"Non-Crossing Dual Neural Network: Joint Value at Risk and Conditional Tail Expectation estimations with non-crossing conditions"

Xenxo Vidal-Llana, Carlos Salort Sánchez, Vincenzo Coia and Montserrat Guillen

UBIREA | Institut de Recerca en Economia Aplicada Regional i Pública

UBIREA | Institut de Recerca en Economia Aplicada Regional i Pública
UNIVERSITAT DE BARCELONA

WEBSITE: www.ub.edu/irea/ • CONTACT: irea@ub.edu

The Research Institute of Applied Economics (IREA) in Barcelona was founded in 2005, as a research institute in applied economics. Three consolidated research groups make up the institute: AQR, RISK and GiM, and a large number of members are involved in the Institute. IREA focuses on four priority lines of investigation: (i) the quantitative study of regional and urban economic activity and analysis of regional and local economic policies, (ii) study of public economic activity in markets, particularly in the fields of empirical evaluation of privatization, the regulation and competition in the markets of public services using state of industrial economy, (iii) risk analysis in finance and insurance, and (iv) the development of micro and macro econometrics applied for the analysis of economic activity, particularly for quantitative evaluation of public policies.

Abstract

When datasets present long conditional tails on their response variables, algorithms based on Quantile Regression have been widely used to assess extreme quantile behaviors. Value at Risk (VaR) and Conditional Tail Expectation (CTE) allow the evaluation of extreme events to be easily interpretable. The state-of-the-art methodologies to estimate VaR and CTE controlled by covariates are mainly based on linear quantile regression, and usually do not have in consideration non-crossing conditions across VaRs and their associated CTEs. We implement a non-crossing neural network that estimates both statistics simultaneously, for several quantile levels and ensuring a list of non-crossing conditions. We illustrate our method with a household energy consumption dataset from 2015 for quantile levels 0.9, 0.925, 0.95, 0.975 and 0.99, and show its improvements against a Monotone Composite Quantile Regression Neural Network approximation.

JEL classification: C31, C45, C52.

Keywords: Risk evaluation, Deep learning, Extreme quantiles.

Xenxo Vidal-Llana (Corresponding author): Universitat de Barcelona. Gran Via de les Corts Catalanes 585. 08007 Barcelona, Spain. Email: juanjose.vidal@ub.edu

Carlos Salort Sánchez: Universitat de Barcelona. Gran Via de les Corts Catalanes 585. 08007 Barcelona, Spain. Email: carlossalort@gmail.com

Vincenzo Coia: University of British Columbia. West Mall 2329. Vancouver, BC Canada. Email: vincen.coia@stat.ubc.ca

Montserrat Guillen: Universitat de Barcelona. Gran Via de les Corts Catalanes 585. 08007 Barcelona, Spain. Email: mguillen@ub.edu

Acknowledgements

We want to have a special acknowledgement to Fundació Banc Sabadell: "Ajudes a la investigació 2022", Fundación BBVA: "Ayudas a proyectos de investigación en Big Data", AGAUR: "PANDÈMIES" grant and the Spanish Ministry of Science grant PID2019-105986GB-C21 for their support to our research.

Non-Crossing Dual Neural Network: Joint Value at Risk and Conditional Tail Expectation estimations with non-crossing conditions

Xenxo Vidal-Llana*† Carlos Salort Sánchez* Vincenzo Coia‡

Montserrat Guillén*

Abstract

When datasets present long conditional tails on their response variables, algorithms based on Quantile Regression have been widely used to assess extreme quantile behaviors. Value at Risk (VaR) and Conditional Tail Expectation (CTE) allow the evaluation of extreme events to be easily interpretable. The state-of-the-art methodologies to estimate VaR and CTE controlled by covariates are mainly based on linear quantile regression, and usually do not have in consideration non-crossing conditions across VaRs and their associated CTEs. We implement a non-crossing neural network that estimates both statistics simultaneously, for several quantile levels and ensuring a list of non-crossing conditions. We illustrate our method with a household energy consumption dataset from 2015 for quantile levels 0.9, 0.925, 0.95, 0.975 and 0.99, and show its improvements against a Monotone Composite Quantile Regression Neural Network approximation.

Keywords: risk evaluation, deep learning, extreme quantiles

1 Introduction

Evaluation of heavy tailed distributions is a crucial part of risk assessment. The incorporation of Value at Risk (VaR) to the analyst toolbox triggered a decisive way to evaluate extreme quantiles from the right (or left) part of the distribution. Commonly, the quantile levels used are greater than 0.9 (or lower than 0.1 when interested in the left part of the tail). Ever since Koenker and Bassett Jr, 1978 proposed the

*Universitat de Barcelona. Gran Via de les Corts Catalanes 585. 08007 Barcelona, Spain.

†Corresponding author: juanjose.vidal@ub.edu

‡University of British Columbia. West Mall 2329. Vancouver, BC Canada

Quantile Regression specification, the use of extreme estimators became a necessity more than an optional tool. Despite finance being the main user of VaR for evaluating future risks (Bodnar et al., 1998), it is not the only field that uses quantile regression. Estimation of rainfall curves for storm forecasts (Cannon, 2018), car telematics data for price drivers in a pay-as-you-go scheme (Guillen et al., 2020) and riverbed widths for predicting floods (Zhang, 2016) are some examples outside the financial topic where the VaR has a powerful role as a statistic to predict future risk.

In recent years, part of the attention of the researchers has moved from the VaR to the Conditional Tail Expectation (CTE), since the Basel III committee (see Basel Committee on Banking Supervision, 2016) motivates its use to evaluate risk on portfolios and reserves. While the VaR focuses on the expected loss given a certain probability, the CTE prioritizes the expected value over its related VaR, estimating the expected cost of an unlikely distress. But moving from the VaR to the CTE adds a complication: elicitability. The definition of elicitability can be reduced into the existence of a scoring function that is strictly consistent (Gneiting, 2011). While the VaR is elicitable (Koenker and Bassett Jr, 1978), the CTE alone is not, and needs the VaR to be estimated jointly in order to be elicitable (Fissler and Ziegel, 2016). Fissler and Ziegel, 2016 also propose a consistent scoring function for this pair, which is a generalization of the one proposed before by Acerbi and Szekely, 2014 (but without opening the discussion of elicitability). Nolde and Ziegel, 2017 compare traditional backtesting for high elicitable estimators against a specification of the Fissler and Ziegel, 2016 scoring formula, concluding a better performance and flexibility from the latter one. Afterwards, literature studied and proposed some other specifications of this loss function and evaluate performance across quantile levels and datasets (e.g. Taylor, 2019; Patton et al., 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.

On a different series of events, He, 1997 and Yu et al., 2003 created another line of research: the problem of crossing quantiles when calculating several ones jointly. Namely, for each observation y_i and each pair of quantile levels q_1, q_2, $q_1 < q_2$, this literature seeks models to ensure that $\text{VaR}_{q_1}(y_i) \leq \text{VaR}_{q_2}(y_i)$. Well known studies on this matter are Takeuchi and Furuhashi, 2004, Dette and Volgushev, 2008, Bondell et al., 2010, Chernozhukov et al., 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 more classical methods (see Cannon, 2018 and Moon et al., 2021).

Directly related literature to CTE estimation and non-crossing quantile regression are the experiments to estimate VaR and CTE without crossing between them. The scoring methods proposed beforehand do not solve a crucial condition when put into practice: for every observation y_i and a quantile q, $\text{VaR}_q(y_i) \leq \text{CTE}_q(y_i)$. One successful attempt was made by Acerbi and Szekely, 2014, in which they add a constant inside the loss function that theoretically ensures this condition to happen, but in practice, finding that constant becomes non-trivial. In other literature, a reparametrization of the CTE as a positive excess of the VaR is commonly used (see Guillen et al., 2021), which ensures the desired output but only one quantile level at a time. Even with its importance as a topic given the numerous available techniques, there is a clear lack of studies about non-crossing VaR and their respective CTE when several quantile levels come into play.

We present a neural network that estimates the VaR and the CTE together, that estimates several quantile levels at the same time, and improves precision against several models, using a significantly lower amount of computational time. Furthermore, our model holds the non-crossing conditions presented in Property 1, that have not been studied in literature altogether.

We use an energy consumption dataset for 4,517 U.S. households from 2015 and evaluate quantile levels 0.9, 0.925, 0.95, 0.975 and 0.99 to simulate an analysis of an energy firm that aims to predict future consumption peaks with use of VaR and CTE corresponding to such quantile levels, that includes a weighting variable which use evokes a model that can be extrapolated to the whole U.S.

This paper is organized as follows. Section 2 presents the definitions for Quantile Regression, Value at Risk, Conditional Tail Expectation, our main contribution, the Non-Crossing Dual Neural Network, and the model used to compare results. Section 3 provides a description of the dataset and presents descriptives of the response and covariates used. Section 4 presents the main assessed results, with comparisons between the estimation methods. Section 5 concludes our work.

2 Methodology

2.1 Quantile Regression and Value at Risk

The Value at Risk at level q for a random variable Y represents the maximum amount of loss (or profit) that a variable is expected to reach with probability q. It is joint together with a period so it is defined for a certain interval, but for our case we will omit the temporality part. Its definition is the following:

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

where F_Y is the distribution function of the random continuous variable Y.

We predict the VaR using Quantile Regression, an extension of linear regression whose objective is to fit the quantile of the response variable using a set of covariates [Koenker and Bassett Jr, 1978, Koenker, 2017]. Let Y be a random variable with a conditional distribution function $F_{Y|X}$ that depends on k covariates defined in vector X, $X' = \{X_1, X_2, \ldots, X_k\}$, where the $'$ denotes the transposed vector.

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

with parameter $\beta_{(q)} = \arg\min_{\beta} \mathbb{E}[\rho_q(X_i'\beta, Y_i)]$, and ρ corresponds to the scoring function that [Koenker and Bassett Jr, 1978] also proposed:

$$\rho_q(r_1, y) = (q - \mathbb{1}_{\{y - r_1 < 0\}})(y - r_1) \tag{3}$$

where r_1 is the quantile prediction, y the observed random variable, and $\mathbb{1}_{\{\cdot\}}$ is the indicator function, with a value equal to 1 when the subscript is true and 0 otherwise.

2.2 Conditional Tail Expectation and scoring

The Conditional Tail Expectation (CTE) for a level q is defined as:

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

The CTE, known alternatively as Tail Conditional Expectation (TCE), Expected Shortfall (ES) or Tail Value at Risk (TVaR), is a risk measure that is the expected loss conditioned on exceeding its related VaR. Its main motivation comes from the [Basel Committee on Banking Supervision, 2016], in which they present it as an alternative to VaR to evaluate future risks of a portfolio.

In literature, it is common to suppose only negative or positive values for the random variable Y, for example, Fissler and Ziegel, 2016 define the scoring function for the couple (VaR, CTE) for the negative part of the tail, aiming to low quantile levels, but as we are interested in the positive part, we will use Nolde and Ziegel, 2017 specification, which is the following:

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

For our study, r_1 represents the VaR_q, and r_2 represents the CTE_q. $G_1,\ G_2,\ \mathcal{G}_2 : \mathbb{R} \to \mathbb{R}$. To ensure consistency on Equation (5) as a scoring function, Fissler and Ziegel, 2016 prove that G_1 must be an increasing function, $\mathcal{G}_2$ must be increasing and concave, and $\mathcal{G}_2' = G_2$. While those conditions are meet, S_q serves as a consistent scoring function for the pair ($\text{VaR}_q,\ \text{CTE}_q$).

There are several options proposed in 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, proving minimal differences between some of them. Taylor, 2020 summarizes previously proposed combinations more in depth. We will use $G_1(x) = 0$ and $G_2(x) = 1/x$ because of its simplicity, which renders the following scoring function:

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

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

Although there is evidence in literature that results are non-similar between all different proposed choices (see Dimitriadis and Bayer, 2019), we introduce the Murphy diagrams in section 4 for comparing different models, which compares two-point estimations without relying on the scoring function choosing.

2.3 The Non-Crossing Dual Neural Network

Deep learning LeCun et al., 2015 is a family of machine learning models based on neural networks. They have been used successfully in a multitude of fields, including computer vision Voulodimos et al., 2018 and natural language processing Vaswani et al., 2017. In this paper we propose Non-Crossing Dual Neural Network, a deep learning model that can be used to predict jointly VaR and CTE for several quantile levels while asserting certain non-crossing conditions. We decided to use neural networks for their flexibility when creating custom model structures, but other Machine Learning models could be tested for further

exploration, like Random Forests (Ho, 1995) or Gradient Boosting (Freund and Schapire, 1995).

The main component of our architecture are fully connected (FC) layers. Each layer l is represented as a non-linear function, defined as

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

where x_l is a vector, W_l is a weights matrix, b_l is the bias vector, and σ is a non-linear function. The architecture of the neural network is composed by stacking several layers and uses as input the vector of covariates. 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, the 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 an update of the weights is performed by computing a conjugate gradient variation (backpropagation).

In regards of our main problem, we define non-crossing conditions in Property 1.

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

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. $CTE_{q_j}(y_i) \leq CTE_{q_{j+1}}(y_i)$.

Our proposed network can calculate multiple VaR and CTE simultaneously, while maintaining a series of non-crossing conditions (Property 1)[1]. To achieve this goal, we follow a model architecture that consists of two almost identical feed-forward neural networks, one for calculating VaR and one for calculating CTE. We aim to calculate VaR and CTE as exceedances of previous predictions, following the intuition from Figure 1. In this figure we present the distribution of an observed variable, and predictions of VaR and CTE for two quantile levels. We also present two possible scenarios that motivate the definition of the model. We estimate 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. Two possibilities

[1] A repository with code and documentation will be available soon for download and testing.

arise before calculating the next CTE (CTE_{q_1}), $\text{VaR}_{q_1} < \text{CTE}_{q_0}$ or the contrary. Both possibilities are represented in both panels of Figure 1. To fulfill non-crossing conditions, we calculate the second CTE (CTE_{q_1}) as an excess of the maximum between its correspondent VaR (VaR_{q_1}) and the previous CTE (CTE_{q_0}). Following this scheme, the non-crossing property will be fulfilled for a grid of quantile levels.

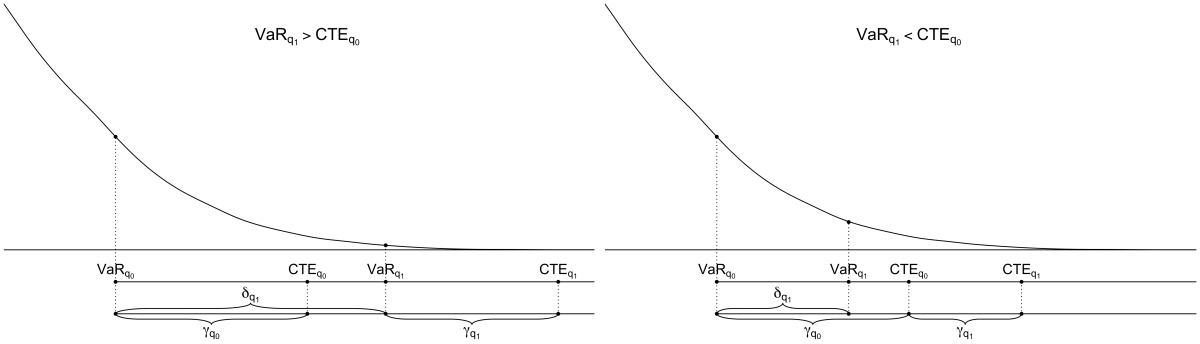

Figure 1: Theoretical representation of our model for motivation purposes.

We omit in the notation the random variable, but every calculation is done elementwise, so it will be 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.

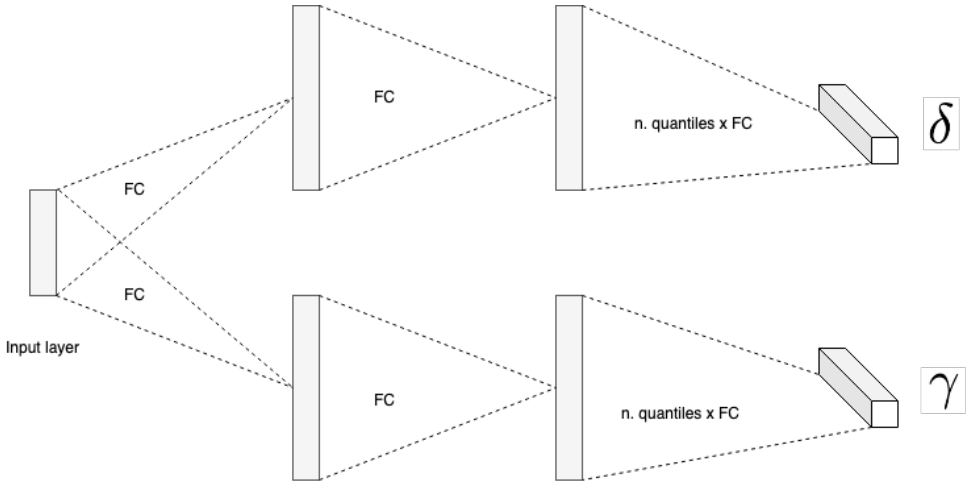

Figure 2: Architectural diagrams of NCDNN. The last layer is composed of as many FC layers as quantiles are calculated simultaneously. FC corresponds to a fully connected layer.

We use the outputs of the NCDNN to predict VaR_{q_j} and CTE_{q_j}, $\forall j$. We will prove that using our definition, Property 1 is held.

Theorem 1. *Given* $\{q_j\}_{j=0}^{J}$ *an ordered set of quantile levels,* $J > 0$, $0 < q_j < 1$ $\forall i, j$, *and* δ_{q_j}, $\gamma_{q_j} > 0$ *the outputs of a neural network with structure similar to Figure 2. We define:*

$$VaR_{q_{j+1}} := VaR_{q_j} + \delta_{q_{j+1}}$$

$$CTE_{q_{j+1}} := VaR_{q_{j+1}} + max(0, CTE_{q_j} - VaR_{q_{j+1}}) + \gamma_{q_{j+1}}$$

Then the predictions for VaR and CTE 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.) $\text{VaR}_{q_j}(y_i) \leq \text{CTE}_{q_j}(y_i)$

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

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

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

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

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

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

□

2.4 Monotone Composite Quantile Regression Neural Network and other tested models

Our main objective is to find a model which can predict several quantiles and their correspondent CTE together while ensuring non-crossing conditions of Property 1. We tested other models like linear and generalized linear regression for VaR and CTE separately and simultaneously, predicting a linear quantile regression and the average exceedances with a Machine Learning model, like Random Forest or Gradient Boosting, and finally the Monotone Composite Quantile Regression Neural Network (MCQRNN). This latter model is the one on which we will focus the discussion, as it is the only model we found that satisfies all required non-crossing conditions (Property 1).

We discarded models used in part of the literature based on time series like GARCH or CAViaR as our database is cross-sectional and not temporal, although our model could be improved with recursive layers to catch momentum trends, and we plan to add those in future research.

The MCQRNN, proposed by [Cannon, 2018], is an expansion of the Quantile Regression Neural Networks initially proposed by [White, 1992] with several improvements, like the non-crossing quantile estimation and optional monotonicity and sign constraints for the model. We use the MCQRNN to predict an extensive grid of quantile levels of the right part of the tail, which correspond to the different VaR. Afterwards, we estimate each CTE by averaging the greater or equal quantile predictions to their correspondent VaR. As the model ensures non-crossing quantile estimation, the CTE estimated will inherit this property and be also consecutive, fulfilling the non-crossing Property 1.

3 Data

We apply the proposed methodology to the Residential Energy Consumption Survey (RECS) [EIA, 2015], which is an American poll that collects information related to energy consumption, expenditure and characteristics of housing units occupied as a primary residence and the households that live in them. More precisely, we use the 2015 recollection for this analysis, which constitutes a sample of 4,517 US households, selected randomly based on an area-probability design. The dataset presents a weighting factor, whose application to the loss in the training phase of a model, can extend its results to the total population of the 118.2 million US households.

The variables used in the study are the following. DOLLAREL is the response variable, which represents the total cost of the electricity of the household in 2015. The covariates used are: SWIMPOOL is a dummy variable with value equal to one if there is pool in the property. SOLAR is a dummy factor with value one if the household has on-site electricity generation from solar, and zero otherwise. The TOTSQFT_EN, TOTCSQFT and TOTHSQFT stand for the total, cooled and heated total square footage. BEDROOMS indicate the number of bedrooms. TVCOLOR means the number of televisions used in the household. NUMSMPHONE refers to the number of smart phones. NUMFLOORFAN denote the number of floor, window or table fans used. Finally, NHSLDMEM represents the number of household members.

In Figure 3 we show the distribution of the variables used in the analysis. We observe how the

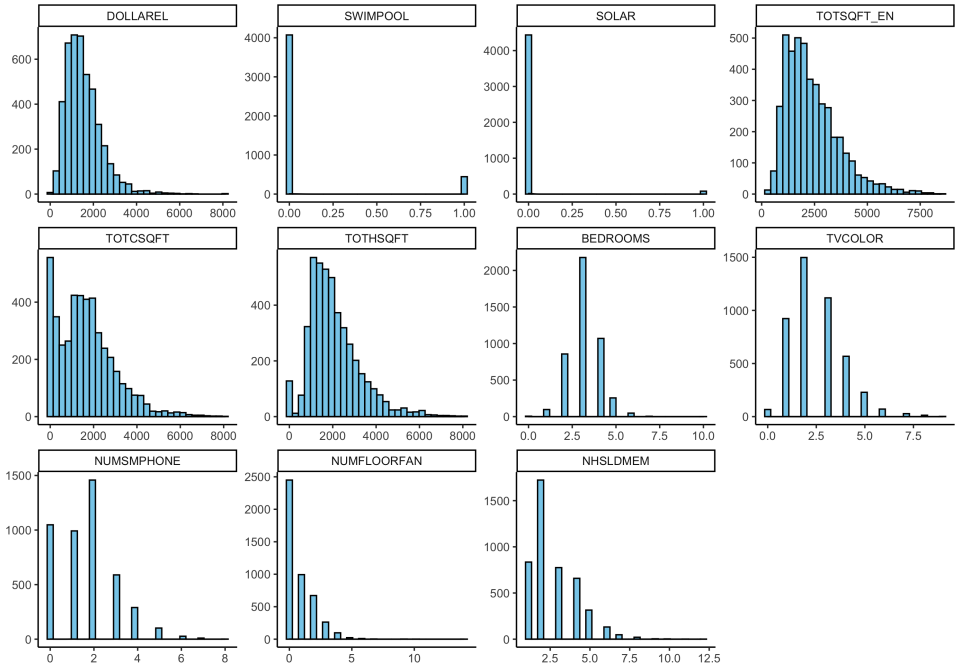

Figure 3: Distributions of the response (DOLLAREL) and covariates of the RECS dataset used in the study.

response variable (DOLLAREL) is strictly positive, with a heavy right tail. The dummy variables SWIMPOOL and SOLAR are zero almost everywhere, meaning that not a big number of households have swimming pools (10%) and solar electricity generation (2%). The square feet footage (TOTSQFT_EN) and the heated and cooled parts have the same distribution as the response, positive and high tail to the right. For the cooled and heated indicators, the variables are not strictly positive, as 0-valued observations appear represented. For the remaining covariates, we want to remark their median, which can provide an "average" representation of the U.S. population: households with 3 bedrooms, 2 television, 2 smartphones, 0 fans and 2 household members.

4 Results

We present the results obtained from our proposed model, the Non-Crossing Dual Neural Network (NCDNN), and compare it against the Monotone Composite Quantile Regression Neural Network (MCQRNN) for quantile levels 0.9, 0.925, 0.95, 0.975 and 0.99.

We took some decisions for the technical implementation: We implemented an early stopping criteria based on the gradient evaluation of the second part of the structure, the CTE calculation, as we are not

using a train-test environment. Secondly, we weighted the loss evaluation of our model, making the use of a weighting factor available for our model too. We also trained 50 different seeds for the NCDNN to show consistency across random initial values (more detail in Section 4.3). The choosing of the best model was done by averaging normalized losses across quantile levels and taking the best performing one.

4.1 Murphy diagrams for model comparison

Selecting a fair scoring function is a critical decision for model comparison, as different choices will affect the ranking of different models (Patton, 2020). For solving this issue, Ehm et al., 2016 presented the Murphy diagrams to compare model predictions without relying into the scoring function. The Murphy diagrams represent an evaluation of each point estimation across the space of consistent scoring loss functions, making possible a comparison between models that do not rely on the scoring structure chosen. The parameter θ present in the x axis of those plots represent a threshold of the comparison of both predictions and the response variable. For a more detailed explanation, see Ehm et al., 2016.

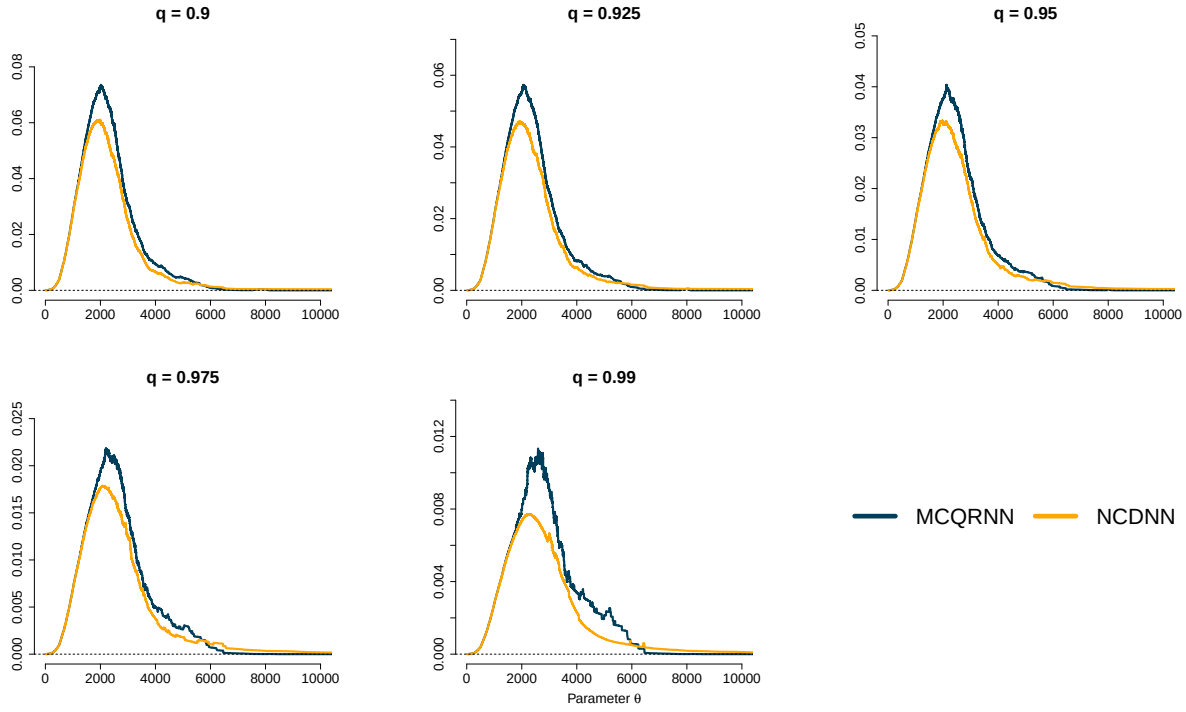

Figure 4: Murphy diagrams for estimated CTE associated with quantile levels 0.9, 0.925, 0.95, 0.975 and 0.99 to compare both models MCQRNN and NCDNN. Lower values of the distribution mean better approximation.

Figure 4 presents the Murphy diagrams for the calculated quantile levels. A lower value states a better prediction. We observe that for all quantile levels, the NCDNN appears below the MCQRNN, which indicates that, consistently, our model predicts more accurately the CTE of the observed data.

4.2 Comparison of losses

Despite the usefulness of the Murphy diagrams for model choosing, we cannot quantify the amount of improvement on performance of using the NCDNN against the MCQRNN, as the Murphy diagram is only used as a graphical tool for determine a better model. For solving this issue, we evaluate the loss functions (3) and (6) over both model predictions and aggregate adding their loss. It is important to remember that, for both scoring functions, a lower value represents a better approximation.

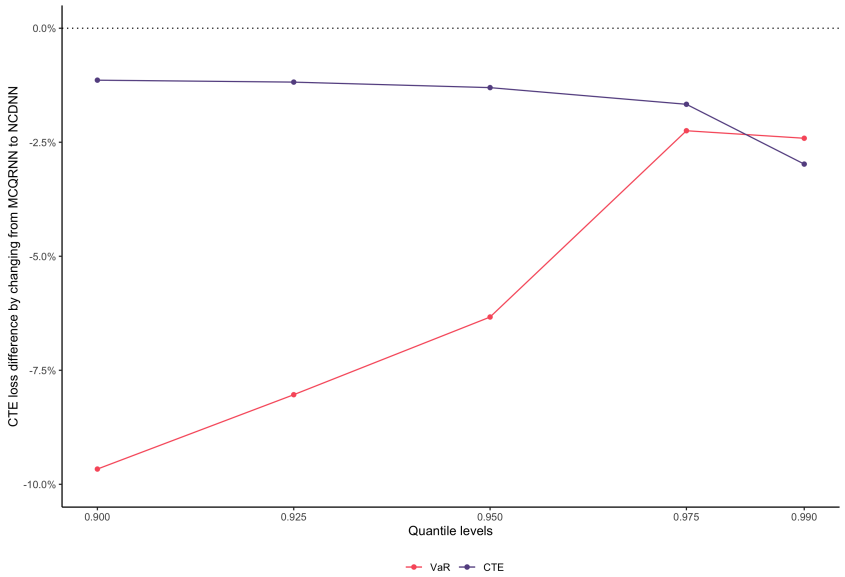

Figure 5: Percentage difference of overall loss calculated across the whole dataset between both models MCQRNN and NCDNN, for predicted VaR and CTE and for quantile levels 0.9, 0.925, 0.95, 0.975 and 0.99. Negative values mean NCDNN reduces loss against MCQRNN.

Figure 5 shows the percentage difference between models of the aggregated loss over the whole dataset for each calculated quantile level. More precisely, the values plotted are:

$$\left\{ \frac{\sum_i \rho_q(\mathrm{VaR}_q^{NCDNN},\ y_i)}{\sum_i \rho_q(\mathrm{VaR}_q^{MCQRNN},\ y_i)} - 1 \right\}_q, \quad \left\{ \frac{\sum_i S_q^{AL}(\mathrm{VaR}_q^{NCDNN}, \mathrm{CTE}_q^{NCDNN}, y_i)}{\sum_i S_q^{AL}(\mathrm{VaR}_q^{MCQRNN}, \mathrm{CTE}_q^{MCQRNN}, y_i)} - 1 \right\}_q$$

We plot the difference on percentage values on Figure 5. Negative results represent that NCDNN has a lower loss value across the dataset in comparison with the MCQRNN, meaning a better approximation.

It is shown that NCDNN improves between a 2.5% and a 10% the loss for the VaR and around a 2% the CTE calculated in comparison with the MCQRNN. Is also of note that, for the VaR, the improvement of using the NCDNN against the MCQRNN is reduced when reaching higher quantile levels, starting from an improvement of 10% for quantile level 0.9 and reaching a point of stability improvement on the two latter quantile levels of around a 2.5%. On the opposite, the CTE improvement of the NCDNN gets greater for bigger quantile levels, reaching from a 1% of improvement on quantile level 0.9 to a 3% of improvement on quantile level 0.99.

It is also important to remark the computational time spent for the calculation of all quantile levels. For the MCQRNN, calculating the grid of quantile levels 0.5, 0.6 ,0.7, 0.8, 0.9, 0.905, 0.91, ..., 0.985, 0.99 (23 quantile levels) lasts around 1 hour and 3 minutes to run, with the standard parameters that [Cannon, 2018] use relying on previous literature, most specifically [Xu et al., 2017], which propose a structure of two hidden layers with 4 nodes each one for Composite Quantile Regression with neural networks. Our proposed model improves computational time to the point of about two and a half minutes per model, for calculating the grid of quantile levels 0.5, 0.6 ,0.7, 0.8, 0.9, 0.925, 0.95, 0.975 and 0.99 (9 quantile levels). The difference on grids is not relevant as we tested that the convergence time does not depend on the number of quantile levels for grids with five or more quantile levels. The CTE calculation for MCQRNN depends on the VaR predictions. To compare with MCQRNN in ideal conditions, we duplicated the number of estimations for quantile levels over 0.9, which should result in better CTE predictions coming from the MCQRNN.

4.3 Robustness across quantiles: Distribution of exceedances

As a test of calibration of the models, we calculated the number of observed values for the VaR that surpassed its prediction. Beforehand, we expect $(1-q)\cdot 100\%$ observations to surpass its predicted VaR, in other words, predicting a quantile level q implies that $(1-q)\cdot 100\%$ of the predictions will be expected to be below its observed value. We show the results in Figure 6.

In Figure 6 we observe that, while MCQRNN % of exceedances are below the expected value for quantile levels 0.9, 0.925 and 0.95, NCDNN % of exceedances are above the expected number of exceedances for all calculated quantiles, but no model seems to be closer to the expected value of exceedances. For latter quantile levels, MCQRNN starts to increment the proportion of exceedances to the point of crossing

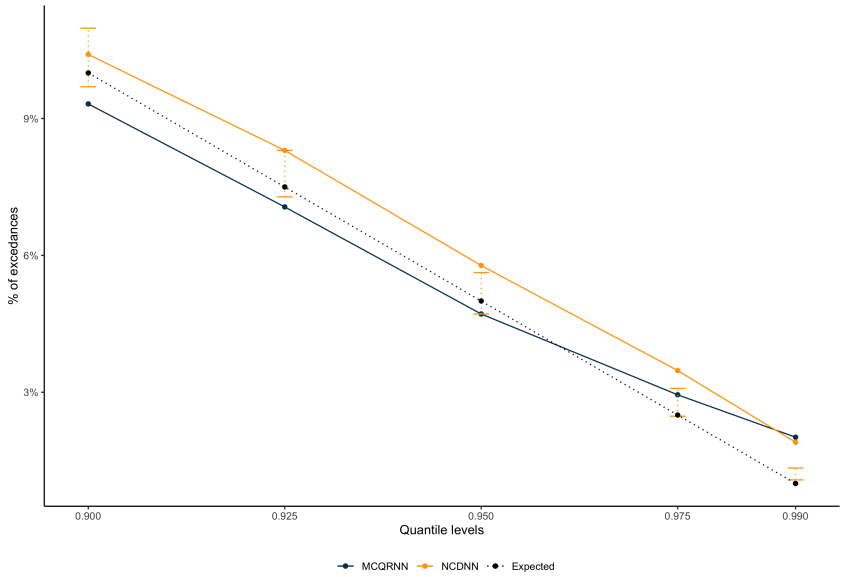

Figure 6: Percentage of exceedances (% of observed values that surpassed its predicted VaR) for models MCQRNN and NCDNN and for quantile levels 0.9, 0.925, 0.95, 0.975 and 0.99. In dotted line appears the expected value $1 - q$.

the NCDNN on quantile level 0.99. While both models appear rational on their quantile calculations, we want to point out that NCDNN shows greater stability than the MCQRNN across quantile levels, as the number of exceedances maintain invariable in comparison of the expected value. We added to the figure confidence intervals indicating the 0.25 and 0.75 quantile levels of the exceedances obtained across the 50 models trained for the NCDNN. While for quantile levels 0.9 and 0.925, the best performing NCDNN model is inside the confidence interval, for the following quantile levels it is above. We can extract two bullets from this result. 1) A random training of the NCDNN will give a consistent result when talking about exceedances of the VaR, as the intervals are near the expected value and get smaller as long as we predict greater quantile levels, and 2) As the NCDNN is optimizing both VaR and CTE, selecting the best performing model can result into a variation from the expected value of the exceedances of the VaR in order to improve the performance of the CTE, as we show in Figure 5, in which the CTE loss improvement against the MCQRNN gets bigger for greater quantile levels.

5 Conclusions

We define non-crossing properties involving several quantile levels for Value at Risk (VaR) and Conditional Tail Expectation (CTE) that have not been approached in literature. Also, to calculate them, we tried several models, linear and generalized linear VaR and CTE separately and jointly and exceedances with Machine Learning Models, like Random Forest and Gradient Boosting. But, the only model that asserts the abovementioned non-crossing conditions is the Monotone Composite Quantile Regression Neural Network (MCQRNN). As the MCQRNN does not directly model the CTE, we propose the Non-Crossing Dual Neural Network (NCDNN), a neural network model that is capable of estimating VaR and CTE for several quantile levels while asserting the non-crossing conditions. We used the MCQRNN and the NCDNN models for predicting VaR and CTE for quantile levels 0.9, 0.925, 0.95, 0.975 and 0.99 for a dataset of energy consumption of U.S. households from 2015.

NCDNN has some advantages when comparing it with the MCQRNN approach to VaR and CTE specification, like the ability to pick different loss functions for the CTE specification, a very reduced computational time, and an improvement in precision for large quantile levels. It should be remarked that MCQRNN was not initially created for calculating CTE and, to the best of our knowledge, there are no models that fulfill all non-crossing conditions and have a joint prediction of VaR and CTE.

We believe that the NCDNN can be useful for different disciplines, like forecasting extreme values of a firm's stock return, predicting a driver's future behavior to create a pay-as-you-go scheme with telematics data, or calculating possible floods on nearby cities by calculating riverbed widths.

Acknowledgements

We want to have a special acknowledgement to Fundació Banc Sabadell: "Ajudes a la investigació 2022", Fundación BBVA: "Ayudas a proyectos de investigación en Big Data", AGAUR: "PANDÈMIES" grant and the Spanish Ministry of Science grant PID2019-105986GB-C21 for their support to our research.

References

[Acerbi and Szekely, 2014] Acerbi, C. and Szekely, B. (2014). Back-testing expected shortfall. *Risk*, 27(11):76–81.

[Basel Committee on Banking Supervision, 2016] Basel Committee on Banking Supervision (2016). Minimum capital requirements for market risk.

[Bodnar et al., 1998] 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.

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

[Cannon, 2018] 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.

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

[Dette and Volgushev, 2008] 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 and Bayer, 2019] Dimitriadis, T. and Bayer, S. (2019). A joint quantile and expected shortfall regression framework. *Electronic Journal of Statistics*, 13(1):1823–1871.

[Ehm et al., 2016] Ehm, W., Gneiting, T., Jordan, A., and Krüger, F. (2016). Of quantiles and expectiles: consistent scoring functions, Choquet representations and forecast rankings. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(3):505–562.

[EIA, 2015] EIA, U. (2015). Residential energy consumption survey (RECS). 2015. https://www.eia.gov/.

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

[Freund and Schapire, 1995] Freund, Y. and Schapire, R. E. (1995). A desicion-theoretic generalization of on-line learning and an application to boosting. *European conference on computational learning theory*, 1:23–37.

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

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

[Guillen et al., 2020] 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.

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

[Ho, 1995] Ho, T. K. (1995). Random decision forests. *Proceedings of 3rd international conference on document analysis and recognition*, 1:278–282.

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

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

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

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

[Moon et al., 2021] 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 and Ziegel, 2017] Nolde, N. and Ziegel, J. F. (2017). Elicitability and backtesting: Perspectives for banking regulation. *The Annals of Applied Statistics*, 11(4):1833–1874.

[Patton, 2020] Patton, A. J. (2020). Comparing possibly misspecified forecasts. *Journal of Business & Economic Statistics*, 38(4):796–809.

[Patton et al., 2019] 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.

[Takeuchi and Furuhashi, 2004] 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, 2019] 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, 2020] Taylor, J. W. (2020). Forecast combinations for value at risk and expected shortfall. *International Journal of Forecasting*, 36(2):428–441.

[Vaswani et al., 2017] 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.

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

[White, 1992] White, H. (1992). *Nonparametric estimation of conditional quantiles using neural networks*. Computing Science and Statistics. Springer. New York, NY.

[Xu et al., 2017] Xu, Q., Deng, K., Jiang, C., Sun, F., and Huang, X. (2017). Composite quantile regression neural network with applications. *Expert Systems with Applications*, 76:129–139.

[Yu et al., 2003] 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, 2016] Zhang, L. (2016). Flood hazards impact on neighborhood house prices: A spatial quantile regression analysis. *Regional Science and Urban Economics*, 60:12–19.

[Ziegel et al., 2020] 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.

UBIREA
Institut de Recerca en Economia Aplicada Regional i Pública
Research Institute of Applied Economics
Universitat de Barcelona
Av. Diagonal, 690 • 08034 Barcelona

WEBSITE: www.ub.edu/irea • **CONTACT:** irea@ub.edu