreg`, which can be used to fit linear quantile and ES regressions in (1) and (4). To fit risk regression on large positive values, the implementation needs to be adapted: the sign of the dependent variable, the level, $(1-\tau)$, and the sign of the resulting parameters have to be reversed.

In addition to the problem of the joint score function to be minimized, another problem arises in practice when fitting *VaR* and *CTE* regressions: numerical instability, i.e. the fact that local minima may be found. This is the reason why Dimitriadis and Bayer (2019) recommend iterative local metaheuristics inspired by Lourenço et al. (2003). The issue is that this optimization is stochastic, because there is a small random noise alteration of the solution to refine the search for a minimum. So, to obtain the same results, one should always remember to use the same seed in the random number generation.

In the following two subsections, we adapt the existing literature on estimation methods for ES regressions to positive, right-side risks.

3.1. Score minimization for VaR and CTE regression

First, we consider the joint score function established by Fissler and Ziegel (2016) in the one response variable setting. In our notation we have positive outcomes and a large τ, for example $\tau = 0.90$, whereas in their original work, the returns were negative and the focus was on low τ levels.

Based on the results obtained by Fissler and Ziegel (2016), the general score function for *VaR* and *CTE* regressions (1) and (4) for a non-negative Y_i and linear predictors $X_i'\beta^\tau$ and $X_i'\gamma^\tau$ is:

$$\rho(Y_i, X_i, \beta^\tau, \gamma^\tau) = I(Y_i \geq X_i'\beta^\tau)\left[G_1(-X_i'\beta^\tau) - G_1(-Y_i)\right] + G_2(-X_i'\gamma^\tau)\left[X_i'\beta^\tau - X_i'\gamma^\tau - \frac{(Y_i - X_i'\beta^\tau)}{1-\tau} I(Y_i \geq X_i'\beta^\tau)\right] - \mathcal{G}_2(-X_i'\gamma^\tau) + a(-Y_i). \quad (7)$$

where G_2 is the first derivative of $\mathcal{G}_2$. Functions $G_1(\cdot)$ and $\mathcal{G}_2$ must satisfy some regularity conditions. Also, $a(\cdot)$ can be eliminated in the optimization procedure, but it should be carefully selected to guarantee that $\rho(Y_i, X_i, \hat{\beta}^\tau, \hat{\gamma}^\tau) > 0$ for the goodness-of-fit calculation (see, Koenker and Machado, 1999). A common choice is $a(z) = (1-\tau)G_1(z) + \mathcal{G}_2(z)$ (see, Dimitriadis and Bayer, 2019).

To obtain joint estimates from (7), the following optimization problem needs to be solved for a sample (Y_i, X_i), $i = 1, \ldots, n$:

$$(\widehat{\beta^\tau}, \widehat{\gamma^\tau}) = \arg\min_{\gamma^\tau \beta^\tau} \sum_{i=1}^{n} \rho(Y_i, X_i, \beta^\tau, \gamma^\tau). \quad (8)$$

The proposal put forward by Acerbi and Szekely (2014) is equivalent to setting $G_1(z) = (-Wz^2/2)$ and $\mathcal{G}_2(z) = (1-\tau)z^2/2$, where W is a constant so that $W\,VaR > CTE$. This guarantees the required regularity conditions, namely that $G_2(\eta)v/(1-\tau) + G_1(v)$ is a strictly increasing function of v, a VaR_τ, and η is its corresponding CTE_τ. If the model does not procure this restriction by specification, the estimation algorithm would find local minima. But the choice of W is unclear. Dimitriadis and Bayer (2019) suggest either $G_1(z) = 0$ or $G_1(z) = z$, like Fissler and Ziegel (2016), and they also propose five options for $\mathcal{G}_2(\cdot)$. They suggest to fix W large enough such that $W\,VaR > CTE$. Some of the choices of G_1 and $\mathcal{G}_2$ were unstable in our implementation and no standard error estimates could be obtained. In empirical applications like the one presented in our paper, one can either try to choose several $W > 1$ and seeing that the results are stable or one can think of a sensible W.

An indirect estimator for *CTE* regression is also presented in Dimitriadis and Bayer (2019), where it is called the *oracle* estimator of *CTE* regression. In our notation, we obtain estimates of β^τvia the quantile regression score in (2), $\hat{\beta}_O^\tau$, and then minimize the sum of squares of conditional residuals $(Y_i - X_i\hat{\gamma}_O^\tau)$ only for those observations that satisfy $Y_i > X_i\hat{\beta}_O^0$. However, this procedure is not recommended for small samples or extreme quantiles, due to the small number of observations beyond the quantile. We will denote the oracle estimator as $(\hat{\beta}_O^\tau, \hat{\gamma}_O^\tau)$.

3.2. Two-step procedure for linear CTE regression

Following Dimitriadis and Bayer (2019) and using our notation, we propose to solve (8) in a two-step process. First, we estimate β^τ via the quantile regression score in (2) as $\hat{\beta}_O^\tau$ and then we find $\widehat{\gamma^\tau}$:

$$\widehat{\gamma^\tau} = \arg\min_{\gamma^\tau} \sum_{i=1}^{n} \rho_{AS}(Y_i, X_i, \hat{\beta}_O^\tau, \gamma^\tau), \quad (9)$$

where the score function is taken from Acerbi and Szekely (2014), and we follow our positive sign convention:

$$\rho_{AS}(Y_i, X_i, \hat{\beta}_O^\tau, \gamma^\tau) = (1-\tau)\left(\frac{(X_i'\gamma^\tau)^2}{2} + W\frac{(X_i'\hat{\beta}_O^\tau)^2}{2} - X_i\gamma^\tau X_i'\hat{\beta}_O^\tau\right) + I(Y_i \geq X_i'\hat{\beta}_O^\tau)\left(-X_i'\gamma^\tau(Y_i - X_i'\hat{\beta}_O^\tau) + W\frac{Y_i^2 - (X_i'\hat{\beta}_O^\tau)^2}{2}\right) + (1-\tau)(W-1)Y_i^2/2, \quad (10)$$

where $I(Y_i \geq X_i\hat{\beta}_O^\tau)$ equals 1 if $Y_i \geq X_i\hat{\beta}_O^\tau$ and equals 0 otherwise. W is a fixed constant that is selected as before, but this has no impact on the minimization.

In our second step, $\widehat{\gamma^\tau}$ is fixed and the minimization of (9) is on β^τ to refine the quantile regression estimate part. However, this should be done carefully to avoid numerical instability, for instance, by using partial gradient descent.

Standard errors for the linear *CTE* regression can be found via bootstrap or with the asymptotic approximation provided by Dimitriadis and Bayer (2019).

Following Theorem 2.6 and the notation in Dimitriadis and Bayer (2019), we can approximate the variance and covariance matrix of the estimator for the linear *CTE* model as follows:

$$\Lambda_{22}^{-1} C_{22} \Lambda_{22}^{-1},$$

$$\Lambda_{22} = \mathbb{E}(X_i X_i' G_2^{(1)}(-X_i'\gamma^\tau)),\ G_2^{(1)} \text{ is the derivative of } G_2,$$

$$C_{22} = X_i G_2^{(1)}(-X_i'\gamma^\tau)^2 \mathbb{V}\left(X_i'\gamma^\tau - X_i'\beta^\tau - (1-\tau)^{-1}(Y_i - X_i'\beta^\tau)I(Y_i - X_i'\beta^\tau)\right)X_i'.$$

Then C_{22}, the asymptotic variance and covariance term of the covariance matrix for the estimator of the *CTE* model using (10) in the linear case, is approximated as:

$$\sum_{i=1}^{n} X_i \mathbb{V}\left(X'\gamma^\tau - X'\beta^\tau - (1-\tau)^{-1}(Y - X'\beta^\tau)I(Y - X'\beta^\tau)\right)X_i'.$$

The scalar term, $\mathbb{V}\left(X'\gamma^\tau - X'\beta^\tau - (1-\tau)^{-1}(Y - X'\beta^\tau)I(Y - X'\beta^\tau)\right)$, can be approximated as:

$$\left(X_i'\gamma^\tau - X_i'\beta^\tau\right)^2 - (1-\tau)^{-2}\mathbb{V}(Y_i - X_i'\beta^\tau)I(Y_i - X_i'\beta^\tau).$$

3.3. Two-step procedure for joint generalized additive VaR and CTE regression

Finally, we need to discuss how to adapt the above methodology to our proposed model for *VaR* and *CTE* regression.

In all the previous settings, an identity link with the linear predictor has been assumed. When we replace $X_i'\beta^\tau$ and $X_i'\gamma^\tau$ by the generalized terms using monotone transformations $F^v(\cdot)$ and $F^e(\cdot)$ in (7) and (10), then we obtain the new score functions to be minimized. The asymptotic statistical theory for the linear case is no longer valid for generalized specifications. In the generalized case, we propose a bootstrap method (Efron and Tibshirani, 1994). To obtain the bootstrap estimates, B samples are generated, that is, for each $b = 1, \ldots, B$, a resample of the original data (Y_i, X_i) is considered for all $i = 1, \ldots, n$. Then the bootstrapped parameter estimate is the average of the estimates obtained in the replication process and the bootstrapped covariance matrix is given by the sample covariance over all bootstrapped parameter estimates.

4. Case study: Predicting the risk of driving over the speed limit

An increasing number of companies are starting to work with telematics data in order to fit a better price for motor insurance by analysing driving patterns. For this study, we used a database containing information about 9,618 car drivers aged between

Table 1
Definition of variables in the telematics data set for 2010.

Variable	Description[a]
Speed_km[b]	Total number of kilometres driven over the speed limit
lnKm	Logarithm of the total number of kilometres driven
P_urban	Percentage of kilometres driven in urban areas
P_night	Percentage of kilometres driven at night
Age	Age of the driver
Male	Gender of the driver (1 = male, 0 = female)

[a]Distances driven are measured over one year.
[b]P_speed is the proportion (percentage) of total kilometres driven above the speed limit. P_speed = 100 × Speed_km/exp(lnKm).

Table 2
Descriptive analysis of the continuous variables in the telematics data set for 2010 ($n = 9618$).

	Mean	Median	Min.	Max.	Std. Dev.	Skewness
Speed_km	1398.21	689.23	0.00	23 500.19	1995.37	3.64
lnKm	9.27	9.37	−0.37	10.96	0.75	−1.87
P_urban	26.29	23.39	0.00	100.00	14.18	1.03
P_night	7.02	5.31	0.00	78.56	6.13	1.68
Age	24.78	24.63	18.11	35.00	2.82	0.11

18 and 35 years in 2010. The data contain information on the distance driven over one year, the type of roads, the time of day and the distance driven above the posted speed limit. The definitions of the variables appear in Table 1. The data have been used in previous studies together with claims information. Boucher et al. (2017) have analysed the simultaneous effect of the distance travelled and exposure time on the risk of accident by using Generalized Additive Models (GAM), while Ayuso et al. (2016) have compared the driving patterns between male and female drivers and Guillen et al. (2019) have proposed new methods to calculate the price of motor insurance. Pitarque et al. (2019) have used quantile regression to analyse the risk of having an accident and Pérez-Marín et al. (2019) have analysed speedy driving.

Our variable of interest is the total number of kilometres driven over the speed limit, Speed_km, which is highly positive skewed. Note that for the covariate lnKm we have used the log-transformed variable, which is a standard transformation to consider exposure to risk in insurance data. This makes interpretation of models with exponential link much more straightforward. The descriptive statistics appear in Table 2. There are 4873 male and 4741 female drivers in the sample. Our objective is to show the pitfalls of existing methods and the advantage of our proposal in an illustrative example.

4.1. Results for a bivariate analysis

First, we present a simple model with one covariate. We model the percentage of kilometres driven above the speed limit as a function of the percentage driving in urban areas. So, our initial predictive model for risk establishes a linear relationship between the percentage of total distance driven above the legal speed limit, P_speed, computed as (Speed_km × 100)/exp(lnKm), and the percentage driven in urban areas, P_urban. A simple linear regression (the details are omitted) finds a negative relationship between P_speed and P_urban, since the slope equals −0.178 (p-value < 0.001), which means that the higher the proportion of driving in urban areas, the lower the proportion of driving above the speed limit. This was expected, because urban areas tend to be more congested than non-urban areas and the possibility of exceeding the speed limit is therefore reduced by traffic. However, we also expect the slope and the intercept to change when looking at the median regression and quantiles with

Table 3
Model results for the percentage of distance driven above the speed limit, at quantile levels $\tau = 0.50, 0.75, 0.90$ and 0.95, as a function of the percentage of urban driving. Identity link (upper) and exponential link (lower). Standard errors in parenthesis.

$VaR_\tau(P\_speed_i|P\_urban_i) = \beta_0^\tau + \beta_1^\tau P\_urban_i$
$CTE_\tau(P\_speed_i|P\_urban_i) = \gamma_0^\tau + \gamma_1^\tau P\_urban_i$

	τ 0.5	0.75	0.90	0.95
$\hat\beta_0$	9.322***	18.329***	29.793***	37.334 ***
	(0.116)	(0.146)	(0.306)	(0.382)
$\hat\beta_1$	−0.107***	−0.221***	−0.353***	−0.434***
	(0.002)	(0.004)	(0.007)	(0.009)
$\hat\gamma_0$	22.822***	31.538***	41.549***	47.472***
	(0.226)	(0.295)	(0.417)	(0.501)
$\hat\gamma_1$	−0.300***	−0.413***	−0.522***	−0.565***
	(0.006)	(0.008)	(0.012)	(0.015)
Goodness-of-fit (R^2)	0.003	0.022	0.088	0.189
Score × 10^3	44,711.26	23,147.33	9167.40	4229.30
$Score_0$ × 10^3	44,867.09	23,669.75	10,055.20	5217.83

$VaR_\tau(P\_speed_i|P\_urban_i) = \exp(\beta_0^\tau + \beta_1^\tau P\_urban_i)$
$CTE_\tau(P\_speed_i|P\_urban_i) = \exp(\gamma_0^\tau + \gamma_1^\tau P\_urban_i)$

	τ 0.5	0.75	0.90	0.95
$\hat\beta_0$	2.412***	3.109***	3.580***	3.798***
	(0.018)	(0.015)	(0.014)	(0.015)
$\hat\beta_1$	−0.022***	−0.024***	−0.023***	−0.023***
	(0.001)	(0.001)	(0.001)	(0.001)
$\hat\gamma_0$	3.310***	3.639***	3.903***	4.028***
	(0.015)	(0.015)	(0.015)	(0.016)
$\hat\gamma_1$	−0.025***	−0.026***	−0.025***	−0.023***
	(0.001)	(0.001)	(0.001)	(0.001)
Goodness-of-fit (R^2)	0.012	0.023	0.087	0.155
Score × 10^3	3601.57	1657.90	505.68	176.99
$Score_0$ × 10^3	3645.02	1696.77	554.16	209.46

$Score_0$ is the value of the score function in a model with only intercepts.
p-value <1% ***, <5% ** and <10% *.

$\tau > 0.5$. Table 3 shows the results for the identity link corresponding to model (1) for *VaR* (linear quantile regression) and to model (4) for *CTE* (linear *CTE* regression) and for the exponential link corresponding to model (3) for *VaR* (generalized quantile regression) and model (5) for *CTE* (generalized *CTE* regression). *CTE* is elicitable when *VaR* is known. So, our approach is first to fit a generalized quantile regression and then we fit the *CTE* regression. The parameter estimates together with their standard errors have been found using our estimation approach.[1]

Hereinafter, we have used the R^2 proposed by Koenker and Machado (1999): $R^2 = 1 - \rho(Y_i, X_i, \hat\beta^\tau, \hat\gamma^\tau)/\rho(Y_i, X_i, \hat\beta_0^\tau, \hat\gamma_0^\tau)$.

As Fig. 1 (left) shows, the 90th-quantile regression finds a linear relationship between the exogenous variable and the response. In the left plot, the problem appears at the extremely large values of the exogenous variable where the predicted values of both risk measures in the linear quantile regression and the linear *CTE* regression are sometimes negative (0.15% of cases for *VaR* and 0.24% of cases for *CTE*) and also where *CTE* is predicted to be lower than *VaR*, 1.02% of cases. The right plot presents the results of the generalized quantile regression (3) and the generalized *CTE* regression (5). As can be observed on the right in Fig. 1, the main difference is that it is impossible to have negative predictions. In addition, note that the generalized regressions does not guarantee the *CTE* > *VaR* restriction. However, in this particular case, we observed all *CTE* predictions above *VaR* predictions, and

[1] A table showing the results obtained with the `esreg` package for linear models and the oracle estimator for $\tau = 0.50, 0.75, 0.90, 0.95$ is available from the authors.

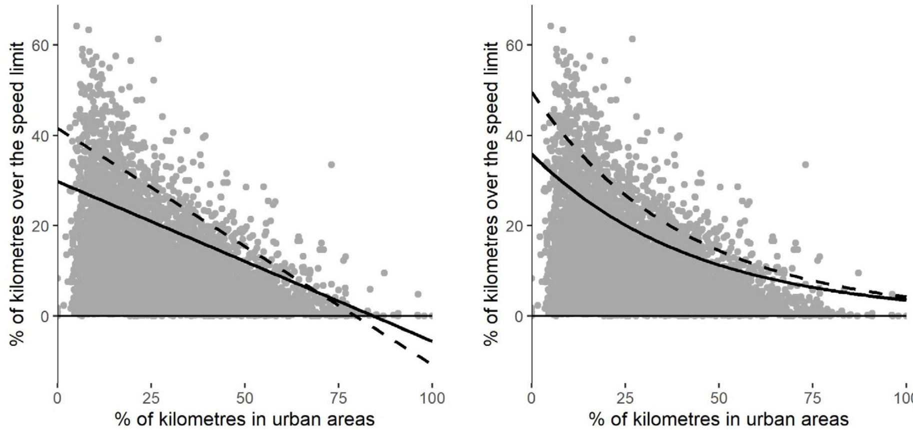

Fig. 1. Linear (left) and generalized (right) quantile regression for *VaR* (solid) and *CTE* (dashed) of the percentage of distance driven above the speed limit as a function of the percentage of urban driving, at $\tau = 0.9$.

we choose not to fit the generalized additive regression model here. In the next subsection, using more covariates, generalized regressions predictions fail to comply with such a restriction and a generalized additive regression will be fitted to overcome this limitation.

4.2. Results for a multivariate analysis

Our aim is now to model the total kilometres driven above the legal speed limit to identify risky drivers who exceed the legal speed limit by considering all other covariates. First, we use a generalized quantile and generalized *CTE* regression, because we do not want to have negative predictions. So, we use an exponential link as in (3) and (5). In addition, we estimate the model where the *CTE* regression part is an additive term, using the specification presented in (6) in order to ensure the $CTE > VaR$ restriction. We also prefer the latter for interpreting the effects of covariates on the tail average, as opposed to the quantile effects.

Table 4 presents the parameter estimates for the generalized quantile regression in (3) and the generalized *CTE* regression in (5). We omit the results for the linear case because they produce predictions that are out of scope (1.93% of predicted cases for *VaR* and 3.60% of cases for *CTE* are negative and *CTE* is predicted to be lower than *VaR*, 6.48% of cases). Table 5 presents the generalized additive *VaR* and *CTE* regression in (6), so that we can interpret the quantile effects and the additional effects for the tail conditional expectation.

An important factor that must be considered when jointly modelling two different risk measures like *VaR* and *CTE* is that there is a possibility that an explanatory variable has an impact on one but not the other. In other words, when considering the mean of the worst cases, *CTE* does not necessarily depend on the same factors as *VaR*. In Table 4, we see that gender has a positive coefficient in the quantile part, meaning that male drivers have a higher predicted quantile than women at the analysed levels, but at the 90th and 95th quantiles we see no significant difference between men and women in the tail expectation. So, in the top decile, the quantile parameter is higher for males than for females (quantile parameter positive and significant), the tail average distance driven above the speed limit does not differ for the two groups of drivers (*CTE* parameter not significant).

Table 5 presents the generalized additive *VaR* and *CTE* regression as in (6) for the 90th quantile. We want to interpret the results for the top decile of risky drivers so this is the reason why we fix $\tau = 0.90$. Here, $CTE_{0.9}(Y_i|X_i) = \exp(X_i'\beta^{0.9}) + \exp(X_i'\eta^{0.9})$. We argue that with this specification we can see the additional influence of each regressor on the tail average. For example, when looking at the results in Table 5, we conclude that an increment of 1% of the total distance (lnKm) causes an increase of 1.141% in the $VaR_{0.9}$ of kilometres driven over the speed limit and an additional increase of 1.784% in the mean kilometres for those drivers exceeding the $VaR_{0.9}$, all other variables being equal. In addition, we see that the effect of age is negative on the *CTE* regression part, meaning that the average distance driven above the speed limit by drivers in the top decile, $\tau = 0.9$, diminishes with age, whereas age does not preclude them from being in the top risk decile. i.e. the age parameter is not a significant parameter in the quantile regression part. Here again, we see the impact of gender with opposite signs on the quantile part and the *CTE* additive term part, which indicates as before that in the

Table 4
Model results for distance driven above the speed limit as a function of total distance driven, percent night driving, percent urban driving, age and gender at quantile levels $\tau = 0.50, 0.75, 0.90$ and 0.95. Standard errors in parenthesis.

$VaR_\tau(Y_i|X_i) = \exp(X_i'\beta^\tau)$
$CTE_\tau(Y_i|X_i) = \exp(X_i'\gamma^\tau)$

	τ			
	0.5	0.75	0.9	0.95
$\beta_{\text{Intercept}}$	-5.247^{***}	-3.541^{***}	-2.494^{***}	-1.884^{***}
	(0.157)	(0.168)	(0.163)	(0.125)
β_{lnKm}	1.320^{***}	1.207^{***}	1.141^{***}	1.094^{***}
	(0.014)	(0.014)	(0.014)	(0.010)
$\beta_{\text{P\_urban}}$	-0.015^{***}	-0.019^{***}	-0.020^{***}	-0.020^{***}
	(0.001)	(0.001)	(0.001)	(0.001)
$\beta_{\text{P\_night}}$	0.004^{**}	0.003^{**}	0.001	0.002
	(0.001)	(0.001)	(0.001)	(0.002)
β_{Age}	-0.011^{***}	-0.007^{**}	−0.001	0.002
	(0.003)	(0.003)	(0.002)	(0.002)
β_{Male}	0.290^{***}	0.246^{***}	0.174^{***}	0.123^{***}
	(0.014)	(0.014)	(0.015)	(0.015)
$\gamma_{\text{Intercept}}$	-4.385^{***}	-3.529^{***}	-2.802^{***}	-2.279^{***}
	(0.351)	(0.372)	(0.380)	(0.403)
γ_{lnKm}	1.364^{***}	1.303^{***}	1.237^{***}	1.180^{***}
	(0.038)	(0.040)	(0.041)	(0.044)
$\gamma_{\text{P\_urban}}$	-0.021^{***}	-0.022^{***}	-0.021^{***}	-0.020^{***}
	(0.001)	(0.001)	(0.002)	(0.002)
$\gamma_{\text{P\_night}}$	−0.002	−0.002	−0.002	−0.001
	(0.002)	(0.002)	(0.002)	(0.002)
γ_{Age}	-0.023^{***}	-0.019^{***}	-0.011^{***}	−0.005
	(0.004)	(0.004)	(0.004)	(0.005)
γ_{Male}	0.140^{***}	0.086^{**}	0.043	0.049
	(0.026)	(0.028)	(0.028)	(0.027)
Goodness-of-fit (R^2)	0.029	0.110	0.472	0.996

p-value <1% ***, <5% ** and <10% *.

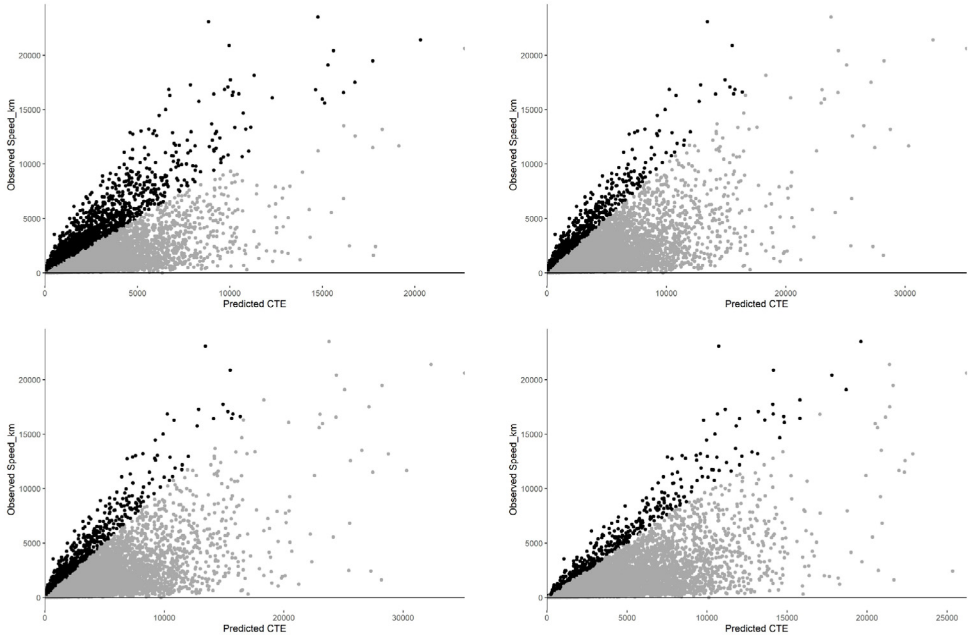

Fig. 2. Observed total distance driven above the speed limit (y-axis) versus predicted *CTE* (x-axis) at $\tau = 0.5$ (top left), 0.75 (top right), 0.90 (bottom left), 0.95 (bottom right). Black dots indicate drivers whose observed distance exceeds the corresponding *CTE* prediction. Other drivers are displayed in grey.

Table 5
Model results for distance driven above the speed limit, as a function of total distance driven, percent night driving, percent urban driving, age and gender at quantile level $\tau = 0.90$. Standard errors in parenthesis.

$VaR_{0.9}(Y_i|X_i) = \exp(X_i'\beta^{0.9})$
$CTE_{0.9}(Y_i|X_i) = \exp(X_i'\beta^{0.9}) + \exp(X_i'\eta^{0.9})$

Variable	$\hat{\beta}^{0.9}$	$\hat{\eta}^{0.9}$
Intercept	-2.494^{***}	-8.500^{***}
	(0.242)	(0.238)
lnKm	1.141^{***}	0.784^{***}
	(0.021)	(0.025)
P_urban	-0.020^{***}	-0.018^{***}
	(0.001)	(0.001)
P_night	0.001	-0.018^{***}
	(0.001)	(0.001)
Age	-0.001	-0.045^{***}
	(0.004)	(0.002)
Male	0.174^{***}	-0.373^{***}
	(0.026)	(0.015)
Goodness-of-fit (R^2)	0.467	

p-value <1% ***, <5% ** and <10% *.

Table 6
Observed distance driven above the speed limit over one year, predicted $VaR_{0.9}$ and $CTE_{0.9}$ for the first six observations in the telematics data set.

Observation	Speed_km	Predicted $VaR_{0.9}$	Predicted $CTE_{0.9}$
1	4212.34	9875.67	12,897.10
2	3647.30	4902.82	6,405.09
3	808.59	5913.95	7,101.61
4	966.69	7743.66	9,632.31
5	2009.42	1681.38	2,077.91
6	187.67	1024.24	1,093.68

top decile, the difference in the average distance above the speed limits between males than for females at the top decile vanishes. Both the percentage of night driving and the percentage of urban driving have negative effects on the tail average, so the higher the percentage of night driving and urban driving, the lower the tail average distance driven above the speed limit in the top decile, given that we have set $\tau = 0.90$.

A part from this new interpretation, it is worth mentioning that the generalized additive model ensures predictions for *CTE* to be greater than for *VaR*. In the generalized models from Table 4, and for $\tau = 0.90$, *CTE* is predicted to be lower than *VaR* 19.09% of cases.

In Table 6, *VaR* and *CTE* are predicted at level $\tau = 0.9$ using the generalized additive model (6) for the first six observations in our dataset. First, note that each driver has a different 90th quantile and *CTE* prediction because they depend on the driver's characteristics. Note also that the fifth observation stands out. That particular driver has an observed total speeding distance equal to 2009.42, which is well above the predicted 90th quantile for drivers with his same characteristics and his observation is almost equal to the tail conditional expectation at level 0.9. This can be used as an indicator of risky driving, as it is widely known that speedy driving is positively correlated with accident occurrence. The situation is quite different for all other drivers and especially for the third, fourth and sixth drivers, who drive at a much less risky speed than the predicted 90th quantile.

In Fig. 2, all the observations versus the *CTE* predictions are compared at different τ levels. The black dots indicate the observations that exceed the mean of the worst cases, $(1 - \tau)$. This serves to identify risky drivers. These drivers have more distance driven above the speed limit than the average of the tail, at the 50th (top left), 75th (top right), 90th (bottom left) and 95th (bottom right) quantile levels. The grey dots indicate the remaining observations.

5. Conclusions

This paper has proposed solutions to the prediction of *VaR* and *CTE* for positive losses. In our view, *CTE* considers values at the extremes and is therefore more informative than *VaR*. When we adjusted the linear regression versions, we observed that there were predictions that did not fall within a plausible range of the response variable and that the predictions for *CTE* were greater than for *VaR*. To overcome these limitations, we propose a joint generalized additive *VaR* and *CTE* regression as the best option to model positive losses.

We have shown that *CTE* predictive modelling is helpful to locate risky drivers in a telematics data set. These methods are easy to implement and can guide risk analysis when there is exogenous information to be considered on the right side of the distribution of a positive response variable.

Our case study shows that risk regression can be applied to the identification of bad drivers and may guide portfolio selection in motor insurance companies once a level of risk appetite has been chosen.

The paper also opens up new lines of research. If it is possible to estimate the effect of covariates for a non-elicitable risk measure such as *CTE*, it should be possible to follow a similar process to predict other risk measures, or to implement other machine learning methodologies to identify the effects of covariates on a risk measure.

References

Acerbi, C., Szekely, B., 2014. Back-testing expected shortfall. Risk 27 (11), 76–81.
Ayuso, M., Guillen, M., Pérez-Marín, A.M., 2016. Telematics and gender discrimination: some usage-based evidence on whether men's risk of accidents differs from women's. Risks 4 (2), 10.
Boucher, J.-P., Côté, S., Guillen, M., 2017. Exposure as duration and distance in telematics motor insurance using generalized additive models. Risks 5 (4), 54.
Chernick, M.R., 2011. Bootstrap Methods: A Guide for Practitioners and Researchers, Vol. 619. John Wiley & Sons.
Dimitriadis, T., Bayer, S., 2019. A joint quantile and expected shortfall regression framework. Electron. J. Stat. 13 (1), 1823–1871.
Dimitriadis, T., Schnaitmann, J., 2019. Forecast encompassing tests for the expected shortfall. Available at SSRN 3436728.
Efron, B., Tibshirani, R.J., 1994. An Introduction to the Bootstrap. CRC press.
Fissler, T., Ziegel, J.F., 2016. Higher order elicitability and Osband's principle. Ann. Statist. 44 (4), 1680–1707.
Gneiting, T., 2011. Making and evaluating point forecasts. J. Amer. Statist. Assoc. 106 (494), 746–762.
Guillen, M., Nielsen, J.P., Ayuso, M., Pérez-Marín, A.M., 2019. The use of telematics devices to improve automobile insurance rates. Risk Anal. 39 (3), 662–672.
Hesterberg, T., 2011. Bootstrap. Wiley Interdiscip. Rev. Comput. Stat. 3 (6), 497–526.
Koenker, R., 2017. Quantile regression: 40 years on. Annu. Rev. Econ. 9, 155–176.
Koenker, R., Bassett, G., 1978. Regression quantiles. Econometrica 33–50.
Koenker, R., Bassett, G., 1982. Robust tests for heteroscedasticity based on regression quantiles. Econometrica 43–61.
Koenker, R., Machado, J.A., 1999. Goodness of fit and related inference processes for quantile regression. J. Amer. Statist. Assoc. 94 (448), 1296–1310.
Kou, S., Peng, X., 2016. On the measurement of economic tail risk. Oper. Res. 64 (5), 1056–1072.
Lourenço, H.R., Martin, O.C., Stützle, T., 2003. Iterated local search. In: Handbook of Metaheuristics. Springer, pp. 320–353.
McCullagh, P., Nelder, J., 1989. Generalized Linear Models. Chapman Hall/CRC.
Nadarajah, S., Zhang, B., Chan, S., 2014. Estimation methods for expected shortfall. Quant. Finance 14 (2), 271–291.
Pérez-Marín, A.M., Guillen, M., Alcañiz, M., Bermúdez, L., 2019. Quantile regression with telematics information to assess the risk of driving above the posted speed limit. Risks 7 (3), 80.
Pitarque, A., Pérez Marín, A.M., Guillén, M., 2019. Regresión cuantílica como punto de partida en los modelos predictivos para el riesgo. Anal. Inst. Actuar. Esp. IV (25), 77–117, 2019.
Taylor, J.W., 2019. Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric Laplace distribution. J. Bus. Econom. Statist. 37 (1), 121–133.
Uribe, J.M., Guillen, M., 2020. Quantile Regression for Cross-Sectional and Time Series Data Applications in Energy Markets using R. Springer.
Wang, R., Ziegel, J.F., 2015. Elicitable distortion risk measures: A concise proof. Statist. Probab. Lett. 100, 172–175.