SORT 50 (1) January-June 2026, 43-66 DOI: 10.57645/20.8080.02.31

Non-crossing neural network quantile regression estimation for driving data with telematics

Xenxo Vidal-Llana[1,*], Carlos Salort Sánchez[1], Vincenzo Coia[2] and Montserrat Guillén[3]

Abstract

State-of-the-art methods for estimating extreme quantiles (value at risk, VaR) and their tail expectations (conditional tail expectation, CTE) under covariate control primarily rely on quantile regression but lack explicit constraints for non-crossing conditions. To address this, we introduce the non-crossing dual neural network, a deep learning model that simultaneously estimates VaRs and CTEs across multiple quantile levels, incorporates covariate dependence, and enables the reconstruction of individual conditional distributions and risk profiles while ensuring the natural order of quantile levels. Using a 2015 telematics dataset, the proposed methodology outperforms benchmark models while enforcing previously unaddressed conditions. The model can be used to identify risk within an insurance portfolio and to analyze extreme right-tail behaviour at the individual level.

MSC: 68T07, 62G32, 62P05, 91G70.

Keywords: Risk evaluation, telematics, quantile regression, motor insurance, value at risk, conditional tail expectation.

1. Introduction

Quantile regression aims at modeling conditional quantiles of a response variable given some covariates and a fixed quantile level (Koenker and Bassett, 1978). This framework naturally allows the reconstruction of conditional distributions of the response by considering multiple quantile levels. In insurance, quantile regression is part of a general

* Corresponding author: juan.j.vidal@uv.es

[1] Universitat de València. Avinguda dels Tarongers. 46022 València, Spain.

[2] BGC Engineering Inc. Suite 500, 980 Howe St. Vancouver, BC Canada.

[3] Universitat de Barcelona. Gran Via de les Corts Catalanes 585. 08007 Barcelona, Spain.

Received: February 2025
Accepted: February 2026

framework known as risk regression, where the aim is to model a conditional risk measure of the response. Algorithms to pursue risk evaluation conditional on covariates are difficult to implement because of their inherent technical difficulties. The availability of large data sets allows predicting high quantiles under this regression framework and creates an environment where big losses can be evaluated given the fixed conditions surrounding the response (Guillen, Bermúdez, and Pitarque, 2021a). A natural extension of quantile regression is to model tail expectations. This enables a more granular characterization of risk at the individual level.

For financial studies, value at risk (VaR) appears as a common risk measure used by the analyst to evaluate extreme quantiles from the left part of the distribution. But, in insurance, it is common to evaluate quantiles from the right part of the distribution, as this literature focuses on predicting losses and its potential impact to the insurance company. The usage of quantile regression models has a long history (Koenker and Bassett, 1978), as it provides information on extreme quantiles where mean-based models are inadequate. Quantile regression for modeling VaR conditional on covariates appears in the financial literature (Bodnar, Hayt and Marston, 1998; Vidal-Llana and Guillén, 2022), but other fields use it as well, like weather forecasting (Cannon, 2018), flood prediction (Zhang, 2016) and car telematics data for telematics-based ratemaking (Guillen et al., 2020; Baione and Biancalana, 2021). In the latter field, the VaR has been especially used due to telematic information containing heavy tails (Pérez-Marín et al., 2019). When predicting losses in the insurance sector, estimators like the VaR help insurance firms manage exposure to extreme events, like earthquakes or massive crashes. Mean-based models assume an underlying normal distribution; therefore, they are not recommended for predicting asymmetric response outcomes, thus the importance of risk measurements such as VaR.

In recent years, some attention has been placed on the conditional tail expectation (CTE), since the Basel III committee (Basel Committee on Banking Supervision, 2016) promoted its use to evaluate risk on portfolios and reserves. The main motivation for CTE is the inclusion of more information about the tail than the VaR, as the VaR limits the estimation of the conditional quantile, while the CTE it estimates the expected beyond-quantile loss. But this change carries a complication, the elicitability or existence of a consistent scoring function that allows estimating the CTE (Gneiting, 2011). While the VaR is elicitable (Koenker and Bassett, 1978), CTE cannot be estimated directly, as it needs the VaR to be estimated beforehand. Fissler and Ziegel (2016) propose a consistent scoring function for the pair (VaR, CTE) which is a generalization of a particular case proposed before by Acerbi and Szekely (2014). Nolde and Ziegel (2017) compare traditional backtesting for elicitable estimators against a specification of the Fissler and Ziegel (2016) scoring function, concluding a better performance and flexibility from the latter one. Afterwards, some other specifications of this loss function were studied and performance across quantile levels and datasets was evaluated (cf. Taylor, 2019; Patton, Ziegel and Chen, 2019; Ziegel et al., 2020). A review of specifications of the general loss function and comparison with state-of-the-art methods for

time series datasets can be found in Taylor (2020). Despite this elicitability issue, there is plenty of fields that use CTE for evaluating risks on heavy tailed distributions. In finance, CTE is used for optimizing portfolios (Ciliberti, Kondor and Mézard, 2007), historical simulation of crises (Kourouma et al., 2010) or for regulation (Chen, 2018). In the energy field, Thaler, Grabec and Poredoš (2005) show the importance of forecasting energy consumption for predicting risk of energy demand and González-Pedraz, Moreno and Peña (2014) study the use of extreme estimators as the tail risk measures for energy portfolios that assume normality and underestimate the actual tail risk. Finally, classical insurance use cases for VaR and CTE are risk capital allocation (Asimit et al., 2011) and applications in predicting extreme losses in motor insurance portfolios (Frees, Shi and Valdez, 2009; Guillen et al., 2021a).

Despite VaR and CTE being popular risk measures in areas other than the insurance and financial industry, there exists an understudied issue: the crossing between the estimates. He (1997) and Yu, Lu and Stander (2003) established the issue of crossing quantiles when estimating several quantiles jointly. Namely, for each outcome variable y_i (i refers to the individual) and each pair of quantile levels $q_1,\ q_2,\ 0 < q_1 < q_2 < 1$, this literature seeks models to ensure that the order is preserved, namely that the quantile of y_i at level q_1 is less than or equal to the quantile of y_i at level q_2, $\text{VaR}_{q_1}(y_i) \leq \text{VaR}_{q_2}(y_i)$. For an authoritative treatment of this subject, see Takeuchi and Furuhashi (2004); Dette and Volgushev (2008); Bondell, Reich and Wang (2010); Chernozhukov, Fernández-Val and Galichon (2010); Liu and Wu (2011). More recent advances propose non-crossing quantile estimation using neural networks, which improve computational time and performance in comparison with other classical methods (cf. Cannon, 2018; Moon et al., 2021). The non-crossing condition that affects the VaR and its corresponding CTE, both assessed at the same quantile level, and which represents one of the ordering properties considered in this framework, was addressed by Acerbi and Szekely (2014), who added a constant inside the loss function, but in practice, finding that constant is non-trivial. Guillen et al. (2021a) solve this issue with a reparametrization of the CTE as a positive excess of the VaR that ensures the desired conditions but only for one quantile at a time. Recently, Fissler, Merz and Wüthrich (2023) proposed the deep composite regression architecture for the joint estimation of value at risk and conditional tail expectation above and below at a given quantile level, ensuring internal coherence between the three risk measures. These approaches focus on estimating VaR and CTE above and below at a single quantile level at a time and do not explicitly address ordering constraints when several quantile levels are considered simultaneously. In contrast, the method proposed in this paper aims to jointly estimate VaR and CTE across a set of quantile levels within a single neural-network framework, while enforcing all natural non-crossing conditions-across quantile levels and across risk measures-by construction. We propose a method that estimates both risk measures VaR and CTE for a set of quantile levels while asserting non-crossing conditions.

One of the areas where data are abundant and there is a need for conditional risk assessment is motor insurance. Driving data provide information to the insurer over time,

such as total distance driven or excess speed, and that information can be combined with static characteristics such as age, type of vehicle or main driving zone. This setting requires multivariate methods that incorporate covariates and are computationally feasible. When choosing quantile regression as the basic approach and evaluating the models separately at different quantile levels, we expect that quantiles estimated for a response are naturally ordered, so that, conditional on the same covariate values, a lower quantile level estimate is lower than a higher quantile level estimate. This is known as the *non-crossing condition* (He, 1997), but it does not necessarily hold in practice. Motor insurance telematics therefore provides a natural and practically relevant setting to illustrate how conditional risk measures can be used to reconstruct individual risk profiles in environments characterized by heterogeneity and extreme outcomes.

We present a neural network that estimates the conditional quantile, or VaR, and the expected loss above its quantile, the CTE, simultaneously and at different quantile levels while requesting that non-crossing conditions are satisfied, and covariates may be introduced. We refer to this proposed approach as the non-crossing dual neural network (NCDNN). By jointly estimating VaRs and CTEs across multiple quantile levels, the proposed framework allows the reconstruction of individual conditional distributions and associated risk profiles. The method shows competitive precision and reasonable computational time compared to other machine learning methods presented in the literature. We compare the results of the non-crossing model to a quantile regression model quantile by quantile (as in Guillen et al. (2021a)) and to the monotone composite quantile regression neural network (MCQRNN) (Cannon, 2018), using a telematics dataset from a Spanish insurance company with 9,614 drivers from 2015, where we evaluate quantile levels 0.5, 0.6, 0.7, 0.8, 0.9, 0.925, 0.95, 0.975 and 0.99. We fit the VaRs and the corresponding CTEs.

This paper is organized as follows. First, Section 2 presents the definitions for value at risk, conditional tail expectation, quantile regression and the non-crossing dual neural network. Section 3 provides a description of the dataset and explains the distribution of the response and covariates used. Section 4 shows the main results, with comparisons between the estimation methods in terms of model performance, and a method to estimate individual conditional distributions of a driver's behaviour. Section 6 concludes our work.

2. Methodology

In this section we explicitly write the notation first and then we present the new proposed model. In insurance we define losses as positive real numbers because we refer to losses as amounts that will be compensated to policyholders. So, the risk is present in the right tail, whereas in finance losses are usually expressed as negative returns and the risk is located on the left tail. This sign convention differs across fields, so we state it explicitly for clarity.

We will assume that losses are positive in the rest of this work and so, that our response variable is positive, as discussed in previous related works (Nolde and Ziegel, 2017).

2.1. Value at risk and quantile regression

The VaR at level q for a random variable Y, $\text{VaR}_q(Y)$, is defined as the magnitude of exceedances of the underlying distribution with probability q, $0 < q < 1$. Mathematically,

$$\text{VaR}_q(Y) = \inf\{y \in \mathbb{R} | F_Y(y) > q\} = F_Y^{-1}(1-q) \,, \tag{1}$$

where F_Y is the distribution function of Y.

For predicting VaR we take use of quantile regression, an extension of linear regression whose objective is to fit the quantile of the response variable for a specified level using a set of k covariates (Koenker and Bassett, 1978; Koenker, 2017). Denoting X as the covariate vector, $T(X) = \{X_1, X_2, \ldots, X_k\}$, where the $T(X)$ denotes the transposed vector, the conditional quantile function can be written as a linear combination of the covariates as

$$Q_{Y|X}(q) = \text{VaR}_{(1-q)}(Y|X) = \beta_{(q)0} + \beta_{(q)1}X_1 + \beta_{(q)2}X_2 + \cdots + \beta_{(q)k}X_k = T(X)\beta_{(q)} \,, \tag{2}$$

with parameter vector $\beta_{(q)}$ corresponding to $\arg\min_{\beta} \mathbb{E}[\rho_q(T(X)\beta, Y)]$, and ρ being the scoring function proposed by Koenker and Bassett (1978):

$$\rho_q(r,y) = (q - \mathbb{1}_{\{y-r<0\}})(y-r) \,, \tag{3}$$

where r is the quantile prediction, y the observed random variable Y, and $\mathbb{1}_{\{\cdot\}}$ is the indicator function, with a value equal to 1 when the subscript is true and 0 otherwise.

In settings characterized by asymmetric response distributions with heavy right tails, classical generalized linear models, which focus on modeling the conditional mean, may provide an incomplete representation of risk and can lead to underpricing of extreme outcomes in insurance applications. By contrast, quantile regression allows for the direct modeling of different parts of the conditional distribution, making it better suited to capture tail behaviour and heterogeneity across risk profiles

2.2. Conditional tail expectation and scoring functions

The CTE for a level q, $0 < q < 1$, is defined as:

$$\text{CTE}_q(Y) = \mathbb{E}[Y | Y \geq \text{VaR}_q(Y)] \,. \tag{4}$$

The CTE, also known as the expected shortfall (ES), tail conditional expectation (TCE), or tail value at risk (TVaR), is a risk measure that evaluates the expected loss conditioned on exceeding its corresponding VaR.

As mentioned above, it is common to assume only positive (or negative) values for the random variable Y, for example, Fissler and Ziegel (2016) define the scoring function for the pair (VaR, CTE) for the negative part of the tail, aiming to low quantile levels, but as for this work, we are interested in the positive part. So, we will use the specification as in Nolde and Ziegel (2017):

$$\begin{aligned} S_q(r_1, r_2, y) = \mathbb{1}_{\{y>r_1\}}\big(-G_1(r_1) + G_1(y) - G_2(r_2)(r_1 - y)\big) + \\ (1-q)\big(G_1(r_1) - G_2(r_2)(r_2 - r_1) + \mathcal{G}_2(r_2)\big)\,, \end{aligned} \tag{5}$$

where r_1 represents the VaR_q, r_2 represents the CTE_q, and $G_1,\ G_2,\ \mathcal{G}_2 : \mathbb{R} \to \mathbb{R}$ are real functions. To ensure consistency as a scoring function for the pair $(\text{VaR}_q,\ \text{CTE}_q)$, Fissler and Ziegel (2016) prove that G_1 must be an increasing function, $\mathcal{G}_2$ must be increasing and concave, and $\mathcal{G}_2' = G_2$. It is important to note that this definition holds for absolutely continuous distributions.

There are several options proposed in the literature for choosing G_1 and G_2. Dimitriadis and Bayer (2019) propose two options for G_1 and five options for G_2 and compare their performance, showing minimal differences between some of them. Taylor (2020) summarizes previously proposed combinations in more depth. We will use $G_1(x) = 0$ and $G_2(x) = 1/x$ for simplicity, which renders the following scoring function:

$$S_q^{AL}(r_1, r_2, y) = \mathbb{1}_{\{y>r_1\}}\frac{y - r_1}{r_2} + (1-q)\left(\frac{r_1 - r_2}{r_2} + \log(r_2)\right)\,. \tag{6}$$

The scoring function represents the negative log-likelihood of an Asymmetric Laplace, hence its superscript *AL* (Taylor, 2019).

Although there is evidence to suggest that some scoring functions are more appropriate than others (Dimitriadis and Bayer, 2019) and that alternative specifications of the functions G_1 and G_2 may yield slightly improved predictive performance, such improvements often come at the cost of a substantial increase in computational time (Vidal-Llana, Coia and Guillen, 2022). In light of these trade-offs, we will retain the study to this scoring functions specification for simplicity.

2.3. The non-crossing condition and the non-crossing dual neural network

Empirical applications have shown that crossings between VaRs are possible. He (1997) and Yu et al. (2003) studied that issue. The fact that a VaR could possibly exceed its corresponding CTE was analyzed by (Acerbi and Szekely, 2014; Guillen et al., 2021a). While certain non-crossing conditions related to VaR or the relationship between VaR and CTE have been previously discussed in the literature, to the best of our knowledge a joint formulation imposing coherent non-crossing constraints simultaneously on VaRs and CTEs across multiple quantile levels has not been explicitly stated. We define the non-crossing conditions for VaR and CTE across several quantile levels with the following property.

Property 1. *For each observation y_i, $i \in \{1,\dots,N\}$, and a pair of consecutive quantile levels from an increasing succession $\{q_j\}_{j=0}^{J}$, $J > 0$, $0 < q_j < 1$ $\forall j$, the (VaR, CTE) predictions ensure non-crossing properties if, $\forall i$:*

1. $VaR_{q_j}(y_i) \leq VaR_{q_{j+1}}(y_i)$,
2. $VaR_{q_j}(y_i) \leq CTE_{q_j}(y_i)$,
3. $CTE_{q_j}(y_i) \leq CTE_{q_{j+1}}(y_i)$.

so, a model that jointly estimates VaR and CTE with a consistent scoring function of the form of (5), while asserting Property 1 produces predictions satisfying the non-crossing property. The model that we present in this study that assures the non-crossing property is the non-crossing dual neural network (NCDNN), a deep learning model that creates an environment in which VaR and CTE predictions maintain the natural ordering of their corresponding quantile levels.

Deep learning is a family of machine learning models based on neural networks (LeCun, Bengio and Hinton, 2015). These models have been used successfully in a variety of fields, including computer vision (Voulodimos et al., 2018) and natural language processing (Vaswani et al., 2017). The architecture of a basic neural network, also called fully connected, consists of L "fully connected" layers. Each layer $l \in \{1,\dots,L\}$, with dimension $h_l > 0$, is defined as:

$$x_{l+1} = Y_l(x_l) = \sigma(W_l \cdot x_l + b_l) \tag{7}$$

where $x_l \in \mathbb{R}^{h_l}$ is the $l-1$-layer output, x_1 is the input vector of covariates ($h_1 = k$), and x_L denotes the output of the final layer of the network, representing the J predicted quantile levels. $W_l \in \mathbb{R}^{h_{l+1} \times h_l}$ denotes the weights matrix, $b_l \in \mathbb{R}^{h_l}$ corresponds to the bias vector and $\sigma(\cdot)$ is a non-linear function.

The model learning process can be roughly explained in two steps. 1) Forward step: The input data passes through the layers where, at each layer, Equation (7) is evaluated, i.e. the output of the previous layer is the input for the current one. 2) Backward step: The scoring function is evaluated on the neural network's output and weights are updated by computing a conjugate gradient variation (backpropagation). Neural networks are particularly well suited for risk estimation problems due to their ability to model complex nonlinear relationships between covariates and outputs, while allowing structural constraints to be incorporated directly into the architecture. In this work, this flexibility is exploited to jointly estimate multiple risk measures across several quantile levels and to enforce non-crossing conditions by construction through the network design.

Our proposed model, the NCDNN, consists of a deep learning model that jointly predicts VaR and CTE for several quantile levels while asserting the non-crossing conditions defined in Property 1. These conditions are met with a model architecture that consists of two fully connected neural networks, one for calculating VaRs and another for calculating CTEs, that estimate consecutive VaR and CTE as exceedances of previous predictions, thus ensuring the natural ordering of the resulting estimates. Many

authors propose the calculation of each VaR as an excess shift from the previous one (cf. Granger, 2010; Schmidt and Zhu, 2016; Catania and Luati, 2022) or estimate VaR and its corresponding CTE above and below (Fissler et al., 2023). Another way to proceed is to use neural networks to predict different quantile levels without crossing (cf. Cannon, 2018; Moon et al., 2021) and this architecture allows incorporating CTE estimation across several quantile levels while enforcing ordering constraints.

For illustrating the definition of the NCDNN, Figure 1 illustrates the distribution of an observed variable together with predictions of VaR and CTE for two quantile levels. We present two possible scenarios. We estimate directly using (3) the VaR for the first quantile level (VaR_{q_0}), and we calculate the second VaR (VaR_{q_1}) and the first CTE (CTE_{q_0}) separately as exceedances of this first VaR. Before computing the next CTE (CTE_{q_1}), two possibilities arise: $\text{VaR}_{q_1} < \text{CTE}_{q_0}$ or the opposite. Both possibilities are represented in the left and right panels of Figure 1, respectively. To ensure non-crossing conditions, we calculate the second CTE (CTE_{q_1}) as an excess of the maximum between its corresponding VaR (VaR_{q_1}) and the previous CTE (CTE_{q_0}). Following this scheme, the non-crossing property holds over a grid of quantile levels. We will show that, under the NCDNN architecture, Property 1 holds.

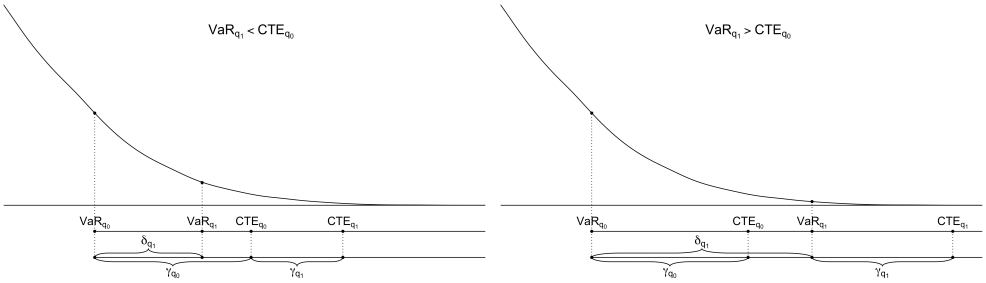

Figure 1. *Representation of NCDNN model for motivation purposes for quantile levels q_0, q_1, $q_0 < q_1$.*

Lemma 1. *Given $\{q_j\}_{j=0}^{J}$, an ordered set of quantile levels, $J > 0$, $0 < q_j < 1 \; \forall j$, and let δ_{q_j}, $\gamma_{q_j} > 0$, be the outputs of a neural network. We define:*

$$\begin{aligned} VaR_{q_{j+1}} &:= VaR_{q_j} + \delta_{q_{j+1}} \, , \\ CTE_{q_{j+1}} &:= VaR_{q_{j+1}} + max(0, CTE_{q_j} - VaR_{q_{j+1}}) + \gamma_{q_{j+1}} \, . \end{aligned}$$

Then the estimates for VaR and CTE using a strictly consistent scoring function for this pair are non-crossing, i.e., they meet Property 1.

Proof.

1.) $\text{VaR}_{q_j}(y_i) \leq \text{VaR}_{q_{j+1}}(y_i)$

$$\text{VaR}_{q_j+1} = \text{VaR}_{q_j} + \delta_{q_{j+1}} \geq \text{VaR}_{q_j} + 0$$

2.) $\text{VaR}_{q_j}(y_i) \leq \text{CTE}_{q_j}(y_i)$

$$\text{CTE}_{q_j} = \text{VaR}_{q_j} + max(0, \text{CTE}_{q_{j-1}} - \text{VaR}_{q_j}) + \gamma_{q_j} \geq \text{VaR}_{q_j} + 0 + 0$$

3.) $\text{CTE}_{q_j}(y_i) \leq \text{CTE}_{q_{j+1}}(y_i)$

Case 1: $\text{VaR}_{q_{j+1}} \geq \text{CTE}_{q_j}$

$$\text{CTE}_{q_{j+1}} = \text{VaR}_{q_{j+1}} + 0 + \gamma_{q_{j+1}} \overbrace{\geq}^{\textit{Case 1 condition}} \text{CTE}_{q_j} + \gamma_{q_{j+1}} \geq \text{CTE}_{q_j} + 0$$

Case 2: $\text{VaR}_{q_{j+1}} < \text{CTE}_{q_j}$

$$\text{CTE}_{q_{j+1}} = \cancel{\text{VaR}_{q_{j+1}}} + \text{CTE}_{q_j} - \cancel{\text{VaR}_{q_{j+1}}} + \gamma_{q_{j+1}} = \text{CTE}_{q_j} + \gamma_{q_{j+1}} \geq \text{CTE}_{q_j} + 0$$

■

We omit the random variable in the notation, but every calculation is done elementwise, so Property 1 is ensured to all observations. We name δ_{q_j} and γ_{q_j} the outputs of the neural networks, which represent the exceedances from previous values after having applied a sigmoid activation function to assure positive values. The architectural representation of the model is found in Figure 2.

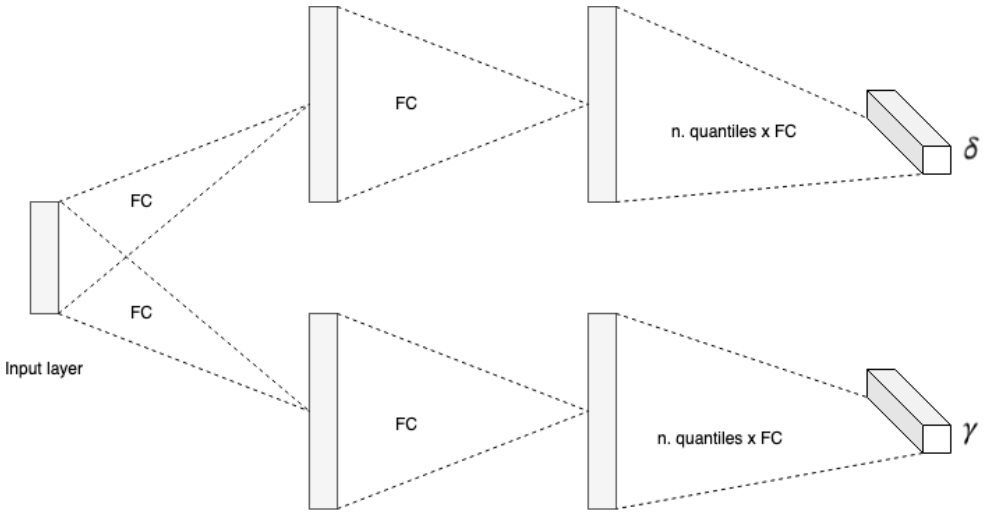

Figure 2. *Architectural diagrams of NCDNN. FC corresponds to a fully connected layer. The last layer is composed of as many FC layers as quantiles are calculated simultaneously.*

For comparison purposes, we are going to show results of the NCDNN compared to a linear quantile regression for the VaR following the same structure as (2) for each quantile level and we optimize a linear model for predicting each corresponding CTE

as an excess. This model structure has been successfully implemented in Acerbi and Szekely (2014); Guillen et al. (2021a). The CTE is defined as:

$$CTE_{Y|X}(q) = VaR_q(Y) + \gamma_{(q)0} + \gamma_{(q)1}X_1 + \gamma_{(q)2}X_2 + \cdots + \gamma_{(q)k}X_k = VaR_q(Y) + T(X)\gamma_{(q)} \,, \tag{8}$$

by using the scoring function (6) for optimizing $\gamma_{(q)}$ and using the previously estimated VaR. This process is repeated for each quantile level, so that predictions meet condition 2 of the non-crossing Property 1 ($VaR_{q_i} < CTE_{q_i}\ \forall i$) as we remove negative values during the estimation process, but the other two non-crossing conditions ($VaR_{q_i} < VaR_{q_{i+1}}\ \forall i$ and $CTE_{q_i} < CTE_{q_{i+1}}\ \forall i$) are not necessarily going to hold and cannot be corrected unless a comprehensive approach is introduced. The code for the NCDNN is publicly available to enable reproducibility of the results and application to other datasets at the following GitHub repository (https://github.com/JuanJoseVidal/ncdnn).

While monotonicity mechanisms based on cumulative positive increments have been previously used in quantile regression and neural network settings to prevent crossings across quantile levels, the contribution of this work does not lie in the monotonicity mechanism itself. Rather, the novelty of the proposed approach is the joint estimation of VaR and CTE over multiple quantile levels within a single neural network architecture, where all natural ordering relationships-across quantile levels and across risk measures-are simultaneously enforced by construction. To the best of our knowledge, existing approaches either address quantile crossing alone or treat VaR and CTE separately, without guaranteeing all non-crossing conditions in a unified framework.

We present the results obtained from our proposed model, the NCDNN, and compare them to two models: 1) a quantile regression estimation results for VaR and the subsequent optimization that leads to the estimation of the CTE one quantile level at a time. This two-step approach is called in this work QR+CTE, and it is implemented in the same way as Guillen et al. (2021a), which assures non-crossing between each VaR and its corresponding CTE (only point 2 of Property 1), but not between different VaRs nor CTEs of different levels. And 2), to the monotone composite quantile regression neural network (Cannon, 2018) for estimating the different VaR and estimating the CTE as averages of all posterior VaRs. We center our study on quantile levels 0.5, 0.6, 0.7, 0.8, 0.9, 0.925, 0.95, 0.975 and 0.99. The particular neural network used in this work consists of 5 neurons in the input layer, corresponding to the number of covariates, and 64 neurons in each fully connected hidden layer, following the architecture presented in Figure 2. Finally, the output layer has a number of neurons equal to the number of quantiles, where the network output represents the increments between the estimated VaRs and CTEs. The stopping criteria used is training was terminated early using a patience-based criterion: if the CTE loss component did not exhibit a strictly lower value than its previously observed minimum for 100 consecutive epochs, optimization was halted and the model parameters from the last epoch were returned. For the three models, we trained 50 different seeds for each model and chose the one that best performs on average for all quantile levels. The best performing selection criteria is based on the average of the normalized errors across quantiles, meaning that the errors for each quantile across

seeds are divided by the maximum, and then for each seed the average is performed. We select the seed with the lower one. Mean and standard deviation of the loss across quantile levels are reported in the Appendix in Table 3, indicating stable performance of the proposed NCDNN under stochastic optimization. All models use scoring functions (3) and (6) for estimating VaR and CTE, respectively, and were trained on a MacBook Pro (13-inch, M1, 2020, 8GB RAM); average training times per random seed were 0.71 minutes for the QR+CTE model, 35.76 minutes for MCQRNN, and 4.10 minutes for the proposed NCDNN.

3. Telematics data from a spanish insurance company

This illustration takes a sample of 9,614 drivers' information from a spanish insurance company consisting of four aggregated factors about their driving behaviour during the year 2015 and two personal factors. Although the data presents annual observations, the same exercise can be applied to monthly, weekly, or even daily observations for creating a more dynamic approach, but putting in risk the consistency of the study, as more granular data implies more volatile information.

The variables used in the study are the percentage of kilometers driven above speed limit (*% Above speed limit*) as the response variable, the percentage of night driving (*% Night*), the logarithm of kilometers driven (*log(Km)*) and the percentage of urban driving (*% Urban*) as telematic factors. Other factors used but only related to personal driver's information are the age (*Age*) and gender (*Gender*). The usage of percentage of kilometers above the speed limit is motivated by the relation of this factor with claim frequency (Pérez-Marín et al., 2019; Guillen et al., 2021a, Guillen, Pérez-Marín and Alcañiz, 2021b).

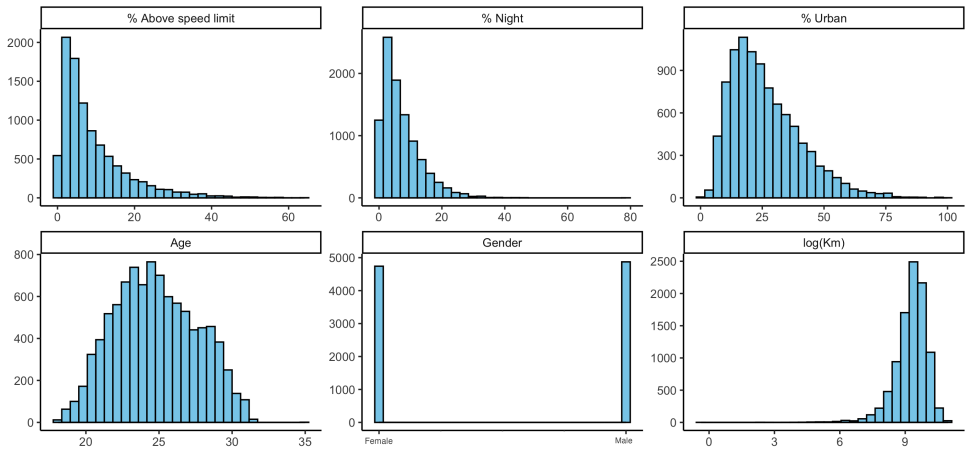

Figure 3. *Distributions of the response variable (%* Above speed limit*) and covariates of the 2015 recollection of telematic information from a spanish insurance company. The total number of drivers is 9,614 drivers in this study.*

Figure 3 shows the distribution of the variables used in the empirical analysis. We observe that the response variable (*% above speed limit*) presents a heavy right tail. Speed violations are directly related to claim frequency, therefore the extreme observations located on the right part of the tail represent a large loss to an insurance company, which motivates our study of extreme quantiles. Other factors also present heavy tails, both percentage or urban driving and percentage of night driving show right tails, and the logarithm of kilometers to the left (which would be right tailed without the logarithm transformation). The average driver on this dataset, by looking at the median values, would be a 25-year-old male, that drives around 11,700 kilometers per year, which are 23.4% on urban areas, 5.3% during the night and 6% above the speed limit.

4. Results

We observe the number of crossings for each pair of consecutive quantile levels and by type of estimator (between consecutive VaRs and consecutive CTEs) in Table 1. This table corresponds to the predictions using the quantile regression for VaR and CTE optimization, one quantile level at a time (QR+CTE). We did not include the table generated by the other two models (MCQRNN and NCDNN) as it is a table where all entries are equal to zero.

4.1. Crossings

Table 1. *Number of crossings for the quantile regression and CTE estimation one quantile level at a time (QR+CTE) (in parenthesis the percentage of observations against the whole sample, 9,614 observations).*

q_i - q_{i+1}	0.5 - 0.6	0.6 - 0.7	0.7 - 0.8	0.8 - 0.9	0.9 - 0.925	0.925 - 0.95	0.95 - 0-975	0.975 - 0.99
$VaR_{q_i} > VaR_{q_{i+1}}$	47 (0.5%)	29 (0.3%)	13 (0.1%)	6 (0.1%)	3 (0.0%)	1 (0.0%)	2 (0.0%)	2 (0.0%)
$CTE_{q_i} > CTE_{q_{i+1}}$	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	265 (2.8%)	33 (0.3%)	14 (0.1%)	764 (7.9%)

Table 1 shows the number of crossings obtained in the illustration data set by a model that considers one quantile level at a time. The first row shows a small number of crossings compared to the second one, 47 cases at most, 0.5% of the dataset. Nonetheless, we must remind that, for example, a value equal to 3 when comparing $VaR_{0.9}$ with $VaR_{0.925}$ means that three observations have been estimated with a $VaR_{0.9}$ greater than a $VaR_{0.925}$, which would be considered unfeasible estimations as they would not follow a natural ordering. We note that those are the type of estimation approaches based on quantile regression that are regularly used nowadays for estimating VaRs for financial firms and insurance companies. If we look at the second row of Table 1, the crossing issue is even more obvious, as CTE crossings reaches 16% for quantile levels 0.95 and 0.975, and although the scoring functions used for estimating CTE are strictly consistent, the

crossing issue must be considered during the model training process by using models like the NCDNN.

As an illustrative example, we consider an observation in which the QR+CTE model has a crossing between $CTE_{0.95}$ and $CTE_{0.99}$, which are two common quantile levels studied in financial and insurance reserving literature. We have an observation that corresponds to a 20-year-old male, that drove 6,800 kilometers, a 75% on urban areas, 3.7% night driving and only 1% above speed limits. For this particular observation, we plot in Figure 4 the predictions for the VaR and CTE for quantile levels 0.95 and 0.99 for both models, the QR+CTE and the NCDNN.

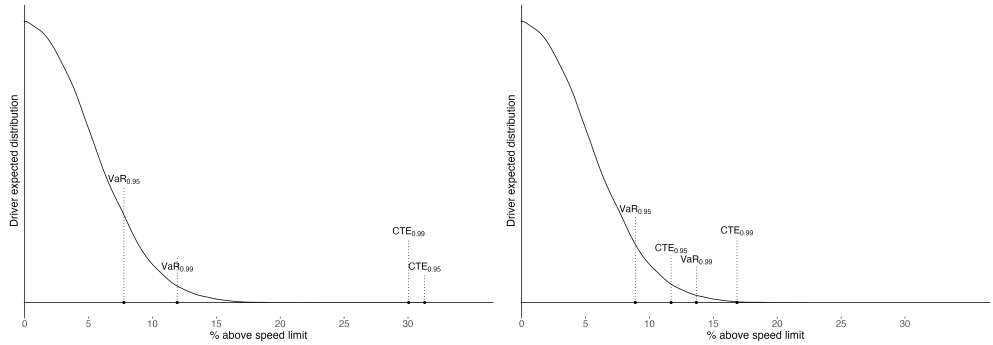

Figure 4. *VaR and CTE predictions for observation 26 of the dataset used in this study, for quantile levels 0.95 and 0.99 and for models QR+CTE (left) and NCDNN (right).*

We observe in Figure 4 how the QR+CTE model predicts a $CTE_{0.95}$ greater than the $CTE_{0.99}$ (left panel), which makes this prediction unfeasible for a direct application into pricing. Also, QR+CTE predictions are more spread than the ones provided by the NCDNN, and apparently the difference between estimates when different quantile levels are considered at the same time, show some difference between the two models. Quantile regression results are accumulated around value 10 and the CTE estimates are located around value 30, while the NCDNN is capable of finding more patterns of these extreme quantile levels and escape from inherited linear predictors that seem to be usual for the traditional QR+CTE models. We argue that this may be due to the non-linear structure that the neural networks are capable of adopting. We also calculated the scoring function (6) for CTE predictions for both quantile levels and for both models to show which is the improvement of changing from a model that admits crossing (QR+CTE) to a model that does not allow it (NCDNN). For quantile level 0.95 the scoring function changes from a 0.13 to a 0.11 (-17% loss), and for quantile level 0.99 the scores change from a 0.028 to a 0.026 (-6% loss). More information about the scoring results across the whole dataset is provided in Section 4.2.

4.2. Model performance

We quantify the improvement on performance of the NCDNN against the QR+CTE and the MCQRNN using scoring functions ρ_q and S_q^{AL} defined in (3) and (6). Results for VaR and CTE scorings are presented in Table 2.

Table 2. *Sum of scoring values for estimated VaR and CTE over the whole dataset for the three models: quantile regression for VaR and CTE optimized one quantile level at a time (QR+CTE), monotone composite quantile regression neural network (MCQRNN), and non-crossing dual neural network (NCDNN), across quantile levels 0.5 to 0.99. Best-performing models for each measure and quantile level are shown in bold.*

	Model	0.5	0.6	0.7	0.8	0.9	0.925	0.95	0.975	0.99
	QR+CTE	27154.01	28709.86	28277.52	25136.43	17671.66	14704.90	11128.86	6631.66	3189.81
VaR	MCQRNN	**26802.60**	28261.37	27719.30	24477.65	16942.51	14083.83	10598.19	6283.45	3010.47
	NCDNN	26802.62	**28257.24**	**27674.48**	**24408.12**	**16911.10**	**14063.51**	**10589.10**	**6269.75**	**3009.66**
	QR+CTE	12734.08	10642.63	8371.30	5893.78	3158.79	2422.24	1658.31	862.27	357.78
CTE	MCQRNN	12777.47	10632.43	8323.67	5832.35	3111.43	2385.44	1632.56	848.20	354.45
	NCDNN	**12604.96**	**10526.57**	**8269.67**	**5815.07**	**3109.02**	**2383.84**	**1631.62**	**846.76**	**352.85**

Observing Table 2, consistently for all quantile levels, and for VaR and CTE, NCDNN overperforms the QR+CTE model, ranging from a 1.3% loss improvement in VaR for median quantile levels to 6% loss improvement in VaR for extreme quantile levels. The improvement in CTE ranges between 1% and 2%. On top of that, QR+CTE model does not follow natural ordering of the estimated quantiles as seen in Table 1, while the other two models meet the non-crossing property.

Regarding the MCQRNN, we observe a lower improvement than with the QR+CTE when comparing to our proposed model. Table 2 shows that only for quantile level 0.5, the MCQRNN overperforms the NCDNN, but for extreme quantile levels, levels that are more of the interest of a risk analyst, the NCDNN situates as the best model, reducing losses to a 0.3% in comparison to the MCQRNN. Regarding the scorings for the CTE estimations, presents in Table 2, we observe a clear improvement for all quantile levels ranging from a 0.1% and 1.3%, which comes mainly due to the ability of the NCDNN to estimate the CTE by optimizing the loss function, and not being approximated by the MCQRNN directly as the observed average of predictions.

4.3. Estimation of the individual distribution of a driver

The non-crossing property allows us to consistently estimate the tail risk of a driver based on covariates and evaluate individual behaviour based on their profile. By removing the restriction on crossing quantiles, we can approximate the conditional distribution of a profile by estimating a set of quantile levels based on covariates. Furthermore, we can

compare observed values to the predicted distribution to assess the risk an individual contributes to a portfolio, not only in comparison to other drivers but also to the expected values of their profile. This practice, in a similar vein, has been recently employed by Bolancé, Cao and Guillen (2024), who used simple-index models to estimate the marginal effect of the heavy tail of claim costs from an insurance company in order to identify risky drivers.

In insurance, this result opens the discussion on assessing a driver's behaviour by combining driving mechanics with their profile. Considering only expected values, a driver who frequently exceeds speed limits would always be deemed high-risk and thus highly priced. However, not all drivers are the same; professional drivers and common drivers exhibit different behaviours and risk levels.

With the non-crossing dual neural network, we can estimate each individual's distribution and compare it to their observed value, thereby ranking driving behaviour and risk based on their covariates. We present an example of a good driver in Figure 5, driver # 54, colored in blue, where we estimated the values at risk, i.e., the quantiles, of the percentage of kilometers driven above speed limits. The observed value is plotted as a vertical line; an observed value below the 0.5 quantile indicates an overperforming driver contributing less risk than expected. Pricing based on these results should consider this good behaviour, potentially reducing or maintaining the risk premium to retain such customers.

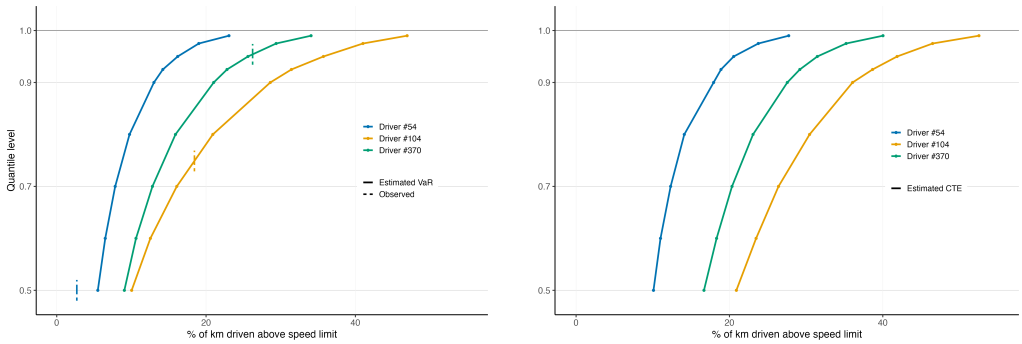

Figure 5. *Estimation of the conditional distribution of the % of kilometers driven above speed limitations for drivers #54, #104 and #370 using VaR (left panel) and CTE (right panel) estimations calculated with the non-crossing dual neural network. In vertical line the observed value.*

In Figure 5, we observe the behaviour of a bad driver, driver #104, colored in yellow, with the percentage of kilometers driven above speed limits falling between the 0.7 and 0.8 estimated quantiles. By incorporating the conditional tail expectation, we can estimate the expected value for a driver with this profile and behaviour. Since this driver's observed value is between the 0.7 and 0.8 quantiles (18% and 20% of kilometers driven above speed limits), it is expected that a driver with a similar profile and behaviour would drive between 25% and 30% of kilometers above speed limits. There-

fore, this driver should be priced higher than a model predicting the expected value, as their profile suggests they should be driving fewer kilometers above speed limits than they actually are.

Figure 5 also illustrates a driver exhibiting very poor behaviour, driver #370 colored in green. The observed value for this driver falls between the 0.95 and 0.975 predicted quantile levels, indicating significantly bad behaviour. This outcome serves as a proxy for identifying customers who are not potentially profitable for the company. In the event that such drivers file a claim, we can expect a high associated cost. Therefore, identifying and appropriately pricing these high-risk drivers is crucial to minimizing potential losses and ensuring the profitability of the insurance portfolio.

5. Discussion

The actuarial literature on usage-based insurance (UBI) is extensive and it has been intensively developed in the last twenty years. Eling and Kraft (2020) made an important contribution to summarize the corpus of knowledge by systematically reviewing 52 academic studies and industry papers published from 2000 to 2019. Many of these articles deal with the role of the distance driven as the key telematics variable for claim frequency estimation. Specifically, Lemaire, Park and Wang (2016) found that annual mileage is an extremely powerful predictor of the number of claims at-fault. In its origins, usage-based insurance was solely based on the total distance driven variable (see Litman (2007), where different types of distance-based insurance price structures are discussed).

In the machine learning literature, there are also many contributions in the field of driving pattern recognition, and how it can be transferred to building accident safety scores and to enhance insurance pricing. Weidner, Transchel and Weidner (2016) found that maneuvers, trips and trip sections as well as the total insurance period can be analyzed to gain significantly scoreable insights into individual driving behaviour. Wüthrich (2017) used high-frequency GPS location data and applied new algorithms to differentiate varying driving styles. He showed how driving styles can be categorized, so that they can be used for a regression analysis in car insurance pricing. Gao and Wüthrich (2018) introduced speed and acceleration heatmaps, categorized with the K-means algorithm to differentiate varying driving styles. Later, Gao, Meng and Wüthrich (2019) investigated the predictive power of covariates extracted from telematics car driving data using the speed-acceleration heat maps. These telematics covariates include K-means classification, principal components, and bottleneck activations from a bottleneck neural network. They found that the first principal component and the bottleneck activations give a better out-of-sample prediction for claims frequencies than other traditional pricing factors such as driver's age. Based on these numerical examples the authors recommended the use of some of these telematics covariates for car insurance pricing. The same authors (see Gao and Wüthrich (2019)) extracted feature information about speeds, acceleration, deceleration, and changes of direction from this high-frequency GPS location data and

used this information to allocate individual car driving trips to selected drivers. Geyer, Kremslehner and Muermann (geyer2020asymmetric) defined a driving factor based on overall distance driven, number of car rides, and speeding and found that speed driving factor has a significant effect on risk. More recently, Meng et al. (2022) calculated risk scores by using a supervised driving risk scoring neural network model and showed that incorporating risk scores improves the prediction performance of classical Poisson regression models for the claim frequency. With these recent references, among many others, we find evidence that risk scoring is becoming central in auto insurance together with big data analytics in this context.

We illustrate this study with a telematics dataset from a spanish insurance company. The case study is similar to what any insurer would do as a first step to score drivers, namely, to analyze the risk of a positive response associated with the risk of a claim given some covariates. Additionally, the model results could be implemented to find the factors that are significantly associated with higher or lower risk.

There are two main advantages of using NCDNN versus risk regressions. First and foremost, the non-crossing property preservation. When conditional risk evaluation is used for pricing, the analyst must estimate the VaR or the CTE given covariates at different quantile levels. Naturally, the estimations should be ordered by quantile level, but this condition does not always hold in practice. A lower quantile level fit that exceeds a higher quantile level fit would be unacceptable, especially if the fitted VaRs or CTEs are being used for pricing or loading a premium. The NCDNN solves this issue by including the natural ordering restriction on VaR and CTE estimations. Secondly, an advantage of NCDNN is the flexibility of neural networks to include cross effects and non-linearities of the covariates, that is superior to what baseline models can provide with linear specifications.

The main disadvantages of implementing a NCDNN are, however, two. On the one hand, the response variable is assumed to be a continuous response. That rules out considering discrete counts for the number of claims as a response, but it could be used for modeling the risk of severities or claim amounts. On the other hand, neural networks are still seen only as exploratory tools in insurance. The fact that they are black boxes that do not serve to write an explicit formula, makes neural network tools inappropriate to the eyes of external stakeholders and incompatible with current regulation requirements. Regulators demand that risk assessment for pricing purposes can explicitly be expressed as a function of rating factors. So, current practice does not allow a neural network fit to be used for pricing, even if explainable artificial intelligence is able to provide an approximation of the variable importance for the output. However, this area is growing very fast, and we believe that deep learning methods combined with explainable artificial intelligence will be helpful to select and even suggest possible non-linearities that explicit classical models can handle. Such procedures could be accepted by regulators as internal steps towards selecting final ratemaking factors. So, we advocate for a cooperation between methods, in particular new machine learning methods, with old methods, rather than just looking at new proposals as opposites or competitors. We also want to

mention that the exploratory tools that are offered by these new methods will certainly be helpful in the discussion of causality and discrimination in insurance, because the new models can cope with many more complexities and can help identifying the role of covariates, even if the role differs among individuals.

We believe that the merits of our proposal can also be expanded in other directions. For example, driving data will certainly be analyzed at increasingly more frequent time intervals than yearly data. That will induce more volatility in the data and, perhaps, methods that are aimed at preserving the natural orderings that we have been presenting here will be welcome if too many inconsistencies are found in the fits otherwise. While still many policyholders pay annual premiums, we are moving into pay-as-you-go schemes that permit monthly rates. Even daily or per trip rates seem plausible, nowadays so we feel that the proposed methods will be helpful in this emerging landscape. We note that the work of Guillen et al. (2021b) explains how risk regressions can be turned into conditional risk scoring that can later be used to summarize driving behaviour very efficiently. This is why we foresee a direct connection of our work to the growing area of insurance telematics.

From an applied perspective, the proposed framework could also be integrated with classical actuarial modeling approaches. In particular, the outputs of the NCDNN may be used to construct derived covariates summarizing individual risk profiles, which could then be incorporated into traditional claims-count models. Such an approach may help bridge telematics-based risk scoring and existing regulatory or operational constraints that favor standard actuarial models. Moreover, while the present work focuses on continuous risk measures, extending neural-network-based architectures to directly model discrete outcomes, such as claim counts, represents a promising direction for future research.

Moreover, while the present work focuses on point estimation of continuous risk measures such as VaR and CTE, uncertainty quantification remains an important topic for practical risk management. In principle, uncertainty around the estimated risk measures could be assessed using resampling-based approaches, such as bootstrap methods, or through Bayesian extensions of the proposed architecture, for example by introducing prior distributions on network parameters or by adopting Bayesian neural network formulations. Exploring such approaches to obtain interval estimates or posterior uncertainty measures represents a promising direction for future research and is beyond the scope of the present work. Finally, while the present study concentrates on continuous risk measures, extending neural-network-based architectures to directly model discrete outcomes, such as claim counts, also represents an appealing avenue for future research.

6. Conclusions

This paper presents a model for estimating the VaR and CTE of a response variable given a set of covariates and involving several quantile levels. The classical approach based on quantile regression is compared to a neural network approach with non-crossing prop-

erties called the NCDNN. Baseline risk regression models do not assure that the non-crossing condition holds, because estimating the VaR and the corresponding CTE at each quantile level are done separately for each quantile level at a time.

This type of analysis brings to light two main points in pricing and risk assessment. First, estimating a driver's risk solely by predicting their expected value is insufficient to capture all potential costs a company may incur. Second, a non-crossing quantile-based approach to risk estimation allows for evaluating how each driver behaves within their profile by locating their observed value between the estimated quantiles. Additionally, the estimation of Conditional Tail Expectations provides a more quantitative approach to understanding the expected value of the percentage of kilometers driven above speed limits when a specific quantile is surpassed. More examples are shown in Figure 6 in the Appendix for further analysis.

It is important to note that this algorithm is not limited to motor insurance but can be applied to any target factors with heavy-tailed distributions where extreme values are of interest to analysts. For example, we are working on estimating energy consumption using household factors, enabling energy companies to provide sufficient resources and manage reserves to prevent energy shortages. Another area of study involves applying the estimation of non-crossing quantiles to predict the conditional distribution of observations in asset pricing. Evaluating stock performance, particularly extreme values in stock prices, can be a valuable tool for identifying investment opportunities and determining underperforming or overperforming stocks. This information allows investors and regulators to manage investment portfolios more effectively in line with their expectations and objectives.

For all these reasons, we think that considering non-crossing algorithms when dealing with heavy tailed response variables in datasets that include fast dynamics, like trip driving data or daily data will be necessary.

Acknowledgements

We would like to acknowledge the Fundació Banc Sabadell: "Ajudes a la investigació 2022", the Spanish Ministry of Science grant PID2019-105986GB-C21/AEI/10.13039/501100011033, the Federación Española de Municipios y Provincias (FEMP) contract number OTR2025-26067SERVI and the NextGenerationEU grant number TED2021-130187B-I00 for their support of our research. We also thank the two anonymous reviewers for their insightful comments and suggestions, which have significantly improved the quality of this manuscript.

References

Acerbi, C. and Szekely, B. (2014). Back-testing expected shortfall. *Risk*, 27(11):76–81.
Asimit, A. V., Furman, E., Tang, Q., and Vernic, R. (2011). Asymptotics for risk cap-

ital allocations based on conditional tail expectation. *Insurance: Mathematics and Economics*, 49(3):310–324.

Baione, F. and Biancalana, D. (2021). An application of parametric quantile regression to extend the two-stage quantile regression for ratemaking. *Scandinavian Actuarial Journal*, 2021(2):156–170.

Basel Committee on Banking Supervision (2016). Minimum capital requirements for market risk. https://www.bis.org/bcbs/publ/d352.htm.

Bodnar, G. M., Hayt, G. S., and Marston, R. C. (1998). 1998 Wharton survey of financial risk management by US non-financial firms. *Financial Management*, 27(4):70–91.

Bolancé, C., Cao, R., and Guillen, M. (2024). Conditional likelihood based inference on single-index models for motor insurance claim severity. *SORT*, 48:235–258.

Bondell, H. D., Reich, B. J., and Wang, H. (2010). Noncrossing quantile regression curve estimation. *Biometrika*, 97(4):825–838.

Cannon, A. J. (2018). Non-crossing nonlinear regression quantiles by monotone composite quantile regression neural network, with application to rainfall extremes. *Stochastic Environmental Research and Risk Assessment*, 32(11):3207–3225.

Catania, L. and Luati, A. (2022). Semiparametric modeling of multiple quantiles. *Journal of Econometrics*. https://doi.org/10.1016/j.jeconom.2022.11.002.

Chen, J. M. (2018). On exactitude in financial regulation: Value-at-risk, expected shortfall, and expectiles. *Risks*, 6(2):61.

Chernozhukov, V., Fernández-Val, I., and Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica*, 78(3):1093–1125.

Ciliberti, S., Kondor, I., and Mézard, M. (2007). On the feasibility of portfolio optimization under expected shortfall. *Quantitative Finance*, 7(4):389–396.

Dette, H. and Volgushev, S. (2008). Non-crossing non-parametric estimates of quantile curves. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(3):609–627.

Dimitriadis, T. and Bayer, S. (2019). A joint quantile and expected shortfall regression framework. *Electronic Journal of Statistics*, 13(1):1823–1871.

Eling, M. and Kraft, M. (2020). The impact of telematics on the insurability of risks. *The Journal of Risk Finance*, 21(2):77–109.

Fissler, T., Merz, M., and Wüthrich, M. V. (2023). Deep quantile and deep composite triplet regression. *Insurance: Mathematics and Economics*, 109:94–112.

Fissler, T. and Ziegel, J. F. (2016). Higher order elicitability and Osband's principle. *The Annals of Statistics*, 44(4):1680–1707.

Frees, E. W., Shi, P., and Valdez, E. A. (2009). Actuarial applications of a hierarchical insurance claims model. *ASTIN Bulletin: The Journal of the IAA*, 39(1):165–197.

Gao, G., Meng, S., and Wüthrich, M. V. (2019). Claims frequency modeling using telematics car driving data. *Scandinavian Actuarial Journal*, 2019(2):143–162.

Gao, G. and Wüthrich, M. V. (2018). Feature extraction from telematics car driving heatmaps. *European Actuarial Journal*, 8:383–406.

Gao, G. and Wüthrich, M. V. (2019). Convolutional neural network classification of telematics car driving data. *Risks*, 7(1):6.

Geyer, A., Kremslehner, D., and Muermann, A. (2020). Asymmetric information in automobile insurance: Evidence from driving behavior. *Journal of Risk and Insurance*, 87(4):969–995.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762.

González-Pedraz, C., Moreno, M., and Peña, J. I. (2014). Tail risk in energy portfolios. *Energy Economics*, 46:422–434.

Granger, C. W. (2010). Some thoughts on the development of cointegration. *Journal of econometrics*, 158(1):3–6.

Guillen, M., Bermúdez, L., and Pitarque, A. (2021a). Joint generalized quantile and conditional tail expectation regression for insurance risk analysis. *Insurance: Mathematics and Economics*, 99:1–8.

Guillen, M., Nielsen, J. P., Pérez-Marín, A. M., and Elpidorou, V. (2020). Can automobile insurance telematics predict the risk of near-miss events? *North American Actuarial Journal*, 24(1):141–152.

Guillen, M., Pérez-Marín, A. M., and Alcañiz, M. (2021b). Percentile charts for speeding based on telematics information. *Accident Analysis & Prevention*, 150:105865.

He, X. (1997). Quantile curves without crossing. *The American Statistician*, 51(2):186–192.

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

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society*, 46(1):33–50.

Kourouma, L., Dupre, D., Sanfilippo, G., and Taramasco, O. (2010). Extreme value at risk and expected shortfall during financial crisis. *Available at SSRN 1744091*.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

Lemaire, J., Park, S. C., and Wang, K. C. (2016). The use of annual mileage as a rating variable. *ASTIN Bulletin: The Journal of the IAA*, 46(1):39–69.

Litman, T. (2007). Distance-based vehicle insurance feasibility, costs and benefits. *Victoria*, 11.

Liu, Y. and Wu, Y. (2011). Simultaneous multiple non-crossing quantile regression estimation using kernel constraints. *Journal of Nonparametric Statistics*, 23(2):415–437.

Meng, S., Wang, H., Shi, Y., and Gao, G. (2022). Improving automobile insurance claims frequency prediction with telematics car driving data. *ASTIN Bulletin: The Journal of the IAA*, 52(2):363–391.

Moon, S. J., Jeon, J.-J., Lee, J. S. H., and Kim, Y. (2021). Learning multiple quantiles with neural networks. *Journal of Computational and Graphical Statistics*, 30(4):1238–1248.

Nolde, N. and Ziegel, J. F. (2017). Elicitability and backtesting: Perspectives for banking regulation. *The Annals of Applied Statistics*, 11(4):1833–1874.

Patton, A. J., Ziegel, J. F., and Chen, R. (2019). Dynamic semiparametric models for expected shortfall (and Value-at-Risk). *Journal of Econometrics*, 211(2):388–413.

Pérez-Marín, A. M., Guillen, M., Alcañiz, M., and Bermúdez, L. (2019). Quantile regression with telematics information to assess the risk of driving above the posted speed limit. *Risks*, 7(3):80.

Schmidt, L. and Zhu, Y. (2016). Quantile spacings: A simple method for the joint estimation of multiple quantiles without crossing. *Available at SSRN 2220901*.

Takeuchi, I. and Furuhashi, T. (2004). Non-crossing quantile regressions by svm. *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541)*, 1:401–406.

Taylor, J. W. (2019). Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric Laplace distribution. *Journal of Business & Economic Statistics*, 37(1):121–133.

Taylor, J. W. (2020). Forecast combinations for value at risk and expected shortfall. *International Journal of Forecasting*, 36(2):428–441.

Thaler, M., Grabec, I., and Poredoš, A. (2005). Prediction of energy consumption and risk of excess demand in a distribution system. *Physica A: Statistical mechanics and its applications*, 355(1):46–53.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.

Vidal-Llana, X., Coia, V., and Guillen, M. (2022). Alternative scoring function specifications for estimating conditional tail expectation regression via neural networks. In *Contributions to Risk Analysis: RISK 2022*, pages 139–148. Cuadernos de la Fundación. Fundación Mapfre. https://www.fundacionmapfre.org/publicaciones/todas/contributions-risk-analysis-2022/.

Vidal-Llana, X. and Guillén, M. (2022). Cross-sectional quantile regression for estimating conditional VaR of returns during periods of high volatility. *The North American Journal of Economics and Finance*, 63:101835.

Voulodimos, A., Doulamis, N., Doulamis, A., and Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Computational Intelligence and Neuroscience*, 2018.

Weidner, W., Transchel, F. W., and Weidner, R. (2016). Classification of scale-sensitive telematic observables for riskindividual pricing. *European Actuarial Journal*, 6:3–24.

Wüthrich, M. V. (2017). Covariate selection from telematics car driving data. *European Actuarial Journal*, 7:89–108.

Yu, K., Lu, Z., and Stander, J. (2003). Quantile regression: applications and current research areas. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 52(3):331–350.

Zhang, L. (2016). Flood hazards impact on neighborhood house prices: A spatial quantile regression analysis. *Regional Science and Urban Economics*, 60:12–19.

Ziegel, J. F., Krüger, F., Jordan, A., and Fasciati, F. (2020). Robust forecast evaluation of expected shortfall. *Journal of Financial Econometrics*, 18(1):95–120.

Appendix

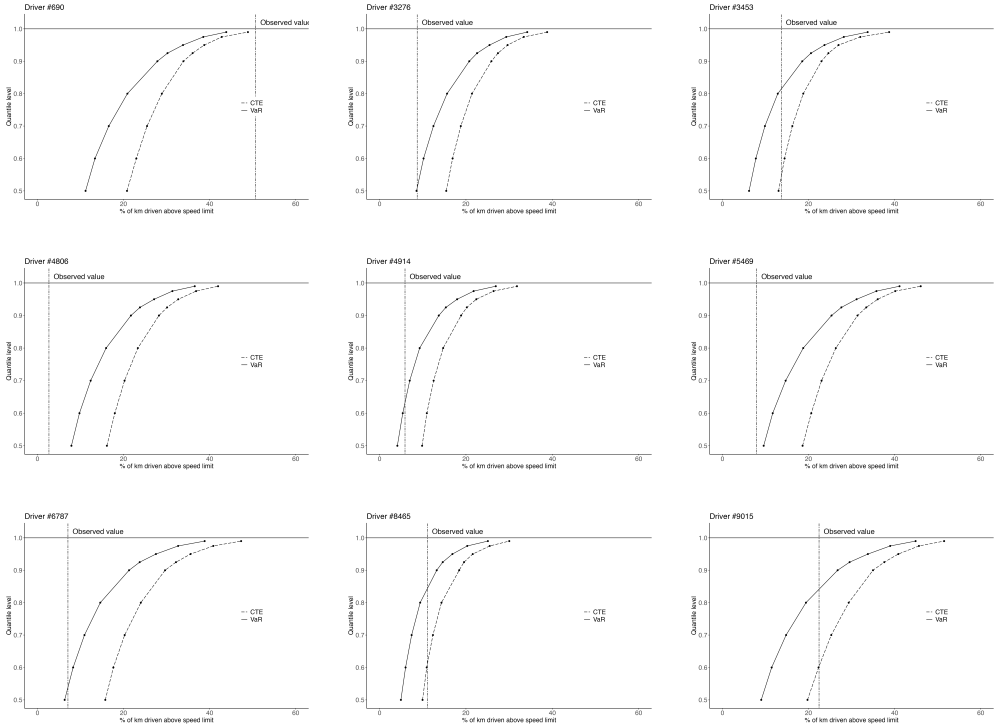

Figure 6. *Estimation of the conditional distribution of the % of kilometers driven above speed limitations for different drivers of the database using VaR and CTE estimations calculated with the non-crossing dual neural network. In vertical line the observed value.*

.

Table 3. *Mean and standard deviance of errors for the NCDNN across 50 seeds.*

Loss NCDNN	Quantile Level								
	0.5	0.6	0.7	0.8	0.9	0.925	0.95	0.975	0.99
Mean	13743.1	11391.5	8884.8	6203.5	3305.4	2523.9	1722.4	895.3	372.7
Std. Dev.	2698.5	2031.3	1429.9	879.2	401.0	285.6	179.7	90.1	35.1