Gulf Journal of Mathematics
Vol 20, Issue 2 (2025) 147-161

ISSN: 2309-4966
https://doi.org/10.56947/gjom.v20i2.3255

NEURAL NETWORK AND LEAST SQUARE ESTIMATOR

GHANIA IDIOU[1*] and HOUDA BOUREZAZ[2] and ILHEM LAROUSSI[3]

Abstract. We propose an alternative and a methodology that joins the expressive power of neural networks with the robustness of least squares estimation to address regression problems involving a twice-censored setting. Within the context of modern statistical learning, this approach is particularly well suited for capturing complex and non-linear relationships. We establish mean squared error norm convergence of the proposed estimator and present a comparative simulation study demonstrating its superior performance over standard neural network methods.

*Keywords.*Neural Networks, Least Squares Estimation, Mixed Censored Data, Regression Function, Nonlinear Modeling, Survival Analysis.
2020 Mathematics Subject Classification. Primary 46L55; Secondary 44B20

1. Introduction

Artificial neural networks have become essential tools in modern data processing. Designed to learn from observations and make predictions, they owe their success largely to their capacity to model difficult and nonlinear relationships. The first mathematical model, introduced by McCulloch and Pitts (1943) [14], aimed to represent the behavior of biological neurons. In 1958 [20], Rosenblatt's perceptron marked an important milestone, despite its limitations in handling nonlinear problems. The limitations pointed out by Minsky and Papert (1969) [16] contributed to a temporary decline in neural network research. However, the following decades saw the emergence of new approaches, such as adaptive linear neurons (Widrow and Hoff, 1960 [20]), and, most notably, The revival of neural networks was largely due to the backpropagation algorithm introduced by Rumelhart and al. (1986) [21], which enabled the efficient training of multilayer networks. While it is impossible

Date: Received: Jul 10, 2025; Accepted: Aug 3, 2025.
*Corresponding author

to cite the extensive body of work that followed, we highlight a few landmark contributions that significantly advanced the field: Hinton (2006) [8] on deep belief networks, Krizhevsky and al. (2012) [11] with ImageNet item, Goodfellow and al. (2014) [6] with generative adversarial networks, and Vaswani and al. (2017) [22] with the Transformer architecture. These developments have collectively pushed the boundaries of artificial intelligence. In the same time with these advances, adapting neural networks to handle censored data has become a growing part of interest, particularly in survival analysis and reliability studies, where censoring including right, left, or interval censoring presents a major challenge. For example, Nagpal and al. (2020) [17] established Deep Survival Machines for censored data with competing risks, while Pearce and al. (2022) [24] suggested a quantile regression model in nonparametric case. More recently, Cordeiro and al. (2024) [3] presented a regression method for interval-censored data with automatic variable selection through sparse architectures. These works illustrate the emergent diversity and efficiency of neural networks in handling censored data, improving predictions and the utilization of partial information. Several studies have extended the use of least squares estimation to various types of censored data by adapting it to different functional classes. Bagirov and al. (2009) [2] proposed modified least squares estimators for bounded random variables, accounting for the effects of censoring. Kebabi and al. (2011) [9] developed least squares estimators specifically designed for mixed censoring situations, and established their convergence in the L_2 norm. Contributions that are more recent have concentrated on expanding the range of function spaces used. In 2021, Laroussi [12] introduced a penalized smoothing spline estimator for censored regression, while Douas and al. (2023) [4] proposed a wavelet-based estimator adapted to twice-censored data. In 2024, Laroussi [13] advanced this line of work by developing an estimator based on B-spline functions, along with a detailed asymptotic risk analysis. These approaches demonstrate the versatility and robustness of least squares estimation across different censoring mechanisms and functional frameworks. Inspired by years of development in neural network architectures and their recent applications to censored data, this study presents a novel methodology that combines the suppleness of neural networks with the robustness of least squares estimation for regression problems involving mixed censoring. We begin by presenting the structure of neural function spaces, followed by the construction of a least squares estimator that includes censoring mechanisms. Theoretical properties, including convergence in the L_2 norm, are established, and a comparative simulation study is conducted to evaluate the estimator's performance against existing methods.

2. Neural network space

Consider a random vector X in $\mathbb{R}^d$, and consider Y as a random variable that is not negative and is bounded by K. We specify a neural network model that has one hidden layer that contains k_n neurons, represented by the function:

$$\sum_{i=1}^{k_n} \zeta_i \ \sigma(\varpi_i^t x + \nu_i) + \zeta_0, \tag{2.1}$$

the input vector is represented by X and Y denotes a non negative output. Initial weight transpored matrix ϖ_i connects input layer and hidden layer, ν_i is a bias vector of a cells in hidden layer. Similarly, ζ_i represents the weight matrix linking hidden layer and output layer and ζ_0 serves in the form of bias vector for the output layer's cells (see figure 1). We denotes $\sigma(.)$ as the activation function that introduces non-linearity into the model, enabling neural networks to learn complex patterns and relationships in the data. Activation functions are generally classified into three main types: binary step, linear, and non-linear. Binary step and linear functions are simple, while non-linear functions such as the sinus and cosinus functions, sigmoid, arctangent, logistic, and Gaussian squashers offer greater complexity. The choice of activation function depends on a particular application and the neural network's intended behavior.

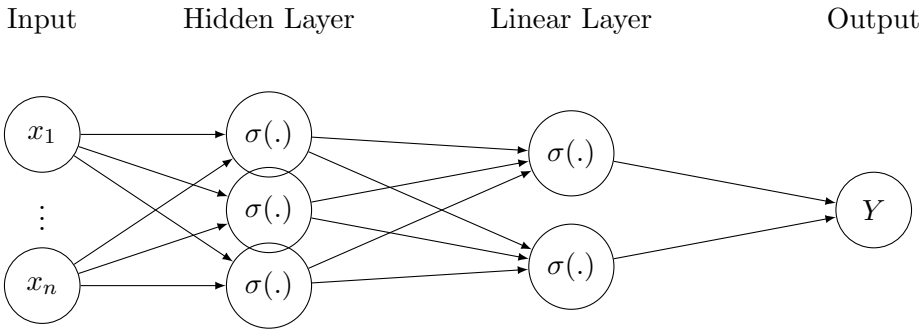

Figure 1. Neural network with hidden layer.

Let's consider the neural network class that is defined by

$$\pounds_n = \left\{ \sum_{i=1}^{k_n} \zeta_i \ \sigma\left(\varpi_i^t X + \nu_i\right) + \zeta_0, \ k_n \in \mathbb{N}, \ \varpi_i^t \in \mathbb{R}^d, \ \nu_i \in \mathbb{R}, \ \sum_{i=0}^{k_n} \zeta_i \leq Ł_n \right\} \tag{2.2}$$

we pose $\varpi_i^t X + \nu_i = \sum_{j=1}^{d} \upsilon_{i,j} \ X^j + \upsilon_{i,0}$ and $\varpi^t = (\upsilon_{i,j})_{i,j}$.

All functions g bounded by $L_n > 0$ are represented by this set. Let R and L be non-negative censoring random variables, and assume that Y is not fully observed. To present estimators of $q(x) = E(Y \mid X = x)$ in the context of i.i.d sampled data $D_n = \{(X_i, Z_i = \max(\min(Y_i, R_i), A_i)); 1 \leq i \leq n\}$, where A is the censoring indicator defined as

$$A = \begin{cases} 0 & \text{if } L < Y < R \\ 1 & \text{if } L < R \leq Y \\ 2 & \text{if } \min(Y, R) \leq L \end{cases}.$$

Subsequently we will denote by F_L and S_R respectively the distribution and the survival functions and $r = \sup\{\tau : S(\tau) > 0\}$ the end point of the support.

The least squares estimator, adapted to censored data, is defined as $\tilde{q}_n(.)$ the function minimizing the empirical loss L_2. This approach, recently introduced by Laroussi (2024) [13], extends traditional least squares estimation to handle incomplete data by incorporating censoring constraints. The estimator is designed to approximate the true regression function while ensuring robustness in the presence of twice-censored observations.

$$\begin{aligned} \tilde{q}_n(.) &= \tilde{q}_n(.\ , D_n) \\ &= \arg\min_{g \in \pounds_n} \frac{1}{n} \sum_{i=1}^{n} 1_{\{A_i = 0\}} \frac{|g(X_i) - Z_i|^2}{S_n(Z_i) F_n(Z_i)}. \end{aligned} \tag{2.3}$$

Where S_n (proposed by Patilea and Rolin (2006) [18]) is an estimation of the survival (reliability) function S_R of R defined as

$$S_n(t) = \prod_{j : Z'_j \leq t} \left(1 - \frac{D_{1j}}{U_{j-1} - N_{j-1}}\right),$$

with $(Z'_j)_{1 \leq j \leq M}$ the list of all unique values you get from the Z_i, sorted from lowest to highest, where $(M \leq n)$. Also

$$U_{j-1} = n \prod_{l=1}^{M} \left(1 - \frac{D_{2l}}{N_l}\right),$$

for

$$D_{kj} = \sum_{i=1}^{n} 1_{\{Z_i = Z'_j, A_i = k\}}, \quad N_j = \sum_{i=1}^{n} 1_{\{Z_i \leq Z'_j\}}.$$

So, the Kaplan-Meier estimator, in the presence of left-censoring, spits out this distribution function usually call it F_L for the random variable L, is defined as follows

$$F_n(t) = \prod_{j:Z'_j>t} \left(1 - 1_{\{A_j=2\}}\right).$$

As Y is almost surely bounded by $0 \leq Y \leq Ł_n < \infty$, $q(x)$ is constrained to the interval $[0, Ł_n]$ for any $x \in \mathbb{R}^d$. To ensure the stability and boundedness of the estimates, we impose suitable constraints on the function space. Specifically, we define

$$q_n(x) = q_n(x, D_n) = \max(0, \min(\tilde{q}_n(x), K)), \tag{2.4}$$

with $0 \leq K \leq Ł_n < \infty$.
In a similar manner of (2.4), we apply a truncated version of our estimator, which is defined by

$$q_n{}^*(x) = \max(0, \min(\tilde{q}_n(x), Ł_n)). \tag{2.5}$$

In the following, we introduce the assumptions that will be considered throughout the analysis.

- **H1**: Y, R and L are assumed to be statistically independent,
- **H2**: The censoring mechanism (R, L) is assumed to be independent of the data (X, Y),
- **H3**: F_L , S_R continuous on $]0, \infty[$,
- **H4**: $\exists\ a = \inf(\tau)_{\tau\in\mathbb{R}^+}$ such that $F_L(a) > 0, \exists b = \sup(\tau)_{\tau\in\mathbb{R}^+}$ and $S_R(b) > 0$, and $\forall i, a < Z_i < b$ for $A_i = 0$.

Hypothesis $(H4)$ appears to be reasonable since $a < Z_i < b$ whenever $A_i = 0$. Specifically, this condition ensures that the model parameters can be uniquely identified.

Assuming hypotheses $(H1)$ and $(H4)$ hold, Patilea and Rolin (2006) [18] demonstrated that

$$\lim_{n\to\infty} \sup_{t\in\mathbb{R}_+} |S_n(t) - S_R(t)| = 0 \text{ a.s.},$$

and

$$\lim_{n\to\infty} \sup_{t\in\mathbb{R}_+} |F_n(t) - F_L(t)| = 0 \text{ a.s.}.$$

According to assumption $(H4)$, it follows that, for sufficiently large n , we have $S_n(b) > 0$ and $F_n(a) > 0$ almost surely.

3. Result of the Censored Neural Network Estimator Using the Regression Function

The censored neural network estimator exhibits desirable asymptotic properties.

Theorem 3.1. *Under assumptions H1:H4 and given that k_n and $Ł_n$ satisfy*

$$k_n \to \infty,\ Ł_n \to \infty \text{ and } \frac{k_n\ Ł_n^4\ log(k_n\ Ł_n^2)}{n} \to 0 \tag{3.1}$$

Then $E\int (q_n(x)-q(x))^2\,\wp(dx) \to 0$ a.s, for $n \to \infty$, with $E|Y|^2 < \infty$ and $\wp$ denotes the distribution of X.
In addition, $\exists \lambda > 0$ such that $\frac{Ł_n^4}{n^{1-\lambda}} \to 0$ then, for $n \to \infty$, $\int (q_n(x)-q(x))^2\,\wp(dx) \to 0$ a.s.

Proof. In the context of twice censored data in neural networks, we define the sets as follows:

$$\psi_n^* \pounds_n = \{g \in \pounds_n : 0 \le g(x) \le Ł_n, x \in \mathbb{R}^d\},$$

$$\pounds_n^* \pounds_n = \left\{f : \mathbb{R}^d \to \mathbb{R}_+, f(x) = \max(0, \min(g(x), Ł_n)), for\ some\ g \in \pounds_n\right\},$$

and $\mathcal{H}_n = \{h : \mathbb{R}^d \times [0, Ł_n] \times \{0,1\} \to \mathbb{R}_+ : \exists g \in \pounds_n^* \pounds_n \quad \text{with} \quad h(x,z,A) = 1_{\{A_i=0\}} \frac{|g(x)-z|^2}{S_R(z)F_L(z)}, \forall (x,z,A) \in \mathbb{R}^d \times [0, Ł_n] \times \{0,1\}\}$.

We have

$$\sup_{g \in \pounds_n^* \pounds_n} \left| \frac{1}{n}\sum_{i=1}^n 1_{\{A_i=0\}} \frac{|g(X_i)-Z_i|^2}{S_n(Z_i)F_n(Z_i)} - E|g(X)-Y|^2 \right| = \sup_{h \in \mathcal{H}_n} \left| \frac{1}{n}\sum_{i=1}^n h(W_i) - Eh(W) \right|, \tag{3.2}$$

where

$$Eh(W) = E\left[E\left(1_{\{A_i=0\}} \frac{|g(X)-Z|^2}{S_R(Z)F_L(Z)} \mid X, Y\right)\right] = E(|g(X)-Z|^2).$$

Such that : $W = (X, Z, A), W_1 = (X_1, Z_1, A_1), ..., W_n = (X_n, Z_n, A_n)$, n random vectors i.i.d with W distribution.
The demonstration of the theorem is grounded using the inequality below.

$$\int_{\mathbb{R}^d} |q_n^*(x) - q(x)|^2\,\wp(dx) \le 2 \sup_{g \in \pounds_n^* \pounds_n} \left| \frac{1}{n}\sum_{i=1}^n 1_{\{A_i=0\}} \frac{|g(X_i)-Z_i|^2}{S_n(Z_i)F_n(Z_i)} - E|g(X)-Y|^2 \right| + \inf_{g \in \psi_n^* \pounds_n} \int_{\mathbb{R}^d} (g(x)-q(x))^2\,\wp(dx). \tag{3.3}$$

Let us begin by establishing this inequality. As a first step, we observe that

$$\int_{\mathbb{R}^d} |q_n^*(x) - q(x)|^2 \wp(dx) = \Big\{ E\big(|q_n^*(X) - Y|^2 \mid D_n\big) - \inf_{g \in \psi_n^* \pounds_n} E\big(|g(X) - Y|^2\big) \Big\} + \inf_{g \in \psi_n^* \pounds_n} E\big(|g(X) - Y|^2\big) - E\big(|q(X) - Y|^2\big). \tag{3.4}$$

And the regression function checks

$$\inf_{g \in \psi_n^* \pounds_n} E|g(X) - Y|^2 - E|q(X) - Y|^2 = \inf_{g \in \psi_n^* \pounds_n} \int_{\mathbb{R}^d} (g(x) - q(x))^2 \wp(dx).$$

On the other hand

$$\begin{aligned}
&E\left(|q_n^*(X) - Y|^2 \mid D_n\right) - \inf_{g \in \psi_n^* \pounds_n} E\left(|g(X) - Y|^2\right) \\
&= \sup_{g \in \psi_n^* \pounds_n} \Bigg[E\left(|q_n^*(X) - Y|^2 \mid D_n\right) - \frac{1}{n} \sum_{i=1}^n 1_{\{A_i = 0\}} \frac{|q_n^*(X_i) - Z_i|^2}{S_n(Z_i) F_n(Z_i)} \\
&\quad + \frac{1}{n} \sum_{i=1}^n 1_{\{A_i = 0\}} \frac{|q_n^*(X_i) - Z_i|^2}{S_n(Z_i) F_n(Z_i)} - \frac{1}{n} \sum_{i=1}^n 1_{\{A_i = 0\}} \frac{|\tilde{q}_n(X_i) - Z_i|^2}{S_n(Z_i) F_n(Z_i)} \\
&\quad + \frac{1}{n} \sum_{i=1}^n 1_{\{A_i = 0\}} \frac{|\tilde{q}_n(X_i) - Z_i|^2}{S_n(Z_i) F_n(Z_i)} - \frac{1}{n} \sum_{i=1}^n 1_{\{A_i = 0\}} \frac{|g(X_i) - Z_i|^2}{S_n(Z_i) F_n(Z_i)} \\
&\quad + \frac{1}{n} \sum_{i=1}^n 1_{\{A_i = 0\}} \frac{|g(X_i) - Z_i|^2}{S_n(Z_i) F_n(Z_i)} - E\left(|g(X) - Y|^2\right) \Bigg] \\
&= \sum_{i=1}^4 U_{n,i}.
\end{aligned}$$

We note that $a < Z_i < b$ for $A_i = 0$ and $1 \le i \le n$, then

$$\begin{aligned}
|\tilde{q}_n(X_i) - Z_i| &= |(q_n^*(X_i) - Z_i) + (\tilde{q}_n(X_i) - q_n^*(X_i))| \\
&\le |q_n^*(X_i) - Z_i| + |\tilde{q}_n(X_i) - q_n^*(X_i)|.
\end{aligned}$$

Since $q_n^* \in \pounds_n^* \pounds_n$ and $g \in \psi_n^* \pounds_n \subset \pounds_n^* \pounds_n$, it's clear that

$$\begin{aligned}
U_{n,1} &= \sup_{g \in \psi_n^* \pounds_n} \left[E\left(|q_n^*(X) - Y|^2 \mid D_n\right) - \frac{1}{n} \sum_{i=1}^n 1_{\{A_i = 0\}} \frac{|q_n^*(X_i) - Z_i|^2}{S_n(Z_i) F_n(Z_i)} \right] \\
&\le \sup_{g \in \pounds_n^* \pounds_n} \left| \frac{1}{n} \sum_{i=1}^n 1_{\{A_i = 0\}} \frac{|g(X_i) - Z_i|^2}{S_n(Z_i) F_n(Z_i)} - E|g(X) - Y|^2 \right|,
\end{aligned} \tag{3.5}$$

and

$$U_{n,4} = \sup_{g \in \pounds_n^* \pounds_n} \left| \frac{1}{n} \sum_{i=1}^{n} 1_{\{A_i=0\}} \frac{|g(X_i) - Z_i|^2}{S_n(Z_i) F_n(Z_i)} - E|g(X) - Y|^2 \right|. \tag{3.6}$$

Given that $q_n^*(X_i) \leq Ł_n$ and $Z_i > a$, it follows that

$$1_{\{A_i=0\}} |\tilde{q}_n(X_i) - Z_i| \leq 1_{\{A_i=0\}} |q_n^*(X_i) - Z_i|,$$

This confirms

$$U_{n,2} = \sup_{g \in \psi_n^* \pounds_n} \left[\frac{1}{n} \sum_{i=1}^{n} 1_{\{A_i=0\}} \frac{|q_n^*(X_i) - Z_i|^2}{S_n(Z_i) F_n(Z_i)} - \frac{1}{n} \sum_{i=1}^{n} 1_{\{A_i=0\}} \frac{|\tilde{q}_n(X_i) - Z_i|^2}{S_n(Z_i) F_n(Z_i)} \right] \leq 0. \tag{3.7}$$

Indeed, according to the definition of (2.4), it's evident that

$$U_{n,3} = \sup_{g \in \pounds_n^* \pounds_n} \left[\frac{1}{n} \sum_{i=1}^{n} 1_{\{A_i=0\}} \frac{|\tilde{q}_n(X_i) - Z_i|^2}{S_n(Z_i) F_n(Z_i)} - \frac{1}{n} \sum_{i=1}^{n} 1_{\{A_i=0\}} \frac{|g(X_i) - Z_i|^2}{S_n(Z_i) F_n(Z_i)} \right] \leq 0. \tag{3.8}$$

Then we obtained the inequality (3.4). It remains to show that the two terms on the right-hand side of the equation converge to zero almost surely as $n \to \infty$. Since the right-hand side is independent of censoring, and under assumptions (3.1), we follow the approach of Györfi and al. (2002), p. 271 [7], to establish the result.

Finally, we need to determine that

$$\lim_{n \to \infty} \sup_{g \in \psi_n^* \pounds_n} \left[\frac{1}{n} \sum_{i=1}^{n} 1_{\{A_i=0\}} \frac{|g(X_i) - Z_i|^2}{S_n(Z_i) F_n(Z_i)} - E|g(X) - Y|^2 \right] = 0 \ a.s.$$

As any $x \in \mathbb{R}^d$ and $g \in \pounds_n^* \pounds_n$ implies that $0 \leq g(X) \leq Ł_n$, so, the functions in $\mathcal{H}_n$ are non negative and bounded. Since $Ł_n \geq \beta$, the functions in $\mathcal{H}_n$ justify

$$0 \leq h(x, z, A) \leq \frac{2Ł_n^2 + 2\beta^2}{S_R(a) F_L(b)} \leq \frac{4Ł_n^2}{S_R(b) F_L(b)}.$$

Our goal is to uniformly bound the deviation between an empirical average and its expectation a given function class. For any arbitrary $\xi > 0$, we obtain:

$$
\begin{aligned}
&P\left\{\sup_{g\in\pounds_n}\left|\frac{1}{n}\sum_{i=1}^{n}1_{\{A_i=0\}}\frac{|g(X_i)-Z_i|^2}{S_R(Z_i)F_L(Z_i)}-E\{|g(X)-Z|^2\}\right|>\xi\right\}\\
&=P\left\{\sup_{g\in\mathcal{H}_n}\left|\frac{1}{n}\sum_{i=1}^{n}h(W_i)-E\{h(W)\}\right|>\xi\right\}\\
&\le 8E\mathcal{N}_1\left(\frac{\xi S_R(a)F_L(b)}{16Ł_n},H_n,W_1^n\right)\ e^{-\frac{n\xi^2S_R^2(a)F_L^2(b)}{128Ł_n^4}}\\
&\le 8\left(\frac{64eŁ_n^3}{\xi S_R(a)F_L(b)}\right)^{2\mathcal{V}_{\pounds_n^*\pounds_n^+}}e^{-\frac{n\xi^2S_R^2(a)F_L^2(b)}{128Ł_n^4}}.
\end{aligned}
$$

Let $h_i(x,z,A)=1_{\{A_i=0\}}\dfrac{|g_i(x)-z|^2}{S_R(z)F_L(z)}$ and lets $h_1,h_2\in\mathcal{H}_n$, for g_1,g_2 corresponding functions in $\pounds_n^*\pounds_n$, then we take the derivation in the proof and we get

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left|1_{\{A_i=0\}}\frac{|g_1(X_i)-Z_i|^2}{S_R(Z_i)F_L(Z_i)}-1_{\{A_i=0\}}\frac{|g_2(X_i)-Z_i|^2}{S_R(Z_i)F_L(Z_i)}\right|\\
&\le\frac{1}{S_R(a)F_L(b)}\frac{1}{n}\sum_{i=1}^{n}|(g_1(X_i)-g_2(X_i)-2Z_i)(g_1(X_i)-g_2(X_i))|\\
&\le\frac{4Ł_n}{S_R(a)F_L(b)}\frac{1}{n}\sum_{i=1}^{n}|(g_1(X_i)-g_2(X_i))|.
\end{aligned}
$$

Which implies that

$$
\mathcal{N}(\xi,\mathcal{H}_n,W_1^n)\le\mathcal{N}\left(\frac{\xi S_R(a)F_L(b)}{4Ł_n},\pounds_n^*\pounds_n,X_1^n\right).
$$

Thus

$$
\mathcal{N}_1\left(\frac{\xi}{8},\mathcal{H}_n,W_1^n\right)\le\mathcal{N}_1\left(\frac{\xi S_R(a)F_L(b)}{32Ł_n},\pounds_n^*\pounds_n,X_1^n\right).
$$

Where $\mathcal{N}(\xi, \mathcal{H}_n, W_1^n)$ and $\mathcal{N}_1\left(\frac{\xi}{8}, \mathcal{H}_n, W_1^n\right)$ are the covering number, for ξ arbitrary we have

$$
\begin{aligned}
&P\left\{\sup_{h\in H_n}\left|\frac{1}{n}\sum_{i=1}^{n}h(W_i)-E\{h(W)\}\right|>\xi\right\}\\
&\leq 8E\left\{\mathcal{N}\left(\frac{\xi S_R(a)F_L(b)}{32\text{Ł}_n},\pounds_n^*\pounds_n,X_1^n\right)\right\}e^{-\frac{n\xi^2S_R^2(a)F_L^2(b)}{2048\text{Ł}_n^4}}\\
&\leq 8\left(\frac{128e\text{Ł}_n^3}{\xi S_R(a)F_L(b)}\right)^{2\mathcal{V}_{\pounds_n^*\pounds_n^+}}e^{-\frac{n\xi^2S_R^2(a)F_L^2(b)}{128\text{Ł}_n^4}}\\
&\leq 8\left(\frac{384e\text{Ł}_n^2(k_n+1)}{\xi S_R(a)F_L(b)}\right)^{(2d+5)k_n+1}e^{-\frac{n\xi^2S_R^2(a)F_L^2(b)}{128\text{Ł}_n^4}}.
\end{aligned}
$$

This implies the first result.
For the second point, we use the same justification as in lemma 16.4 and lemma 16.5 of Györfi and al. (2002) [7] about the covering numbers of classes of functions whose members are the sums or products of functions from othernclasses. This implies the final result. □

4. Simulation study

In order to assess the practical performance of the proposed methods, a set of simulation experiments is conducted under twice censoring scenarios for exponential and Weibull laws. Several underlying input distributions (uniform and Gaussian) are used to generate the data, with varying levels of censoring and increasing sample sizes. Two main models are considered in the simulation study

- **First model:** the covariate matrix is generated as $X = \text{unifrnd}(-\pi, \pi, n, 10)$, with two sub-models based on different censoring schemes
  a) $R = \text{exprnd}(6, n, 1)$ and $L = \text{exprnd}(0.5, n, 1) - 11$,
  b) $R = \text{wblrnd}(6, 1, n, 1)$ and $L = \text{wblrnd}(1, 1, n, 1) - 11$.
- **Second model:** the covariates follow a standard normal distribution,$X = \text{normrnd}(0, 1, n, 10)$, with two sub-models
  a) $R = \text{exprnd}(17, n, 1)$ and $L = \text{exprnd}(0.5, n, 1) - 1.5$,
  b) $R = \text{wblrnd}(10, 2, n, 1)$ and $L = \text{wblrnd}(1.5, 8, n, 1) - 1.5$.

These scenarios are designed to reflect a range of censoring intensities, from mild to heavy, and allow us to evaluate the sensitivity and accuracy of the estimators in complex situations.

4.1. **Results.** The table (1) compares the performance of two estimation methods, Weighted Least Squares (WLS) and Neural Networks (NN) for different epochs (1000, 12100 and 15000). In the context of censored data under different censoring distributions (Exponential and Weibull), the performance is measured through the Mean Squared Error (MSE) for various sample sizes ($n = 100, 300, 1000$) and twice censoring schemes (defined by L, R, and the average number of right- and left-censored observations).

TABLE 1. Experimental results according to different distributions

X	L&R	n	Epochs	MSE WLS	MSE NN	Right	Left
			1000	1.3477	22.7548	16	9
		100	12100	0.1661	2.9844	17	8
			15000	0.4247	3.2824	15	9
			1000	1.6809	33.9083	23.67	9.33
	Exponential	300	12100	0.6424	4.741	21.67	12
			15000	0.7059	3.5512	21.33	10
			1000	1.2958	37.5387	19.4	11.2
		1000	12100	1.2185	3.3709	17	12.8
Uniform			15000	1.3116	3.0067	18.6	11.4
			1000	1.3934	31.9446	19	17
		100	12100	0.1425	2.7873	16	12
			15000	1.8490	6.2365	19	10
			1000	1.5396	34.0129	21.67	9.33
	Weibull	300	12100	0.6313	3.9781	18.67	11
			15000	1.8490	6.2365	19	10
			1000	1.3669	34.1693	20.4	10.9
		1000	12100	1.5665	3.4181	19.7	10.6
			15000	1.3170	2.9825	19	11.8
			1000	1.4021	10.5532	22	7
		100	12100	0.3271	3.0479	20	6
			15000	0.1065	3.2506	12	8
			1000	1.5915	12.7334	20.33	6.33
	Exponential	300	12100	1.2085	4.8760	18.67	5
			15000	1.5159	2.3497	21.67	4.33
			1000	1.8835	11.7661	18.8	6.1
		1000	12100	1.6878	2.9100	17.6	5.7
Gaussian			15000	2.1777	3.0019	19.6	7.1
			1000	1.3834	5.8166	20.67	6.33
		100	12100	0.1039	0.2621	18.4	5.8
			15000	0.0538	0.1644	18.4	5.8
			1000	1.0203	7.0328	20.67	6.33
	Weibull	300	12100	0.9319	1.1779	18.4	5.8
			15000	0.8479	1.0993	18.4	5.8
			1000	1.2247	6.9834	20.67	6.33
		1000	12100	1.4772	1.5065	18.4	5.8
			15000	1.2710	1.4202	18.4	5.8

4.2. **Interpretation and conclusion.** Across all censoring distributions, increasing the sample size results in a consistent decrease in MSE for both methods, particularly for the WLS estimator. This reduction in MSE is more regular and noticeable for WLS, suggesting its greater stability in large-sample settings. Moreover, it can be observed that for censoring distributions of the Weibull and exponential types, when the sample size is $n = 1000$, the mean squared error (MSE) is relatively high for neural networks, whereas it remains low for the weighted least squares method, which confirms the efficiency of the latter. However, the effect of the input Gaussian distribution lead to relatively moderate MSEs for both methods, especially in larger sample settings.

In conclusion, we show that the least squares estimator applied to the class of functions represented by neural networks yields consistently good performance, regardless of the distribution of the input variables and in complex censoring situations.

Acknowledgement. The authors would like to extend their gratitude to the editor of the journal and to the reviewers for their constructive comments and suggestions.

References

1. Arik, S. O., Pfister, T. (2021). Tabnet : Attentive interpretable tabular learning. Proceedings of the AAAI Conference on Artificial Intelligence 35(8):6679-6687. https://doi.org/10.48550/arXiv.1908.07442
2. Bagirov, A. M., Clausen, C., Kohler, M. (2009). Estimation of a Regression Function by Maxima of Minima of Linear Functions. IEEE Transactions on Information Theory, 55(2): 833-845. https://doi.org/10.1109/TIT.2008.2009835
3. Cordeiro, G. M., Rodrigues, G. M., Prataviera, F., Ortega, E. (2024). A new quantile regression model with application to human development index. Computational Statistics, 39(6), 2925-2948. http://dx.doi.org/10.1007/s00180-023-01413-w
4. Douas, M., Laroussi, I., & Kharfouchi, S. (2023). Incomplete least squared regression function estimator based on wavelets. Journal of Siberian Federal University. Mathematics & Physics, 16(2), 204-215. https://www.mathnet.ru/eng/jsfu1070
5. Fontenla-Romero, O., Erdogmus, D., Principe, J. C., Alonso-Betanzos, A., & Castillo, E. (2003). Linear Least-Squares Based Methods for Neural Networks Learning. Lecture Notes in Computer Science, 2714, 84-91. https://doi.org/10.1007/3-540-44989-2_11
6. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Networks. Advances in Neural Information Processing Systems (NeurIPS), 27, 2672-2680. https://doi.org/10.1145/3422622
7. Gyôrfi, L., Kohler, M., Krzyżak, A., & Walk, H. (2002). A Distribution-Free Theory of Nonparametric Regression, Springer-Verlag, New York. https://doi.org/10.1007/b97848
8. Hinton, G. E., Osindero, S., and Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. Neural Computation, 18(7), 1527-1554. https://doi.org/10.1162/neco.2006.18.7.1527

9. Kebabi, K., Laroussi, I., & Messaci, F. (2011). Least squares estimators of the regression function with twice censored data. Statistics and Probability Letters, 81(11): 1588-1593. https://doi.org/10.1016/j.spl.2011.06.010
10. Kohler, M., Máthé , K., & Pintér, M. (2002). Prediction from randomly right censored data. Journal of Multivariate Analysis, 80(1): 73-100. https://doi.org/10.1006/jmva.2000.1973
11. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Image net classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25, 84-91. https://doi.org/10.1145/3065386
12. Laroussi, I. (2021). A generalised censored least squares and smoothing spline estimators of regression function. International Journal of Mathematics in Operational Research, 20(4), 506-520. https://doi.org/10.1504/IJMOR.2021.120102
13. Laroussi, I. (2024). B-Spline estimate of the regression function under general censorship model. Jordan Journal of Mathematics and Statistics, 17(1), 2024, pp 179-197. https://doi.org/10.47013/17.1.11
14. McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, 5(4), 115-133. https://link.springer.com/article/10.1007/BF02478259
15. Meixide, G. C., Matabuena, M., Abraham, L., & Kosorok, M. R. (2024). Neural interval-censored survival regression with feature selection. Statistical Analysis and Modeling, 18(2), 123-145. https://doi.org/10.1002/sam.11704
16. Minsky, M., & Papert, S. (1969). Perceptrons: An Introduction to Computational Geometry. Cambridge, MA: MIT Press. https://www.bibsonomy.org/bibtexkey/minsky69perceptrons/sb3000
17. Nagpal, C., Li, X., & Dubrawski, A. (2021). Deep Survival Machines: Fully parametric survival regression and representation learning for censored data with competing risks. IEEE Journal of Biomedical and Health Informatics, 25(8), 3163-3175. https://doi.org/10.1109/JBHI.2021.3052441
18. Patilea, V. & Rolin, J.-M. (2006). Product-limit estimators of the survival function with twice censored data. Annals of Statistics, 34(2), 925-938. https://doi.org/10.1214/009053606000000065
19. Pearce, T. , Leibfried, F., Brintrup, A., Zaki, M., Neely, A. (2020). Uncertainty in neural networks: Approximately Bayesian ensembling. Proceedings of the 23^{rd} International Conference on Artificial Intelligence and Statistics (AISTATS), Volume 108. http://dx.doi.org/10.13140/RG.2.2.12913.43364
20. Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review, 65(6), 386-408. https://doi.org/10.1037/h0042519
21. Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323(6088), 533-536. https://doi.org/10.1038/323533a0
22. Vaswani, A., Shazeer, N., Parmar,N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., Polosukhin, I. (2017). Attention is all you need. 31^{st} Conference on Neural Information Processing Systems (NIPS 2017). https://doi.org/10.48550/arXiv.1706.03762
23. Widrow, B., & Hoff, M. E. (1960). Adaptive Switching Circuits. IRE WESCON Convention Record, Part 4, 96-104. https://doi.org/10.7551/mitpress/4943.003.0012

24. Zadeh, SH. G., Behning, C., Matthias Schmid, M. (2022). An imputation approach using subdistribution weights for deep survival analysis with competing events. 12(3815). https://doi.org/10.1038/s41598-022-07828-7

[1] Department of Mathematics, Mentouri Brothers University of Constantine 1, Constantine, Algeria.
Email address: ghania.idiou@umc.edu.dz

[2] National Polytechnic Institute of Constantine, 25016, Constantine, Algeria. Applied Mathematics and Modelling Laboratory, Mentouri brothers University, 25017, Constantine, Algeria.
Email address: houda.bourezaz@enp-constantine.dz

[3] Sciences Laboratory, Department of Mathematics, Mentouri Brothers University of Constantine 1, Constantine, Algeria.
Email address: ilhem.laroussi@umc.edu.dz