



HAL
open science

Régression sur données censurées avec indicateurs manquants

Feriel Bouhadjera

► **To cite this version:**

Feriel Bouhadjera. Régression sur données censurées avec indicateurs manquants. 55es journées de statistique de la SFDS, May 2024, Bordeaux, France. <hal-05010361>

HAL Id: hal-05010361

<https://hal.science/hal-05010361v1>

Submitted on 28 Mar 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

RÉGRESSION SUR DONNÉES CENSURÉES AVEC INDICATEURS MANQUANTS

Feriel Bouhadjera

*Laboratoire CEDRIC, équipe MSDMA, CNAM, France,
feriel.bouhadjera@lecnam.net*

Résumé. Les données censurées représentent une réelle difficulté dans la modélisation d'un problème de régression. Parmi les méthodes classiques permettant d'estimer la fonction de régression, nous pouvons notamment citer les méthodes à noyaux, dites *non-paramétriques*. Toutefois, dans le cas où l'indicateur de censure est manquant, il devient encore plus complexe d'estimer la fonction de régression. Ce travail porte sur un estimateur à noyaux de la fonction de régression basé sur la méthode locale linéaire lorsque la fonction de perte est l'erreur relative quadratique moyenne dans un contexte de données censurées avec indicateurs de censure manquants. Il s'agit d'une extension de Bouhadjera and Ould Saïd (2022) au contexte de données manquantes aléatoirement. L'approche proposée présente certains avantages notamment la stabilité en présence de données censurées et aberrantes avec indicateurs de censure manquants. Elle présente également l'avantage d'atténuer les effets de bords et de réduire le biais. Ainsi, sous des hypothèses appropriées, nous présentons un résultat de convergence asymptotique.

Outre cette étude théorique, ce travail compare numériquement l'approche proposée avec des compétiteurs de la littérature, adaptés au contexte des données manquantes au hasard, dites MAR (Rubin, 1976). Parmi ces compétiteurs, nous retrouvons: la régression classique (Guessoum and Ould Saïd, 2008), la régression locale linéaire (Bouhadjera et al., 2022), la régression relative (Khardani and Slaoui, 2019) et la régression relative locale linéaire (Bouhadjera and Ould Saïd, 2022). Les simulations sont réalisées sur un jeu de données générées sous différents scénarios dépendants du modèle considéré (linéaire ou non-linéaire), de la taille d'échantillon, du taux de valeurs aberrantes, du taux de censure et du taux de valeurs manquantes. L'étude est complétée par une simulation basée sur un jeu de données réels illustrant les bonnes propriétés de l'approche.

Mots-clés. Données censurées, données aberrantes, indicateurs manquants, lissage à noyaux, mécanisme MAR, régression relative, régression locale linéaire.

Abstract. Censored data represent a real difficulty in modeling a regression problem. Classical methods for estimating the regression function include kernel methods, known as *nonparametric*. However, when the censoring indicator is missing, estimating the regression function becomes even more complex. This work deals with a kernel estimator of the regression function based on the local linear method when the loss function is the mean squared relative error in a context of censored data with missing censoring indicators. It is an extension of Bouhadjera and Ould Saïd (2022) in a context of randomly missing data. The proposed approach has advantages, notably stability in the presence of outliers and censored values with missing indicators. It also has the advantage of reducing edge effects and bias. In addition, under appropriate assumptions, we present an asymptotic convergence result with rate.

In addition to this theoretical study, this work numerically compares the proposed approach with competitors adapted to the context of random missing data, known as MAR (Rubin, 1976). These include: classical regression (Guessoum and Ould Saïd, 2008), local linear regression (Bouhadjera et al., 2022), relative regression (Khardani and Slaoui, 2019) and local linear relative regression (Bouhadjera and Ould Saïd, 2022). Simulations are carried out on a dataset generated under different scenarios depending on the model considered (linear or non-linear), on sample size, on outlier rate, on censoring rate and on missing value rate. The study is completed by a simulation based on a real dataset that exhibits the good properties of the approach.

Keywords. Censored data, outliers, missing indicators, kernel smoothing, MAR mechanism, relative regression, local linear regression.

Les modèles de régression non-paramétriques fournissent des outils statistiques performants pour l'analyse des données. Ces modèles sont désormais populaires et largement utilisés dans les applications. En particuliers, une variété d'approches s'intéressent à cette problématique de régression. Sans être exhaustif, nous pouvons en citer : les spline, les k -plus proches voisins, les polynômes locaux et les noyaux. Ce travail s'intéresse quand à lui à l'étude des estimateurs de type noyaux. Cette méthode est simple, flexible et populaire grâce à son potentiel d'interprétabilité.

En analyse de survie, lorsqu'une variable est sujette à la censure, les méthodes statistiques classiques ne peuvent plus s'appliquer directement. Ce type de données constitue alors une difficulté pour la mise en oeuvre de modèles de régression. Nous nous intéressons dans ce travail à un type de données censurées à droite. Sous ce modèle, différents estimateurs non-paramétriques de la fonction de régression ont été développés dans la littérature. Par exemple, Guessoum and Ould Saïd (2008) ont proposé un estimateur à noyau de la régression classique (RC) pour lequel ils ont montré des résultats de convergence presque sûre et de normalité asymptotique. Kim and Truong (1998), Cai (2003) et Bouhadjera et al (2022) ont proposé, quant à eux, un estimateur de la fonction de régression par une approche locale linéaire (RLL) qui possède les propriétés principales d'atténuer les effets de bords et de réduire le terme du biais. Cependant, tous ces estimateurs sont basés sur la minimisation du critère des moindres carrés demeurent peu performants en présence de données aberrantes, souvent rencontrées en pratique. Certains auteurs se sont alors intéressés à un nouveau critère qui minimise les erreurs relatives quadratiques moyennes, plus robuste en présence de valeurs aberrantes (Park and Stefanski, 1998). En considérant ce nouveau critère, Khardani and Slaoui (2019) ont proposé un estimateur de la fonction de régression relative (RR) pour des données indépendantes et Bouhadjera et al. (2023) ont étendu l'approche dans un contexte de données dépendantes (α -mélangeantes). En combinant les deux approches locale linéaire et régression relative, Bouhadjera and Ould Saïd (2022) ont montré que l'on combinait également leurs avantages. Ces derniers ont proposé un estimateur de la fonction de régression relative par une approche locale linéaire (RLL) robuste aux valeurs aberrantes et censurées pour lequel ils ont prouvé la convergence uniforme presque sûre.

Tous ces travaux antérieurs nécessitent que l'indicateur de censure soit complètement observé. En pratique, ce dernier peut être manquant pour différentes raisons. Par exemple, dans une étude médicale, certaines informations importantes concernant les patients peuvent venir à manquer, comme le certificat de décès, les résultats d'autopsie ou les notes d'hospitalisation (e.g. dûe à l'immigration de patients). De plus, dans le cas où nous nous intéressons à un type de décès

spécifique (par exemple, dûe à une maladie X) en considérant tout autres types de décès comme un mécanisme de censure, les indicateurs de censure des patients qui n'ont pas d'informations sur la cause de décès sont manquants. Pour plus d'exemples sur ce type de données, nous renvoyons le lecteur à Van Der Laan and McKeague (1998). Dans ce contexte de données censurées avec indicateurs de censure manquants, Wang et al. (2009) ont proposé un estimateur de la fonction de densité à noyau pour lequel ils ont établi des résultats asymptotiques de convergence presque sûre et de normalité. Brunel et al. (2013) ont étudié, quant à eux, un estimateur non-paramétrique du taux de hasard conditionnel. D'autre part, Bousselmi et al. (2021) ont élaboré plusieurs estimateurs pour un modèle de régression Poisson. Qiu et al. (2021) ont proposé un estimateur de la fonction de régression quantile basé sur des techniques de pondération. Wang et al. (2022) ont proposé un estimateur de la fonction de régression par approche locale polynomiale pour lequel ils ont établi un résultat de normalité asymptotique. Ainsi, la littérature sur les données censurées avec indicateurs de censure manquants proposent différentes approches pour estimer la fonction de régression. Dans ce travail, nous nous intéressons à développer une méthode non-paramétrique pour estimer la fonction de régression, robuste aux valeurs aberrantes pour des données censurées à droite avec indicateurs de censure manquants.

Le présent document est organisé comme suit. Dans la partie **formalisme**, nous présentons les notations et hypothèses nécessaires à la construction du modèle. Ensuite, la partie **données synthétiques** présente la modélisation des données observées dans un contexte de données censurées à droite avec indicateurs manquants au hasard. De plus, les fonctions de perte considérées, notamment les moindres carrées et les moindres carrés relatifs, ainsi que les estimateurs résultants sont présentés dans la partie **fonctions de perte et estimateurs associés**. Enfin, dans la partie **propriétés**, nous abordons les résultats asymptotiques théorique et numériques de l'approche proposée en comparaison avec des compétiteurs de la littérature, adaptés au contexte des données manquantes au hasard.

Formalisme. Soit (X, Y) un couple de variables aléatoires (v.a.) où Y est une v.a. réelle positive expliquée et X est une v.a. réelle explicative (de densité f_X). On suppose que la densité du couple (X, Y) existe et est notée f_{XY} . On cherche à modéliser le lien entre X et Y représenté par $Y = m(X) + \varepsilon$, où $m(X) = \mathbb{E}[Y|X]$ est la fonction de régression (inconnue) qui apparaît comme une quantité contenant toute l'information sur la structure de dépendance et ε est une v.a. représentant le terme d'erreur. On suppose que ε est centrée, de variance finie et indépendante de X . Classiquement, la fonction de régression $m(\cdot)$ est obtenue en minimisant $\mathbb{E}[(Y - m(X))^2|X]$. Toutefois, cette fonction de perte est inefficace en présence de données aberrantes.

L'estimation de l'erreur relative a été récemment utilisée dans l'analyse de la régression comme une alternative aux restrictions imposées par l'approche classique des moindres carrés. Cette dernière consiste à considérer l'estimation de la fonction de régression $m(\cdot)$ minimisant l'erreur relative quadratique moyenne $\mathbb{E}[Y^{-2}(Y - m(X))^2|X]$. Ce critère a été largement étudié dans un cadre paramétrique (Park and Stefanski, 1998; Chen et al., 2010) et a été récemment considéré dans le cadre non-paramétrique dans différents contextes (Chahad et al., 2017; Attouch et al., 2017; Thiam, 2018; Hu, 2019). Sous l'hypothèse que les deux premiers moments inverses conditionnels de Y sachant X existent et soient finis, (Park and Stefanski, 1998) ont montré que la minimisation de l'erreur relative quadratique moyenne s'exprime sous la forme $m(X) = m_1(X)/m_2(X)$ où

pour $\nu = 1, 2$, $m_\nu(\cdot) = \mathbb{E}[Y^{-\nu}|X = \cdot] = r_\nu(\cdot)/f_X(\cdot)$ avec $r_\nu(\cdot) = \int y^{-\nu} f_{XY}(\cdot, y) dy$.

Dans ce travail, nous nous intéressons à ce dernier modèle de régression relative non-paramétrique en analyse de survie, et en particulier nous nous focalisons sur des situations où la réponse Y est sujette à une censure aléatoire à droite. Plus formellement, soit C une v.a. non-négative et continue, de fonction de survie S . Dans un modèle de censure à droite, le triplet observé est (X, \dot{Y}, δ) , où $\dot{Y} = \min(Y, C)$ est la variable observée et $\delta = \mathbb{1}_{\{Y \leq C\}}$ est une v.a. indicatrice, qui vaut 1 si nous observons le temps d'intérêt Y (c'est-à-dire $\dot{Y} = Y$), et 0 si nous observons le temps de censure C . Nous supposons que le couple (X, Y) et C sont indépendants (\mathbf{H}_0). Cette hypothèse d'indépendance est classiquement utilisée pour les modèles de censure à droite et permet l'identifiabilité du modèle.

À présent dans le cas où, l'indicatrice de censure δ est partiellement manquante, nous introduisons une nouvelle v.a. ξ , qui vaut 1 si la v.a. δ est observée, et 0 si elle est manquante. Nous supposons que les variables ξ et δ sont indépendantes conditionnellement à \dot{Y} (\mathbf{H}_1). Cette hypothèse est classique dans les analyses statistiques qui incluent des données manquantes (Little and Rubin, 2002) et signifie que δ est manquante aléatoirement (Missing At Random "MAR") à la différence du mécanisme de données manquantes complètement au hasard (Missing Completely At Random "MCAR") qui nécessite une hypothèse d'indépendance complémentaire. Sous \mathbf{H}_1 , nous désignons par $\omega(\cdot) \equiv \mathbb{P}[\delta|\dot{Y} = \cdot, \xi = 1] = \mathbb{P}[\delta|\dot{Y} = \cdot]$ la probabilité conditionnelle de l'indicatrice de censure sachant la variable observée \dot{Y} .

Données synthétiques. Dans un modèle de censure à droite avec indicateur de censure sous mécanisme MAR, les données nécessitent une modélisation spécifique. Nous présentons dans cette partie, ce que l'on appelle les données synthétiques, qui remplacent les données d'intérêt qui ne sont pas complètement observées en raison des phénomènes de censure et valeurs manquantes. Les données observées consistent en n réalisations indépendantes et identiquement distribuées (i.i.d.) $O_i \equiv (X_i, \dot{Y}_i, \xi_i = 1, \delta_i)$ ou $(X_i, \dot{Y}_i, \xi_i = 0)$, pour $i = 1, \dots, n$, de même loi que le quadruplet $(X, \dot{Y}, \xi, \delta)$. Pour cet n -échantillon, les données synthétiques sont données par :

$$\tilde{Y}_i^{-\nu} = \left(\xi_i \frac{\delta_i}{S(\dot{Y}_i)} + (1 - \xi_i) \frac{\omega(\dot{Y}_i)}{S(\dot{Y}_i)} \right) \dot{Y}_i^{-\nu}, \quad \text{pour } \nu = 1, 2, \quad (1)$$

où $\omega(\cdot)$ et $S(\cdot)$ sont respectivement la probabilité de présence de l'indicatrice et la fonction de survie de la variable de censure. Dans le cas où $\xi_i = 1$, l'équation (1) se résume aux données synthétiques considérées dans Bouhadjera and Ould Saïd (2022). De plus, en nous basant sur les propriétés de l'espérance conditionnelle et sous les hypothèses d'indépendance \mathbf{H}_0 et \mathbf{H}_1 , nous avons :

$$\mathbb{E} \left[\tilde{Y}_i^{-\nu} | O_i \right] = \mathbb{E} \left[Y_i^{-\nu} | X_i \right] = m_\nu(X_i), \quad (2)$$

pour $\nu = 1, 2$ et $i = 1, \dots, n$. Toutefois, en pratique, les fonctions $\omega(\cdot)$ et $S(\cdot)$ dans (1) sont inconnues. Pour obtenir des données synthétiques parfaitement calculables, nous cherchons à remplacer ses deux fonctions dans (1) par leurs estimateurs respectifs. Concernant, la fonction $\omega(\cdot)$, de façon similaire à Wang et al. (2009) et Wang et al. (2022), nous proposons de l'approcher par un

estimateur de type Nadaraya-Watson (Nadaraya (1964); Watson (1964)) donné par :

$$\hat{\omega}(y) = \frac{\sum_{j=1}^n \xi_j \delta_j W_b(\dot{Y}_j - y)}{\sum_{j=1}^n \xi_j W_b(\dot{Y}_j - y)}, \quad (3)$$

où $W_b(\cdot) := W(\cdot/b)$ est une densité de probabilité appelée *noyau* et $0 < b := b_n \xrightarrow[n \rightarrow \infty]{} 0$ est une suite de nombre réels positifs appelée *fenêtre*. À présent, concernant la fonction de survie S , certains auteurs comme Wang et al. (2022) proposent d'adapter le célèbre estimateur de Kaplan and Meier (1958) en remplaçant directement l'indicatrice de censure manquante par (3), ce qui donne :

$$\hat{S}(y) = \prod_{i: \dot{Y}_i \leq y} \left(\frac{n - R_i}{n - R_i + 1} \right)^{1 - \hat{\omega}(\dot{Y}_i)}, \quad (4)$$

où $R_i = \sum_{j=1}^n \mathbb{1}_{\{Y_j \leq \dot{Y}_i\}}$ pour $i = 1, \dots, n$. En remplaçant, (4) et (3) dans (1), nous obtenons des données synthétiques calculables donnée par :

$$\hat{Y}_i^{-\nu} = \left(\xi_i \frac{\delta_i}{\hat{S}(\dot{Y}_i)} + (1 - \xi_i) \frac{\hat{\omega}(\dot{Y}_i)}{\hat{S}(\dot{Y}_i)} \right) \dot{Y}_i^{-\nu}, \quad (5)$$

pour $\nu = 1, 2$ et $i = 1, \dots, n$.

Fonctions de perte et estimateurs associés. Dans cette présentation, nous nous intéressons à deux critères d'optimisation : les moindres carrés et les moindres carrés relatifs afin d'estimer la fonction de régression $m(\cdot)$. Nous considérons pour ces deux fonctions de perte l'approche locale linéaire dont l'idée principale est d'approximer $m(X)$ par un développement de Taylor d'ordre 1. Ainsi, nous supposons que la première dérivée de $m(\cdot)$ existe et est continue telle que, $m(X) \approx m(x) + m'(x)(X - x) \equiv \beta_0 + \beta_1(X - x)$ où $x \in \mathcal{C}$ avec \mathcal{C} un ensemble compact de $\mathcal{C}_0 = \{x \in \mathbb{R}, f_X(x) > 0\}$.

Critère des moindres carrés. Nous nous intéressons à β_0 qui minimise l'erreur quadratique moyenne pondérée et adaptée en considérant les données synthétiques de l'équation (5) avec $\hat{Y}_i = \hat{Y}_i^{-\nu}$, pour $\nu = -1$, donnée par :

$$\arg \min_{\beta_0 \in \mathbb{R}} \left\{ \sum_{i=1}^n (\hat{Y}_i - \beta_0 - \beta_1(X_i - x))^2 K_h(X_i - x) \right\}, \quad (6)$$

où $K_h(\cdot) := K(\cdot/h)$ et $0 < h := h_n \xrightarrow[n \rightarrow \infty]{} 0$ sont respectivement le noyau et la fenêtre définis précédemment. Par un calcul algébrique, la solution du problème de minimisation (6), nous donne un estimateur de la fonction de RLL pour une variable d'intérêt censurée à droite avec un indicateur de censure manquant aléatoirement dont l'expression est la suivante :

$$\hat{\beta}_0^{RLL} := \frac{\hat{r}_{-1,0}(x) \hat{f}_2(x) - \hat{r}_{-1,1}(x) \hat{f}_1(x)}{\hat{f}_2(x) \hat{f}_0(x) - (\hat{f}_1(x))^2},$$

avec

$$\hat{f}_\eta(x) = \frac{1}{nh} \sum_{i=1}^n (X_i - x)^\eta K_h(X_i - x), \quad (7)$$

et

$$\hat{r}_{\nu,\eta}(x) = \frac{1}{nh} \sum_{i=1}^n \hat{Y}_i^{-\nu} (X_i - x)^\eta K_h(X_i - x), \quad (8)$$

pour $\nu = 1, 2$ et $\eta = 0, 1, 2$. Dans le cas où, les δ_i sont complètement observés pour tout i (c'est-à-dire pour $\xi_i = 1$), la solution du critère (6) a été étudiée par Bouhadjera et al. (2022). Considérons maintenant $\beta_1 = 0$, la minimisation de (6) nous donne un estimateur de la fonction de RC dont l'expression est donnée par :

$$\hat{\beta}_0^{RC} = \frac{\hat{r}_{-1,0}(x)}{\hat{f}_0(x)}.$$

avec $\hat{r}_{\nu,\eta}(x)$ et $\hat{f}_\eta(x)$ pour $\nu = -1$ et $\eta = 0$ sont définis dans (7) et (8) respectivement. Le cas sans indicateurs manquants a été considéré dans (Guessoum and Ould Saïd, 2008).

Critères de moindres carrés relatifs. Nous nous intéressons ici à β_0 qui minimise l'erreur relative quadratique moyenne pondérée et adaptée en considérant les données synthétiques de l'équation (5), donnée par :

$$\arg \min_{\beta_0 \in \mathbb{R}} \left\{ \sum_{i=1}^n \hat{Y}_i^{-2} \left(\hat{Y}_i - \beta_0 - \beta_1 (X_i - x) \right)^2 K_h(X_i - x) \right\}. \quad (9)$$

Comme pour le critère (6), la solution du problème de minimisation (9) nous donne un estimateur de la fonction de RRLM adapté au contexte censuré avec indicateur manquant dont l'expression est la suivante :

$$\hat{\beta}_0^{RRLM} = \frac{\hat{m}_1(x)}{\hat{m}_2(x)}, \quad \text{avec} \quad \hat{m}_\nu(x) = \hat{r}_{2,2}(x) \hat{r}_{\nu,0}(x) - \hat{r}_{2,1}(x) \hat{r}_{\nu,1}(x),$$

où $\hat{r}_{\nu,\eta}(x)$ pour $\nu = 1, 2$ et $\eta = 0, 1, 2$ est défini dans (8). Si δ est complètement observé, l'estimateur ci-dessous a été étudié dans Bouhadjera and Ould Saïd (2022). Le critère (9) peut être réduit au cas où $\beta_1 = 0$. Dans ce cas, la solution de ce problème de minimisation nous donne un estimateur de la fonction RR :

$$\hat{\beta}_0^{RR} = \frac{\hat{r}_{1,0}(x)}{\hat{r}_{2,0}(x)},$$

où $\hat{r}_{\nu,\eta}(x)$ pour $\nu = 1, 2$ et $\eta = 0$ est définie dans (8). Cet estimateur a été proposé et étudié par Khardani and Slaoui (2019) dans le cadre de données censurées classiques, c'est-à-dire lorsque $\xi_i = 1$ pour $i = 1, \dots, n$.

Propriétés. Cette présentation propose d'étudier le comportement asymptotique de $\hat{\beta}_0^{RRLM}$ dans le contexte de données censurées à droite avec indicateurs de censure manquants sous un mécanisme MAR. Un résultat de convergence uniforme presque sûre est proposé sous des hypothèses classiques en estimation non-paramétrique telles que la régularité de fonctions, des hypothèses sur les

fenêtres h et b et finalement sur les noyaux W_b et K_h . En particulier, celui-ci caractérise la vitesse de convergence de l'estimateur en fonction des fenêtres h et b . De plus, une étude numérique est effectuée pour comparer les différents estimateurs de régression ($\hat{\beta}_0^{RC}$, $\hat{\beta}_0^{RLL}$, $\hat{\beta}_0^{RR}$ et $\hat{\beta}_0^{RRLL}$), adaptés au contexte de données censurées avec indicateurs de censure manquants, variant par la nature du mécanisme considéré (MAR ou MCAR). L'étude est réalisée sur des données simulées selon trois modèles (linéaire, quadratique et exponentielle) dont on fait varier la taille d'échantillon, le taux de censure et le taux de valeurs aberrantes. L'étude illustre que la qualité d'estimation est meilleure par approche RLL en comparaison aux approches RC, RLL, RR. Ainsi, ces premiers résultats numériques tendent à montrer que la méthode proposée pour ce contexte de données censurées avec indicateurs manquants est performante.

Bibliographie

- B. Bousselmi, J. F. Dupy and A. Karoui (2021), Censored count data regression with missing censoring information. *Electronic Journal of Statistics*, 15(2): 4343–4383.
- D. Dabrowska (1987), Nonparametric regression with censored survival data. *Scandinavian Journal of Statistics*, 14:181–197.
- E. A. Nadaraya (1964), On estimating regression. *Theory of Probability & Its Applications*, 9:141–142.
- E. Brunel, F. Comte and A. Guillaou (2013), Nonparametric estimation for survival data with censoring indicators missing at random. *Journal of Statistical Planning and Inference*. 143(10):1653–1671.
- E. L. Kaplan and P. Meier (1958), Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53:458–481.
- S. Khardani and Y. Slaoui (2019), Nonparametric relative regression under random censorship model. *Statistics & Probability Letters*, 151:116–122.
- F. Bouhadjera and E. Ould Saïd (2022), Strong consistency of the local linear relative regression estimator for censored data. *Opuscula Mathematica*, 42(6):805–832.
- F. Bouhadjera, E. Ould Saïd and M. R. Remita (2022), Strong consistency of the nonparametric local linear regression estimation under censorship model, *Communications in Statistics - Theory and Methods*, 51(20):7056–7072.
- F. Bouhadjera, E. Ould Saïd & M. R. Remita (2023), Relative error prediction : Strong uniform consistency for censoring time series model, *Communications in Statistics - Theory and Methods*, 52(11):3709–3729.
- G. S. Watson (1964), Smooth regression analysis. *Sankhyà*, 26:359–372.
- H. Park and L. A. Stefanski (1998), Relative-error prediction. *Statistics and probability letters*, 40(3), 227–236.
- H. T. Kim and Y. K. Truong (1998), Nonparametric regression estimates with censored data : local linear smoothers and their applications. *Biometrics*, 54:1434–1444.

- J. Wang, Y. Zhou and J. Tang (2022), Weighted local polynomial estimations of a non-parametric function with censoring indicators missing at random and their applications. *Frontiers of Mathematics in China*, 17:117–139.
- M. J. Van Der Laan and I. W. McKeague (1998), Efficient estimation from right-censored data when failure indicators are missing at random. *Annals of Statistics*, 26(1):164–182.
- K. Chen, S. Guo, Y. Lin and Z. Ying (2010). Least absolute relative error estimation. *Journal of the American Statistical Association*, 105(491):1104–1119.
- Q. Wang, W. Liu and C. Liu (2009), Probability density estimation for survival data with censoring indicators missing at random. *Journal of Multivariate Analysis*, 100:835–850.
- R. Little and D. Rubin (2002). Statistical analysis with missing data, *New-York*.
- D. Rubin (1976), Inference and missing data. *Biometrika*, 63:581–592.
- Z. Cai (2003), Weighted local linear approach to censored nonparametric regression. *Recent Advances and Trends in Nonparametric Statist*, 217–231.
- Z. Guessoum and E. Ould Saïd (2008), On nonparametric estimation of the regression function under random censorship model. *Statistics and Decisions*, 26:1001–1020.
- Z. Qiu, H. Ma, J. Chen and G. E. Dinse (2021), Quantile regression models for survival data with missing censoring indicators. *Statistical Methods in Medical Research*, 30(5):1320–1331.