

INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES
Série des Documents de Travail du CREST
(Centre de Recherche en Economie et Statistique)

n° 2006-10

**Nonlinear Censored Regression
Using Synthetic Data**

M. DELECROIX¹

O. LOPEZ²

V. PATILEA³

Les documents de travail ne reflètent pas la position de l'INSEE et n'engagent que leurs auteurs.

Working papers do not reflect the position of INSEE but only the views of the authors.

¹ CREST-ENSAI and LSTA Paris 6. Email : delecroi@ensai.fr

² IRMAR and CREST-ENSAI. Email : lopez@ensai.fr *Corresponding author*, Address : CREST-ENSAI, Campus de Ker Lann, Rue Blaise Pascal, BP 37203, 35172 BRUZ Cedex, France.

³ CREST-ENSAI. Email : patilea@ensai.fr

Nonlinear censored regression using synthetic data

M. Delecroix* O. Lopez[†] V. Patilea[‡]

June 15, 2006

*CREST-ENSAI and LSTA Paris 6. Email: delecroi@ensai.fr

[†]IRMAR and CREST-ENSAI. Email: lopez@ensai.fr. Corresponding author. Address for correspondence: CREST-ENSAI, Campus de Ker Lann, Rue Blaise Pascal BP 37203, 35172 Bruz cedex, FRANCE

[‡]CREST-ENSAI. Email: patilea@ensai.fr

Abstract

The problem of estimating a nonlinear regression model when the dependent variable is randomly censored is considered. The parameter of the model is estimated by least squares using synthetic data, that is a suitable transformation of the response variables that preserves the conditional expectation. Two such transformations are considered. Consistency and asymptotic normality of the least squares estimators are derived. The proofs are based on a novel approach that uses i.i.d. representation of synthetic data through Kaplan-Meier integrals. The asymptotic results are completed by a small comparative simulation study.

Key words: asymptotic normality, consistency, Kaplan-Meier integral, nonlinear regression, right censoring, synthetic data.

MSC 2000: 62N02, 62J02, 62F12

Résumé

Nous considérons le problème de l'estimation d'un modèle de régression non linéaire, dans le cas où la variable explicative est censurée aléatoirement. Le paramètre de ce modèle est estimé par moindres carrés basés sur des "synthetic data", c'est à dire une certaine transformation des observations qui préserve l'espérance conditionnelle. Deux transformations de ce type sont considérées. Les preuves sont basées sur une nouvelle approche utilisant des représentations i.i.d. des "synthetic data", grâce aux représentations i.i.d. des intégrales Kaplan-Meier. Les résultats sur les propriétés asymptotiques sont enrichis par une étude comparative des différents estimateurs par simulation.

Mots clefs : normalité asymptotique, convergence, intégrale Kaplan-Meier, régression non linéaire, censure à droite, synthetic data.

1 Introduction

The nonlinear regression model is a common way to infer about the dependence between two random variables Y and X . Assuming that

$$E[Y | X] = f(\theta_0, X), \quad \theta_0 \in \Theta \subset \mathbb{R}^k,$$

where f is a known function, and given an i.i.d. sample of Y and X , a classical way to deal with the estimation of the parameter θ_0 consists in using a (nonlinear) least squares procedure. In presence of random censorship, the situation appears to be more complex. Assume that, instead of i.i.d. replications $(X, Y) \in \mathbb{R}^{d+1}$, one only observes n independent copies of (X, δ, T) where

$$T = Y \wedge C, \quad \delta = \mathbf{1}_{\{Y \leq C\}},$$

where C is a censoring random variable and $\mathbf{1}_A$ denotes the indicator function of the set A . Then the least squares procedure must be properly adapted to account for censorship. In the particular case of linear regression (that is when f is linear), several methods have already been studied, mostly using martingales techniques. But inquiring into their theoretical properties, the existing results strongly rely on the particular shape of the linear regression function. Therefore the extension of these techniques to the more general nonlinear regression model does not seem straight.

In the linear case, the existing estimators can be split into two main categories, namely weighted least-squares and synthetic data estimators. None of the two approaches dominates the other, that is, depending on the distribution of C or Y , either one or the other approach have better finite sample and asymptotic properties. The weighted least-squares approach was first considered by Zhou (1992a). Using suitable weights in the least squares criterion, Zhou proposes an estimator that accounts for censorship. The weights he used were then connected by Stute (1999) to the jumps of a distribution function estimating $\mathbb{P}(X \leq x, Y \leq y)$. It is the only method that has been fully extended to the nonlinear model case. An alternative method that we call the synthetic data approach, is based on transformations of $T = Y \wedge C$ which have the same conditional expectation as Y . Such estimation methods were developed in Koul, Susarla and Van Ryzin (1981) and Leurgans (1987). Buckley and James (1979) estimator could be also included in the synthetic data type estimators. However, the transformation proposed by Buckley and James depends also on θ_0 and thus the implementation of this method requires an iterative algorithm. For this reason, this approach will not be considered below.

The aim of this paper is to obtain asymptotic properties for a whole class of synthetic data estimators in the case of nonlinear regression. Our class, which is closely related to the transformations studied by Fan and Gijbels (1994), extends the estimators of Koul, Susarla and Van Ryzin (1981) and Leurgans (1987) to nonlinear regressions. For deriving the asymptotic results, instead of using martingales techniques as it was done in the linear case, we will propose a new way to understand estimators of Koul, Susarla and Van Ryzin (1981) and Leurgans (1987). Then they will be easily connected to Kaplan-Meier integrals, such as those studied by Stute (1993, 1996).

The reminder of the paper is organized as follows. In section 2, the model and the nonlinear least squares estimation method are introduced. Section 3 contains a new representation of empirical means involving synthetic data as Kaplan-Meier integrals plus

a negligible reminder. This representation is the key element for the new approach used to obtain asymptotic results. Consistency and asymptotic normality for the nonlinear least squares estimator are derived in section 4 and 5. In section 6, our theoretical study is completed by an empirical comparison with Stute's estimator (1999). Some concluding remarks are gathered in section 7. The technical proofs are relegated to the appendix.

2 Model and estimator

In regression analysis with censored responses, one must first make some assumptions on Y , C and X and their dependence relationships. Following Stute (1996, 1999), consider the identifiability conditions

- I1. Y and C are independent,
- I2. $\mathbb{P}(Y \leq C \mid Y, X) = \mathbb{P}(Y \leq C \mid Y)$.

Condition I1 is a common identification condition when using the Kaplan-Meier estimator. Conditions like I1 and I2 allow studying Koul, Susarla and Van Ryzin (1981) estimator. This estimator, which will be denoted KSV in the following, relies on the following transformation

$$Y_{KSV}^* = \frac{\delta T}{1 - G(T-)} \quad (2.1)$$

where $G(t) = \mathbb{P}(C \leq t)$. It follows from condition I2 that $E[Y_{KSV}^* \mid X] = E[Y \mid X]$.

Leurgans proposed another transformation of the censored response

$$Y_L^* = \int \left(\frac{\mathbf{1}_{\{T>t\}}}{1 - G(t)} - \mathbf{1}_{\{t<0\}} \right) dt. \quad (2.2)$$

In this case, to obtain $E[Y_L^* \mid X] = E[Y \mid X]$, the following stronger condition is needed:

- I2'. C and (X, Y) are independent.

Weaker assumptions have also been proposed in some particular cases. In the case of polynomial regression with one covariate X , Heuchenne and Van Keilegom (2006) only require that Y and C are independent conditionally to $X \in \mathbb{R}$. Heuchenne and Van Keilegom (2005) extends the analysis to the case of a nonlinear, not necessarily polynomial, univariate regression. However, this latter estimator requires some kernel smoothing and thus happens to have bad performances when the number of covariates increases. Moreover, discrete covariables cannot be considered. These drawbacks do not affect our approach.

If Y^* were available, the parameter θ_0 could be estimated through a standard least squares procedure. Unfortunately, the transformations in (2.1) and (2.2) cannot be directly computed from the data. They would require exact knowledge of the distribution function G . Although G is unknown, it can be estimated from the data, typically by its Kaplan-Meier estimator

$$\hat{G}(t) = 1 - \prod_{i: T_i \leq t} \left(\frac{n - i}{n - i + 1} \right)^{1 - \delta_i}.$$

Define

$$\begin{aligned}
\hat{Y}_{i,KSV}^* &= \frac{\delta_i T_i}{1 - \hat{G}(T_i-)}, & 1 \leq i \leq n, \\
\hat{Y}_{i,L}^* &= \int \left(\frac{\mathbf{1}_{\{T_i > t\}}}{1 - \hat{G}(t)} - \mathbf{1}_{\{t < 0\}} \right) dt, & 1 \leq i \leq n, \\
M_n^{KSV}(\theta) &= \frac{1}{n} \sum_{i=1}^n \left(\hat{Y}_{i,KSV}^* - f(\theta, X_i) \right)^2, \\
M_n^L(\theta) &= \frac{1}{n} \sum_{i=1}^n \left(\hat{Y}_{i,L}^* - f(\theta, X_i) \right)^2, \\
\hat{\theta}_{KSV} &= \arg \min_{\theta \in \Theta} M_n^{KSV}(\theta), \\
\hat{\theta}_L &= \arg \min_{\theta \in \Theta} M_n^L(\theta).
\end{aligned}$$

We call $\hat{\theta}_{KSV}$ and $\hat{\theta}_L$ the KSV and Leurgans estimator, respectively. Other estimators can be computed by using convex combinations of (2.1) and (2.2). For $\alpha \in \mathbb{R}$, follow Fan and Gijbels (1994) and define

$$\begin{aligned}
\hat{Y}_{i,\alpha}^* &= (1 + \alpha) \hat{Y}_{i,L}^* - \alpha \hat{Y}_{i,KSV}^*, \\
M_n^\alpha(\theta) &= \frac{1}{n} \sum_{i=1}^n \left(\hat{Y}_{i,\alpha}^* - f(\theta, X_i) \right)^2, \\
\hat{\theta}_\alpha &= \arg \min_{\theta \in \Theta} M_n^\alpha(\theta).
\end{aligned}$$

See section 3 of Fan and Gijbels (1994) for a motivation of the definition of $\hat{Y}_{i,\alpha}^*$.

In the following, we derive asymptotic properties (consistency and asymptotic normality) for the class of estimators $\hat{\theta}_\alpha$ which contains $\hat{\theta}_{KSV}$ and $\hat{\theta}_L$ as particular cases. Before getting into the details, let us give a flavor of our approach. In the case of uncensored dependent observations, the asymptotics of the nonlinear least-squares estimators can be derived from the behavior of the functionals like

$$\frac{1}{n} \sum_{i=1}^n Y_i \varphi(\theta, X_i) \quad \text{and} \quad \frac{1}{n} \sum_{i=1}^n \psi(\theta, X_i)$$

where φ and ψ are defined by f and its derivatives. The first sum can be seen as an integral with respect to the empirical distribution function of the observations (X, Y) . In the presence of censoring, a natural way to proceed with this integral is to replace the empirical process by a suitable empirical measure that accounts for censored Y . On contrary, as in the framework we consider the covariates X are not censored, the second sum in the last display remains unchanged.

3 Kaplan-Meier integrals for synthetic data

Consider $F_{(X,Y)}(x, y) = \mathbb{P}(X \leq x, Y \leq y)$ the $(d + 1)$ -variate distribution function of (X, Y) . Stute (1993) proposed to estimate $F_{(X,Y)}$ by putting on (X_i, Y_i) the same weights

as the Kaplan-Meier estimator puts on Y_i in the univariate case. Stute's estimator can be written

$$\hat{F}_{(X,Y)}(x,y) = \sum_{i=1}^n W_{i:n} \mathbf{1}_{\{X_i \leq x\}} \mathbf{1}_{\{T_i \leq y\}}$$

where $W_{i:n}$ is the jump of the Kaplan-Meier estimator of $F(t) = \mathbb{P}(Y \leq t)$ at observation i . Stute (1993, 1996) derived strong law of large numbers and distributional convergence for functionals like $\int \phi(x,y) d\hat{F}_{(X,Y)}(x,y)$. The asymptotics of Stute (1999)'s estimator for the nonlinear regression model under random censorship were derived from the asymptotics of such a functional with $\phi(x,y) = (y - f(\theta, x))^2$. The simplicity and intelligibility of this method, along with the possibility to obtain asymptotic variance formulas of the same form for each estimator, invites to inquiring into a similar connection to those multivariate Kaplan-Meier integrals in the case of the synthetic data approach. For this purpose, first note that

$$W_{i:n} = \frac{1}{n} \frac{\delta_i}{1 - \hat{G}(T_i^-)},$$

see equation (2) in Zhou (1992a) or Satten and Datta (2001) for a full proof of this result. A direct consequence of this remark is the following representation lemma which will permit studying KSV estimator.

Lemma 1 *Let $g(x)$ be any measurable function. Then*

$$\frac{1}{n} \sum_{i=1}^n g(X_i) \hat{Y}_{i,KSV}^* = \int g(x) y d\hat{F}_{(X,Y)}(x,y).$$

A similar representation of the empirical means involving Leurgans' synthetic observations \hat{Y}_L^* can be derived. Let $H(t) = \mathbb{P}(T \leq t)$ and $F(t) = \mathbb{P}(Y \leq t)$. Let \hat{H} be the empirical distribution of T_i and \hat{F}_Y the Kaplan-Meier estimator one would obtain from Y_i , $1 \leq i \leq n$. The starting point for the representations with synthetic observations \hat{Y}_L^* is the identity

$$\hat{Y}_L^* = \int_{-\infty}^{\infty} \left(\frac{(1 - \hat{F}(t)) \mathbf{1}_{\{T > t\}}}{1 - \hat{H}(t)} - \mathbf{1}_{\{t < 0\}} \right) dt, \quad (3.1)$$

a direct consequence of the well known property $1 - \hat{H}(t) = (1 - \hat{F}(t))(1 - \hat{G}(t))$.

To avoid complicated expansions, in the following we will assume $T > 0$. This is just for an easier formulation of the results and a better comparison with the proof of asymptotic normality in the linear model case (Zhou, 1992b). This restriction does not affect asymptotic properties, since all difficulties in proving asymptotic results stands in the right tail of T . Note that in equation (3.3) below the measure $\hat{F}_{(X,Y)}$ can be replaced by \hat{F}_Y , the Kaplan-Meier estimator of the distribution function of Y , and thus the first integral in that equation is a one-dimensional integral. For the sake of unifying notation, we prefer to write (3.3) in this way.

Lemma 2 *Let g be a bounded measurable function, $F(t, X) = \mathbb{P}(Y \leq t | X)$ and let*

$$K(g, t) = E[g(X)(1 - F(t, X))] = \frac{E[g(X) \mathbf{1}_{\{T > t\}}]}{1 - G(t)}.$$

Assume $T = Y \wedge C > 0$ and suppose that for some $\varepsilon \in (0, 1/2)$

$$\int \frac{|y|}{(1 - H(y))^{1/2+\varepsilon}} dF_{(X,Y)}(x, y) < \infty. \quad (3.2)$$

Then,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n g(X_i) (\hat{Y}_{i,L}^* - Y_{i,L}^*) &= \int \left(\int_0^y \frac{K(g, t)}{1 - F(t)} dt \right) \left\{ d\hat{F}_{(X,Y)}(x, y) - dF_{(X,Y)}(x, y) \right\} \\ &+ \int \frac{K(g, y)}{1 - H(y)} \left(\frac{1}{n} \sum_{i=1}^n \{\mathbf{1}_{\{T_i \leq y\}} - H(y)\} \right) dy \\ &+ R_n \end{aligned} \quad (3.3)$$

with $R_n = o_P(n^{-1/2})$.

The proof of this lemma is postponed to the appendix. Note that identity (3.3) is related to equation (3.6) in Zhou (1992b). In view of (3.1), the first term of the decomposition (3.3) can be interpreted as the contribution of the estimation of F by \hat{F} , while the second one comes from the approximation of H by \hat{H} . Note that condition (3.2) will be fulfilled for a large class of models including F and G both exponential or both gaussian (with $1 - F$ decreasing faster than $1 - G$), or F being gaussian and G exponential.

4 Consistency

Consistency can be quickly derived using the above representations if f is differentiable with respect to θ . However, we show that consistency can be achieved even without such differentiability conditions. In the following, we shall say that a function $\psi(\theta, X)$ satisfies the property (P) if :

$$(P) \quad \forall \theta_1, \theta_2 \in \Theta, \quad |\psi(\theta_1, X) - \psi(\theta_2, X)| \leq \Phi(X) \|\theta_1 - \theta_2\|^a, \\ \text{with } E[\Phi^2(X)] < \infty \text{ and for some } a > 0.$$

Assumption 1 *i) $E[Y | X] = f(\theta_0, X)$ for some $\theta_0 \in \Theta$, where Θ is a compact subset of \mathbb{R}^k . Moreover, θ_0 is unique with this property.*

ii) The function $f : \Theta \times \mathbb{R}^d \rightarrow \mathbb{R}$ satisfies property (P) and $E[f^2(\theta_0, X)] < \infty$.

Assumption 2 $E[Y_\alpha^{*2}] < \infty$ for $\alpha = -1$ and $\alpha = 0$.

Assumption 3 *The censoring C satisfy conditions I1 and I2' if $\alpha \neq 1$, I1 and I2 if $\alpha = 1$, and $\mathbb{P}(Y = C) = 0$. The distribution functions F and G are continuous and $0 < \tau_H = \tau_F \leq \infty$, where $\tau_L = \inf\{t | L(t) = 1\}$ for any distribution function L .*

Assumption 1 *i)* states that the regression model is correct and guarantees the identification of θ_0 . Assumption 1 *ii)* represents a sufficient condition for $f^2(\theta, X)$ to satisfy a law of large numbers uniformly over Θ . See e.g. Pakes and Pollard (1989) or van

der Vaart and Wellner (1996). Assumption 2 is equivalent to $E[Y_\alpha^{*2}] < \infty$ for any $\alpha \in \mathbb{R}$, and this is a minimal condition for applying nonlinear least squares with the unobserved Y_α^* . This assumption will suffice to prove consistency for $\hat{\theta}_\alpha$ even if Y_α^* is replaced by \hat{Y}_α^* in the definition of the nonlinear least squares estimator. As explained in section 2, the independence between C and (X, Y) imposed by Assumption 3 ensures that $E[Y | X] = E[Y_\alpha^* | X]$, while the condition $\mathbb{P}(Y = C) = 0$ allows to switch the roles of Y and C and thus to estimate G by the classical Kaplan-Meier estimator \hat{G} . When $\tau_F > \tau_H$, there is no way to access to information about the law of Y beyond τ_H , so that, in general, there is no way to consistently estimate θ_0 . The last part of Assumption 3 removes this case. The proof of the following theorem is postponed to the appendix.

Theorem 3 *Under Assumptions 1 to 3, $\hat{\theta}_\alpha \rightarrow \theta_0$ in probability.*

5 Asymptotic normality

In the case of linear regression, Zhou (1992b) gave a proof of asymptotic normality of Leurgans' estimator that strongly relies on the particular shape of the regression function. In fact, in the linear setting the least squares estimator $\hat{\theta}$ is a linear function of the \hat{Y}_i^* . In the nonlinear regression model, in general there is no explicit solution of the equation defining $\hat{\theta}$. In such a case, the i.i.d. representations of the Kaplan-Meier integrals derived by Stute (1996) become a valuable tool. Let us now introduce some notation similar to that of Stute (1996), and recall the asymptotic representation of Kaplan-Meier integrals derived in Stute (1993, 1996). In the following, $\nabla_\theta g$ (resp. $\nabla_\theta^2 g$) denotes the vector (resp. matrix) of first (resp. second) order partial derivatives of a function g with respect to θ .

Assumption 4 *i) The vector valued function $\nabla_\theta f(\theta, \cdot)$ is bounded and its components satisfy property (P).*

ii) The matrix valued function $\nabla_\theta^2 f(\theta, \cdot)$ is bounded and its components satisfy property (P).

iii) $E[(T \gamma_0(T) \delta)^2] < \infty$ where $\gamma_0(y) = (1 - G(y-))^{-1}$.

Let

$$\tilde{H}^0(y) = \mathbb{P}(T \leq y, \delta = 0), \quad \tilde{H}^{11}(x, y) = \mathbb{P}(X \leq x, T \leq y, \delta = 0)$$

and for any measurable function ϕ , let

$$\begin{aligned} \gamma_1^\phi(y) &= (1 - H(y))^{-1} \int \mathbf{1}_{\{y < t\}} \phi(x, t) \gamma_0(t) d\tilde{H}^{11}(x, t), \\ \gamma_2^\phi(y) &= \int \int (1 - H(v))^{-2} \mathbf{1}_{\{v < y, v < t\}} \phi(x, t) \gamma_0(t) d\tilde{H}^0(v) d\tilde{H}^{11}(x, t). \end{aligned}$$

Consider

$$\phi_0(x, y) = y \nabla_\theta f(\theta_0, x) \tag{5.1}$$

and

$$\tilde{\phi}_0(x, y) = \int_0^y (1 - F(t))^{-1} K(\nabla_\theta f(\theta_0, \cdot), t) dt. \tag{5.2}$$

Note that $|\tilde{\phi}_0(x, y)| \leq Cy$, provided that $|\nabla_\theta f(\theta_0, \cdot)| \leq C$.

Under Assumption 4 and condition (3.2), Stute (1993, 1996)'s results imply

$$\int \phi(x, y) \left\{ d\hat{F}_{(X,Y)}(x, y) - dF_{(X,Y)}(x, y) \right\} = n^{-1} \sum_{i=1}^n \left\{ \eta_i^\phi - E \left[\eta_i^\phi \right] \right\} + o_P(n^{-1/2})$$

with

$$\eta_i^\phi = \delta_i \gamma_0(T_i) \phi(X_i, T_i) + (1 - \delta_i) \gamma_1^\phi(T_i) - \gamma_2^\phi(T_i). \quad (5.3)$$

The role of condition (3.2) is to ensure $\int |y| C^{1/2}(y) dF_{(X,Y)}(x, y) < \infty$, where

$$C(y) = \int_0^{y^-} (1 - H(w))^{-1} (1 - G(w))^{-1} dG(w).$$

See equations (1.6) and (1.10) to (1.12) in Stute (1995).

Now, we come to the asymptotic representation theorem for synthetic data estimators.

Theorem 4 *Assume that θ_0 is an interior point of Θ and let*

$$\Omega = E \left[\nabla_\theta f(\theta_0, X) \nabla_\theta f(\theta_0, X)' \right].$$

Suppose that Assumptions 1 to 3 and 4 hold and that condition (3.2) is satisfied. Let $\phi_0(x, y)$ and $\tilde{\phi}_0(x, y)$ be defined as in (5.1) and (5.2). Then for $\alpha \in \mathbb{R}$,

$$\hat{\theta}_\alpha - \theta_0 = \Omega^{-1} \frac{1}{n} \sum_{i=1}^n \left[\alpha \eta_i^{\phi_0} + (1 + \alpha) \Psi_i^{\theta_0} - f(\theta_0, X_i) \nabla_\theta f(\theta_0, X_i) \right] + o_P(n^{-1/2}) \quad (5.4)$$

where function $\eta_i^{\phi_0}$ is defined as in (5.3), and

$$\Psi_i^{\theta_0} = Y_{i,L}^* + \eta_i^{\tilde{\phi}_0} - E \left[\eta_i^{\tilde{\phi}_0} \right] + \int_0^{\tau_H} \frac{(\mathbf{1}_{\{T_i \leq y\}} - H(y))}{1 - H(y)} K(\nabla_\theta f(\theta_0, \cdot), y) dy.$$

In view of the asymptotic representation (5.4), in order to achieve asymptotic normality, the variance of

$$\int K(\nabla_\theta f(\theta_0, \cdot), y) (1 - H(y))^{-1} (\mathbf{1}_{\{T \leq y\}} - H(y)) dy$$

must be finite. This is the case provided that

$$\int_0^{\tau_H} \left(\int_0^u \frac{|K(\nabla_\theta f(\theta_0, \cdot), v)|}{1 - H(v)} dv \right) |K(\nabla_\theta f(\theta_0, \cdot), u)| du < \infty. \quad (5.5)$$

As $\nabla_\theta f(\theta_0, \cdot)$ is bounded, it is easy to check that (5.5) is implied by the condition $\int y (1 - G(y))^{-1} dF(y) < \infty$. Here, it is implied by Assumption 4 and therefore no additional condition will be needed to deduce the following corollary.

Corollary 5 *Under the same hypothesis as in Theorem 4,*

$$\sqrt{n} \left(\hat{\theta}_\alpha - \theta_0 \right) \implies \mathcal{N} \left(0, \Omega^{-1} \Sigma_\alpha \Omega^{-1} \right)$$

with Σ_α the covariance matrix of $\alpha \eta_i^{\phi_0} + (1 + \alpha) \Psi_i^{\theta_0} - f(\theta_0, X_i)$.

To estimate the limit covariance matrix, one can follow the lines of Stute (1999), that is one can replace the random variables γ_i by their empirical counterparts

$$\begin{aligned}\gamma_0^n(y) &= \frac{1}{1 - \hat{G}(y-)} \\ \gamma_1^{n,\phi}(y) &= \left(1 - \hat{H}(y)\right)^{-1} \int \mathbf{1}_{\{y < w\}} \hat{\phi}(x, w) \gamma_0^n(w) d\tilde{H}_n^{11}(x, w) \\ \gamma_2^{n,\phi}(y) &= \int \int \left(1 - \hat{H}(v)\right)^{-2} \mathbf{1}_{\{v < y, v < w\}} \hat{\phi}(x, w) \gamma_0^n(w) d\tilde{H}_n^0(v) d\tilde{H}_n^{11}(x, w)\end{aligned}$$

with $\tilde{H}_n^0(t) = n^{-1} \sum_{i=1}^n \mathbf{1}_{\{T_i \leq t, \delta_i = 0\}}$ and $\tilde{H}_n^{11}(x, t) = n^{-1} \sum_{i=1}^n \mathbf{1}_{\{X \leq x, T_i \leq t, \delta_i = 1\}}$, and $\hat{\phi}$ is obtained by plugging in $\hat{\theta}_\alpha$ instead of θ_0 . The integrals

$$\int_0^{\tau_H} \left(\mathbf{1}_{\{T_i \leq y\}} - H(y)\right) \left(1 - H(y)\right)^{-1} K(\nabla_{\theta} f(\theta_0, X), y) dy,$$

can be estimated by

$$\int_0^{\tau_H} \left(\mathbf{1}_{\{T_i \leq y\}} - \hat{H}(y)\right) \left(1 - \hat{H}(y)\right)^{-1} \hat{K}(\nabla_{\theta} f(\hat{\theta}_\alpha, \cdot), y) dy,$$

where

$$\hat{K}(\nabla_{\theta} f(\hat{\theta}_\alpha, \cdot), y) = \frac{1}{n} \sum_{i=1}^n \nabla_{\theta} f(\hat{\theta}_\alpha, X_i) \frac{\mathbf{1}_{\{T_i > y\}}}{1 - \hat{G}(T_i-)}.$$

Then covariance can be estimated by the empirical covariance of these estimated random variables. Validity of such techniques are discussed in Stute (1996, 1999) for the γ_i , or in Zhou (1992b) for the second part due to Leurgans' estimator (see remark 3.3 in Zhou, 1992b).

6 Empirical evidence

To assess the performances of these estimators for finite sample sizes and compare them to Stute's estimator (1999), we proceeded to a small empirical experiment. We considered the function $f(\theta, x)$ as in Stute (1999),

$$f(\theta, x) = \frac{1}{\theta_1 + \theta_2} \exp(\theta_1 x_1 + \theta_2 x_2).$$

The true parameter is $\theta_0 = (0.5, 0.3)'$ and $X = (X_1, X_2)'$ consists of two independent random variables with uniform distribution on $[0, 1]$. The purpose of the simulations is to estimate the mean squared error (MSE) $E[\|\hat{\theta} - \theta_0\|^2]$ for each of the following four estimators $\hat{\theta}$: KSV, Leurgans, Stute, and FG estimator, that is an estimator $\hat{\theta}_\alpha$ based on the adaptive choice of α inspired by Fan and Gijbels (1994). Let us recall that Stute (1999)'s estimator can be defined as the value of θ that minimizes

$$\int [y - f(\theta, x)]^2 d\hat{F}_{(X,Y)}(x, y).$$

The sample based α we use for FG estimator is

$$\min_{\{i: \delta_i=1, \hat{Y}_{i,L}^* - T_i > 0\}} \frac{\hat{Y}_{i,L}^* - T_i}{\hat{Y}_{i,KSV}^* - \hat{Y}_{i,L}^*}.$$

This choice is closely related to $\hat{\alpha}$ defined by Fan and Gijbels (1994), equation (11), the difference being that in our framework we do not need to use kernel smoothing for estimating the distribution G .

The sample size was taken $n = 30$ and $n = 100$. For each sample size, 5000 samples of $(Y, X)'$ were simulated. Given a sample of $(Y, X)'$ several levels of censoring were considered. We do not report the results for KSV estimators as they were far worse in all cases considered.

Insert Figure 1 here

In the first simulation experiment, the residual $\varepsilon = Y - f(\theta_0, X)$ was generated from a standard normal variable independent of X . As in Stute (1999), the censoring distribution was uniform on $[0, t]$ where t gives the strength of censorship: large (small) values of t produce light (heavy) censoring. Note that, in this particular setting, Assumption 3 is partially violated, because $\tau_H < \tau_F = \infty$. This was already the case in Stute (1999) who also used this kind of assumption. See discussion in Stute (1999), p.1094. The results of this first simulation study are displayed in Figure 1. Leurgans (nonlinear) estimator outperforms the three competitors, especially in presence of heavy censoring. The poor performance of $\hat{\theta}_\alpha$ with the sample based choice of α is due to the failure of the KSV (nonlinear) estimator.

Insert Figure 2 here

For the second simulation, we considered heteroscedastic normal errors ε with conditional variance

$$E(\varepsilon^2 | X) = (1 + X_1 + X_2)^{-2}.$$

Moreover, the censoring variable C was generated from an exponential law of parameter λ defining the level of censoring. In this setup, Assumption 3 is fulfilled. With such a configuration, the observations $(Y_i, X_i)'$ with large $f(\theta, X_i)$ are very informative, but more subject to censoring. By definition, Stute's estimator partly overlooks the information carried by these observations when they are censored. Then, Leurgans (nonlinear) estimator is expected to behave better. This intuition is confirmed by the results plotted in Figure 2: Stute is worse than Leurgans and FG estimator. In this case, FG estimator performs well, perhaps due to the slightly better behavior of KSV estimator.

7 Conclusion

The linear regression model with a right censored dependent variable received extensive attention in survival data analysis. On contrary, nonlinear (parametric) regression with censored responses was quite little investigated. In this paper we propose a new way

to estimate the parameters of the nonlinear regression in the presence of censoring. A class of estimators is introduced. The estimators are obtained by least squares based on synthetic responses. Heteroscedastic errors with unknown conditional variance and discrete covariates are allowed. A small empirical investigation indicates that our estimators outperform the estimator of Stute (1999) in finite samples. This can be explained by the fact that Stute's estimator drops a part of the information carried by the censored observations, while the estimators we propose use this information.

The synthetic responses are obtained from the original observations through a transformation that depends on data. See Koul, Susarla and Van Ryzin (1981) and Leurgans (1987). Consequently, the synthetic observations are not i.i.d. even if the original data were. This complicates the asymptotic analysis of least squares estimators. In the linear regression case, martingale arguments were used to deal with this problem. Such arguments heavily relies on the particular shape of the regression function. Our idea is to derive new representations of empirical means involving synthetic observations using Kaplan-Meier integrals. With at hand such representations and given the i.i.d. representations of the Kaplan-Meier integrals obtained by Stute (1993, 1996), the asymptotic properties of the least squares estimator becomes straight. The extension to a weighted least squares procedure designed to yield less variable estimates in the presence of heteroscedastic errors is also quite easy. We argue that our approach is potentially even more valuable and can be also applied to semi- or nonparametric regression models. This last issue will be investigated in future work.

A Appendix

In the following, to simplify notation we only deal with positive random variables T_i . The general case does not rise any problem since the difficulty stands in the vicinity of τ_H .

Proof of Theorem 3. For the sake of simplicity, we only consider case $\alpha = 0$ and $\alpha = -1$ corresponding to Leurgans and KSV-estimator respectively. The case $\alpha \in \mathbb{R}$ can be directly adapted from this proof.

By Assumption 2, $\theta_0 = \arg \min_{\theta \in \Theta} \Gamma(\theta)$, where $\Gamma(\theta) = E[f^2(\theta, X)] - 2E[Yf(\theta, X)]$, and $\hat{\theta}_\alpha = \arg \min_{\theta \in \Theta} \Gamma_n^\alpha(\theta)$, where $\Gamma_n^\alpha(\theta) = n^{-1} \sum_i f^2(\theta, X_i) - 2n^{-1} \sum_i \hat{Y}_{i,\alpha}^* f(\theta, X_i)$. Assumption 1 ensures that $\{f^2(\theta, \cdot) : \theta \in \Theta\}$ satisfies a uniform law of large numbers (e.g., Pakes and Pollard, 1989, or van der Vaart and Wellner, 1996). Deduce that $n^{-1} \sum_i f^2(\theta, X_i) \rightarrow E[f^2(\theta, X)]$ almost surely (a.s.) uniformly over Θ .

For the case $\alpha = -1$, use the representation of Lemma 1, to write

$$2n^{-1} \sum_i \hat{Y}_{i,KSV}^* f(\theta, X_i) = 2 \int y f(\theta, x) d\hat{F}_{(X,Y)}(x, y).$$

Now, as in Stute (1999), apply an uniform version of Stute (1993)'s law of large numbers to conclude that $\sup_{\theta \in \Theta} |\Gamma_n^\alpha(\theta) - \Gamma(\theta)| \rightarrow 0$ a.s., and thus to obtain the consistency.

For the case $\alpha = 0$, we could also use the asymptotic properties of the Kaplan-Meier integrals. However, if we were to use a uniform version of the representation in Lemma 2 with the reminder term R_n of order $o_P(n^{-1/2})$, the conditions imposed to derive consistency would be too strong. Instead of proving a representation like in Lemma 2

but with a reminder of order $o_P(1)$ uniformly in θ , for the sake of simplicity, we provide a more direct argument in this case. Rewrite

$$\frac{1}{n} \sum_i \hat{Y}_{i,L}^* f(\theta, X_i) = \frac{1}{n} \sum_i \left(\hat{Y}_{i,L}^* - Y_{i,L}^* \right) f(\theta, X_i) + \frac{1}{n} \sum_i Y_{i,L}^* f(\theta, X_i).$$

The second sum converges to the desired limit uniformly over Θ , while the first sum is of order $o_P(1)$ uniformly over Θ as a consequence of Cauchy-Schwarz inequality, Lemma A.1 below and the uniform law of large numbers. ■

Lemma A.1 *Assume that $E[Y_L^{*2}] < \infty$ and G is continuous. Then*

$$\frac{1}{n} \sum_{i=1}^n \left(\hat{Y}_{i,L}^* - Y_{i,L}^* \right)^2 = o_P(1).$$

Proof. Write

$$\begin{aligned} \hat{Y}_{i,L}^* - Y_{i,L}^* &= \int_0^{\tau_H} \frac{\mathbf{1}_{\{T_i > t\}}}{1 - G(t)} \frac{\hat{G}(t) - G(t)}{1 - \hat{G}(t)} dt \\ &\leq \sup_{t \leq T_{(n-k_n)}} \left| \frac{\hat{G}(t) - G(t)}{1 - \hat{G}(t)} \right| \int_0^{T_{(n-k_n)}} \frac{\mathbf{1}_{\{T_i > t\}}}{1 - G(t)} dt \\ &\quad + \sup_{t \leq T_{(n)}} \left| \frac{\hat{G}(t) - G(t)}{1 - \hat{G}(t)} \right| \int_{T_{(n-k_n)}}^{\tau_H} \frac{\mathbf{1}_{\{T_i > t\}}}{1 - G(t)} dt, \end{aligned}$$

where $T_{(j)}$ denotes the j -th order statistics of the sample T_1, \dots, T_n . Then, for each $i = 1, \dots, n$

$$\begin{aligned} \left(\hat{Y}_{i,L}^* - Y_{i,L}^* \right)^2 &\leq 2 \left(\sup_{t \leq T_{(n-k_n)}} \left| \frac{\hat{G}(t) - G(t)}{1 - \hat{G}(t)} \right| \right)^2 \left(\int_0^{T_{(n-k_n)}} \frac{\mathbf{1}_{\{T_i > t\}}}{1 - G(t)} dt \right)^2 \\ &\quad + 2 \left(\sup_{t \leq T_{(n)}} \left| \frac{\hat{G}(t) - G(t)}{1 - \hat{G}(t)} \right| \right)^2 \left(\int_{T_{(n-k_n)}}^{\tau_H} \frac{\mathbf{1}_{\{T_i > t\}}}{1 - G(t)} dt \right)^2. \end{aligned}$$

As $\mathbf{1}_{\{T_i > t\}} (1 - G(t))^{-1}$ is positive, $T_{(n-k_n)}$ can be replaced by τ_H in the first integral of the last display. Summing the inequalities for each i yields

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \left(\hat{Y}_{i,L}^* - Y_{i,L}^* \right)^2 &\leq 2 \left(\sup_{t \leq T_{(n-k_n)}} \left| \frac{\hat{G}(t) - G(t)}{1 - \hat{G}(t)} \right| \right)^2 \sum_{i=1}^n \frac{Y_{i,L}^{*2}}{n} \tag{A.1} \\ &\quad + \frac{2}{n} \left(\sup_{t \leq T_{(n)}} \left| \frac{\hat{G}(t) - G(t)}{1 - \hat{G}(t)} \right| \right)^2 \sum_{i=1}^n \left(\int_{T_{(n-k_n)}}^{\tau_H} \frac{\mathbf{1}_{\{T_i > t\}}}{1 - G(t)} dt \right)^2. \end{aligned}$$

By Csörgö's theorem (Csörgö, 1996)

$$\sup_{x \leq T_{(n-k_n)}} \left| \frac{\hat{G}(x) - G(x)}{1 - \hat{G}(x)} \right| = O_P(k_n^{-1/2}),$$

provided that G is continuous. Applying Csörgö's theorem with $k_n = \lfloor (\log n)^2 \rfloor$ and using the law of large numbers for $Y_{i,L}^{*2}$, the first term on the right hand side of (A.1) is easily seen to converge to zero in probability. Since

$$\sup_{t \leq T_{(n)}} \left| \left(\hat{G}(t) - G(t) \right) \left(1 - \hat{G}(t) \right)^{-1} \right| = O_P(1),$$

(e.g. Zhou, 1992b, equation (5.9)) to prove that the second term on the right hand side of (A.1) converges to zero in probability, it suffices to show

$$S_n = n^{-1} \sum_{i=1}^n \left(\int_{T_{(n-k_n)}}^{\tau_H} \mathbf{1}_{\{T_i > t\}} (1 - G(t))^{-1} dt \right)^2 \rightarrow 0,$$

in probability. Let a_n be a sequence tending to infinity with $\mathbb{P}(T_{(n-k_n)} \geq a_n) \rightarrow 0$. Then, we have with probability tending to 1

$$S_n \leq n^{-1} \sum_{i=1}^n \left(\int_{a_n}^{\tau_H} \mathbf{1}_{\{T_i > t\}} (1 - G(t))^{-1} dt \right)^2.$$

Taking the expectation of the two parts leads to

$$E[S_n] \leq E \left[\left(\int_{a_n}^{\tau_H} \mathbf{1}_{\{T > t\}} (1 - G(t))^{-1} dt \right)^2 \right]$$

with the right part tending to zero with probability 1 in view of $E[Y_L^{*2}] < \infty$ and dominated convergence theorem. As S_n appears to be a sum of positive quantities, it follows that $S_n = o_P(1)$ and this completes the proof. ■

Proof of Lemma 2. First, simplify notation: $\hat{F}(t)$ is the Kaplan-Meier estimator of $F(t) = \mathbf{P}(Y \leq t)$. Write

$$\hat{Y}_L^* - Y_L^* = \int_0^{\tau_H} \frac{(F(t) - \hat{F}(t)) \mathbf{1}_{\{T > t\}}}{1 - H(t)} dt + I(T)$$

with

$$I(T) = \int_0^{\tau_H} \frac{(1 - \hat{F}(t)) \mathbf{1}_{\{T > t\}}}{1 - H(t)} \left(\frac{\hat{H}(t) - H(t)}{1 - \hat{H}(t)} \right) dt.$$

Next, decompose

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n g(X_i) \left(\hat{Y}_{i,L}^* - Y_{i,L}^* \right) &= \int_0^{\tau_H} \frac{K(g, t)}{1 - F(t)} \left(F(t) - \hat{F}(t) \right) dt \\
&+ \frac{1}{n} \sum_{i=1}^n g(X_i) I(T_i) + \int_0^{\tau_H} R_{0n}(t) dt \\
&= \int_0^{\tau_H} \left(\int_0^y \frac{K(g, t)}{1 - F(t)} dt \right) \left\{ d\hat{F}(y) - dF(y) \right\} \\
&+ \frac{1}{n} \sum_{i=1}^n g(X_i) I(T_i) + \int_0^{\tau_H} R_{0n}(t) dt,
\end{aligned}$$

where the last equality can be obtained by Fubini's theorem. On the other hand, rewrite

$$I(T) = \int_0^{\tau_H} \frac{(1 - \hat{F}(t)) \mathbf{1}_{\{T > t\}}}{(1 - H(t))^2} \left(\hat{H}(t) - H(t) \right) dt + R_{1n}(T). \quad (\text{A.2})$$

Replacing \hat{F} by F in equation (A.2) leads to

$$I(T) = \int_0^{\tau_H} \frac{\mathbf{1}_{\{T > t\}}}{(1 - H(t))(1 - G(t))} \left(\hat{H}(t) - H(t) \right) dt + R_{1n}(T) + R_{2n}(T).$$

Now, rewrite

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n g(X_i) I(T_i) &= \int_0^{\tau_H} \frac{n^{-1} \sum_{i=1}^n g(X_i) \mathbf{1}_{\{T_i > t\}}}{(1 - H(t))(1 - G(t))} \left(\hat{H}(t) - H(t) \right) dt \\
&+ \frac{1}{n} \sum_{i=1}^n g(X_i) (R_{1n}(T_i) + R_{2n}(T_i)) \\
&= \int_0^{\tau_H} \frac{K(g, t)}{1 - H(t)} \left(\hat{H}(t) - H(t) \right) dt \\
&+ \int_0^{\tau_H} \frac{R_{3n}(t)}{(1 - H(t))(1 - G(t))} \left(\hat{H}(t) - H(t) \right) dt \\
&+ \frac{1}{n} \sum_{i=1}^n g(X_i) (R_{1n}(T_i) + R_{2n}(T_i)).
\end{aligned}$$

We can write

$$\int_0^{\tau_H} \frac{K(g, t)}{1 - H(t)} \left(\hat{H}(t) - H(t) \right) dt = - \int_0^{\tau_H} \frac{K(g, t)}{1 - H(t)} \left(\{1 - \hat{H}(t)\} - \{1 - H(t)\} \right) dt$$

which shows that the integral on the left side is almost surely finite. It remains to prove that the reminders R_{0n} to R_{3n} are negligible. For this, recall that

$$\text{for any } a > 0, \quad \sup_t \left| \hat{H}(t) - H(t) \right| (1 - H(t))^{-1/2+a} = O_P(n^{-1/2}), \quad (\text{A.3})$$

$$\sup_t (1 - H(t)) \left(1 - \hat{H}(t) \right)^{-1} = O_P(1), \quad (\text{A.4})$$

$$\sup_t \left(1 - \hat{F}(t) \right) (1 - F(t))^{-1} = O_P(1). \quad (\text{A.5})$$

See, for instance, Shorack and Wellner (1986) and Zhou (1992b). Moreover, notice that condition (3.2) implies

$$\int_0^{\tau_H} \frac{(1 - F(t))^{1-\varepsilon}}{(1 - H(t))^{1/2}} dt < \infty \text{ and } \int_0^{\tau_H} \frac{(1 - F(t))^{1/2}}{(1 - G(t))^{1/2+\varepsilon}} dt < \infty, \quad (\text{A.6})$$

for some $\varepsilon > 0$. Indeed,

$$\begin{aligned} J &\leq \int_0^{\tau_H} \frac{(1 - F(t))^{1/2-\varepsilon}}{(1 - G(t))^{1/2+\varepsilon}} dt = \left(\frac{1}{2} - \varepsilon\right) \int_0^{\tau_H} \int_0^t \frac{ds}{(1 - H(s))^{1/2+\varepsilon}} dF(t) \\ &\leq \left(\frac{1}{2} - \varepsilon\right) \int_0^{\tau_H} \frac{tdF(t)}{(1 - H(t))^{1/2+\varepsilon}} < \infty \end{aligned}$$

with J any of the integrals in (A.6).

The order of R_{0n} . We have to prove that

$$\int_0^{\tau_H} \frac{n^{-1} \sum_i g(X_i) \mathbf{1}_{\{T_i > t\}} - E[g(X) \mathbf{1}_{\{T > t\}}]}{1 - H(t)} \left(F(t) - \hat{F}(t)\right) dt = o_P(n^{-1/2}).$$

For some fixed $\varepsilon > 0$, the family of functions

$$\left\{ (x, t) \rightarrow g(x) \mathbf{1}_{\{t > s\}} (1 - H(s))^{-1/2+\varepsilon} : s \geq 0 \right\}$$

is Euclidean for a square integrable envelope (see Pakes and Pollard, 1989, for the definition of a Euclidean family of functions; see also example 2.6.23 in van der Vaart and Wellner, 1996), so that

$$n^{-1} \sum_i [g(X_i) \mathbf{1}_{\{T_i > t\}} - E(g(X) \mathbf{1}_{\{T > t\}})] (1 - H(t))^{-1/2+\varepsilon} = O_P(n^{-1/2})$$

uniformly in t . From Stute (1993) and equation (A.5) above,

$$\begin{aligned} \sup_t \frac{|F(t) - \hat{F}(t)|}{(1 - F(t))^{1-\varepsilon}} &\leq \left[\sup_t |F(t) - \hat{F}(t)| \right]^\varepsilon \left[\sup_t \frac{|F(t) - \hat{F}(t)|}{1 - F(t)} \right]^{1-\varepsilon} \\ &= o_P(1) O_P(1) = o_P(1). \end{aligned} \quad (\text{A.7})$$

Therefore

$$\left| \int_0^{\tau_H} R_{0n}(t) \right| \leq o_P(n^{-1/2}) \int_0^{\tau_H} \frac{(1 - F(t))^{1/2-2\varepsilon}}{(1 - G(t))^{1/2+2\varepsilon}} dt$$

and the last integral above is finite by the arguments used to prove (A.6).

The order of R_{1n} . We have

$$R_{1n}(T) = \int_0^{\tau_H} \frac{(1 - \hat{F}(t)) \mathbf{1}_{\{T > t\}}}{(1 - H(t))^2} \left(\frac{\hat{H}(t) - H(t)}{\sqrt{1 - \hat{H}(t)}} \right)^2 dt.$$

Rewrite, for any $\varepsilon \in (0, 1)$

$$R_{1n}(T) = \int_0^{\tau_H} \left(\frac{\hat{H}(t) - H(t)}{\{1 - H(t)\}^{1/2 - \varepsilon/4}} \right)^{1+\varepsilon} \frac{(1 - F(t)) \mathbf{1}_{\{T > t\}}}{(1 - H(t))^{3/2 + \varepsilon/2 + (1+\varepsilon)\varepsilon/4}} \left(\frac{1 - \hat{F}(t)}{1 - F(t)} \right) \\ \times \left(\frac{\hat{H}(t) - H(t)}{1 - H(t)} \right)^{1-\varepsilon} \left(\frac{1 - H(t)}{1 - \hat{H}(t)} \right) dt.$$

By equations (A.3) to (A.5),

$$\left| \frac{1}{n} \sum_{i=1}^n g(X_i) R_{1n}(T_i) \right| \leq O_P(n^{-1/2 - \varepsilon/2}) \int_0^{\tau_H} \frac{(1 - F(t)) n^{-1} \sum_{i=1}^n |g(X_i)| \mathbf{1}_{\{T_i > t\}}}{(1 - H(t))^{3/2 + \varepsilon/2 + (1+\varepsilon)\varepsilon/4}} dt.$$

Now, observe that for a given $\varepsilon \in (0, 1)$, the family of functions

$$h_s(X, T) = g(X) \frac{\mathbf{1}_{\{T > s\}}}{(1 - H(s))^{1-\varepsilon}} - E \left[g(X) \frac{(1 - H(t, X))}{(1 - H(t))^{1-\varepsilon}} \right], \quad s \geq 0, \quad (\text{A.8})$$

is an Euclidean family for the integrable envelope $g(X) (1 - H(T))^{-1+\varepsilon}$ (one more time, see van der Vaart and Wellner, 1996). By the uniform law of large numbers,

$$\sup_{s \geq 0} \left| \frac{1}{n} \sum_{i=1}^n h_s(X_i, T_i) \right| \rightarrow 0$$

almost surely. As g is bounded and $E[1 - H(t, X)] = 1 - H(t)$, by (A.6)

$$\left| \frac{1}{n} \sum_{i=1}^n g(X_i) R_{1n}(T_i) \right| \leq O_P(n^{-1/2 - \varepsilon/2}) \left\{ o_P(1) + \int_0^{\tau_H} \frac{(1 - F(t))^{1/2}}{(1 - G(t))^{1/2 + \varepsilon}} dt \right\} \\ = O_P(n^{-1/2 - \varepsilon/2}) = o_P(n^{-1/2}).$$

The order of R_{2n} . As g is bounded, for some $C > 0$

$$\left| \frac{1}{n} \sum_{i=1}^n g(X_i) R_{2n}(T_i) \right| \leq C \int_0^{\tau_H} \frac{|F(t) - \hat{F}(t)|}{(1 - H(t))^{1/2 + \varepsilon}} \frac{(1 - \hat{H}(t))}{(1 - H(t))} \frac{|\hat{H}(t) - H(t)|}{(1 - H(t))^{1/2 - \varepsilon}} dt.$$

By (A.3) and (A.4), for any $\varepsilon \in (0, 1)$

$$\left| \frac{1}{n} \sum_{i=1}^n g(X_i) R_{2n}(T_i) \right| \leq O_P(n^{-1/2}) \int_0^{\tau_H} \frac{|F(t) - \hat{F}(t)|}{(1 - F(t))^{1-\varepsilon}} \frac{(1 - F(t))^{1-\varepsilon}}{(1 - H(t))^{1/2 + \varepsilon}} dt \\ = o_P(n^{-1/2})$$

where for the last equality we use (A.6) and (A.7).

The order of R_{3n} . By the uniform law of large numbers applied with the family defined in (A.8)

$$\sup_{s \geq 0} \frac{|R_{3n}(s)|}{(1 - H(s))^{1-\varepsilon}} = o_P(1).$$

Now, by (A.3) and (A.6)

$$\begin{aligned}
\left| \int_0^{\tau_H} \frac{R_{3n}(t)}{1-G(t)} \frac{\hat{H}(t) - H(t)}{1-H(t)} dt \right| &\leq o_P(n^{-1/2}) \int_0^{\tau_H} \frac{(1-H(t))^{1/2-\varepsilon/2}}{(1-G(t))} dt \\
&= o_P(n^{-1/2}) \int_0^{\tau_H} \frac{(1-F(t))^{1/2-\varepsilon/2}}{(1-G(t))^{1/2+\varepsilon/2}} dt \\
&= o_P(n^{-1/2}).
\end{aligned}$$

■

Proof of Theorem 4. For simplicity, only the case $\theta \in \mathbb{R}$ is considered, the extension to multidimensional parameter being straight. Moreover, it suffices to consider $\alpha = 0$ and $\alpha = -1$, the general case following by a simple convex combination. A Taylor expansion of $\nabla_{\theta} M_n^{\alpha}$ yields

$$\hat{\theta}_{\alpha} - \theta_0 = - (\nabla_{\theta}^2 M_n^{\alpha}(\theta_{1n}))^{-1} \nabla_{\theta} M_n^{\alpha}(\theta_0),$$

for some θ_{1n} between θ_0 and $\hat{\theta}_{\alpha}$. The proof is complete if for $\alpha = -1$ and $\alpha = 0$

$$\nabla_{\theta}^2 M_n^{\alpha}(\theta_{1n}) = \frac{2}{n} \sum_{i=1}^n [\nabla_{\theta} f(\theta_{1n}, X_i)]^2 - \frac{2}{n} \sum_{i=1}^n \nabla_{\theta}^2 f(\theta_{1n}, X_i) (\hat{Y}_{i,\alpha}^* - f(\theta_{1n}, X_i)) \rightarrow \Omega$$

in probability and $\nabla_{\theta} M_n^{\alpha}(\theta_0)$ admits an i.i.d. representation. For $\alpha = -1$,

$$\begin{aligned}
\nabla_{\theta}^2 M_n^{\alpha}(\theta_{1n}) &= 2 \int \{[\nabla_{\theta} f(\theta_{1n}, x)]^2 + \nabla_{\theta}^2 f(\theta_{1n}, x) f(\theta_{1n}, x)\} d\hat{F}_X(x) \\
&\quad - 2 \int y \nabla_{\theta}^2 f(\theta_{1n}, x) d\hat{F}_{(X,Y)}(x, y).
\end{aligned}$$

As for the first integral in the last display we can apply a classical uniform law of large numbers for uncensored data, to obtain the convergence of $\nabla_{\theta}^2 M_n^{\alpha}(\theta_{1n})$ towards Ω in probability it remains to investigate the second integral. By Assumption 4(ii), the convergence of Kaplan-Meier integrals and the consistency of $\hat{\theta}_{\alpha}$ we have

$$\begin{aligned}
\left| \int y \nabla_{\theta}^2 f(\theta_{1n}, x) d\hat{F}_{(X,Y)}(x, y) \right| &\leq \int y |\nabla_{\theta}^2 f(\theta_{1n}, x) - \nabla_{\theta}^2 f(\theta_0, x)| d\hat{F}_{(X,Y)}(x, y) \\
&\quad + E [Y \nabla_{\theta}^2 f(\theta_0, X)] + o_P(1) \\
&\leq |\theta_{1n} - \theta_0| \int y \Phi(x) d\hat{F}_{(X,Y)}(x, y) \\
&\quad + E [Y \nabla_{\theta}^2 f(\theta_0, X)] + o_P(1) \\
&= |\theta_{1n} - \theta_0| \{E[|Y| \Phi(X)] + o_P(1)\} \\
&\quad + E [Y \nabla_{\theta}^2 f(\theta_0, X)] + o_P(1) \\
&= E [Y \nabla_{\theta}^2 f(\theta_0, X)] + o_P(1).
\end{aligned}$$

Deduce that $\nabla_{\theta}^2 M_n^{\alpha}(\theta_{1n}) \rightarrow \Omega$ in probability when $\alpha = -1$. For $\alpha = 0$, proceed as in Theorem 3 and use Lemma A.1 to obtain

$$\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^n \nabla_{\theta}^2 f(\theta, X_i) \hat{Y}_{i,L}^* - \frac{1}{n} \sum_{i=1}^n \nabla_{\theta}^2 f(\theta, X_i) Y_{i,L}^* \right| = o_P(1).$$

As $\theta_{1n} \rightarrow \theta_0$ in probability, deduce that $n^{-1} \sum_i \nabla_{\theta}^2 f(\theta_{1n}, X_i) \hat{Y}_{i,L}^* \rightarrow E[\nabla_{\theta}^2 f(\theta_0, X_i) Y_L^*]$. Therefore $\nabla_{\theta}^2 M_n^{\alpha}(\theta_{1n}) \rightarrow \Omega$ in probability.

On the other hand, by our Lemma 1 and the results of Stute (1996)

$$\nabla_{\theta} M_n^{\alpha}(\theta_0) = \frac{1}{n} \sum_{i=1}^n \left(\eta_i^{\phi_0} - f(\theta_0, X_i) \nabla_{\theta} f(\theta_0, X_i) \right) + o_P(n^{-1/2})$$

when $\alpha = -1$. For $\alpha = 0$, apply Lemma 2 to obtain

$$\begin{aligned} \nabla_{\theta} M_n^{\alpha}(\theta_0) &= \frac{1}{n} \sum_{i=1}^n (Y_{i,L}^* - f(\theta_0, X_i)) \nabla_{\theta} f(\theta_0, X_i) \\ &\quad + \int \left(\int_0^y \frac{K(\nabla_{\theta} f(\theta_0, \cdot), s)}{1 - F(s)} ds \right) (d\hat{F}_Y(y) - dF_Y(y)) \\ &\quad + \int \frac{K(\nabla_{\theta} f(\theta_0, \cdot), y)}{1 - H(y)} \left(\frac{1}{n} \sum_{i=1}^n \{ \mathbf{1}_{\{T_i \leq y\}} - H(y) \} \right) dy \\ &\quad + o_P(n^{-1/2}). \end{aligned}$$

Use the i.i.d. representation of Kaplan-Meier integrals (Stute, 1996) to complete the proof. ■

References

- [1] Buckley, J., and James, I. R. (1979). Linear regression with censored data. *Biometrika* **66**, 429–436.
- [2] Csörgő, S. (1996). Universal Gaussian approximations under random censorship. *Ann. Statist.* **24**, 2744–2778.
- [3] Fan, J., and Gijbels, I. (1994). Censored regression: local linear approximations and their applications. *J. Amer. Statist. Assoc.* **89**, 560–570.
- [4] Heuchenne, C., and Van Keilegom, I. (2005). Estimation in nonparametric location-scale regression models withensored data. Discussion Paper (DP 0518), Institute of Statistics, Louvain-la-Neuve.
- [5] Heuchenne, C., and Van Keilegom, I. (2006). Polynomial regression with censored data based on preliminary nonparametric estimation. *Ann. Inst. Statist. Math.* (to appear).
- [6] Koul, H., Susarla, V., and Van Ryzin, J. (1981). Regression analysis with randomly right censored data, *Ann. Statist.* **9**, 1276–1288.
- [7] Leurgans, S. (1987). Linear models, random censoring and synthetic data. *Biometrika* **74**, 301–309.
- [8] Pakes, A., and Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica* **57**, 1027–1057.

- [9] Satten, G. A., and Datta, S. (2001). The Kaplan-Meier estimator as an inverse-probability-of-censoring weighted average. *Amer. Statist.* **55**, 207–210.
- [10] Shorack, G. R., and Wellner, J. A. (1986). *Empirical processes with applications to statistics*. Wiley, New York.
- [11] Stute, W. (1993). Consistent estimation under random censorship when covariables are present. *J. Multivariate Anal.* **45**, 89–103.
- [12] Stute, W. (1995). The central limit theorem under random censorship. *Ann. Statist.* **23**, 422–439.
- [13] Stute, W. (1996). Distributional convergence under random censorship when covariables are present. *Scand. J. Statist.* **23**, 461–471.
- [14] Stute, W. (1999). Nonlinear censored regression. *Statistica Sinica* **9**, 1089–1102.
- [15] van der Vaart, A.W., and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag, New-York.
- [16] Zhou, M. (1992a). M-estimation in censored linear models. *Biometrika* **79**, 837–841.
- [17] Zhou, M. (1992b). Asymptotic normality of the "synthetic data" regression estimator for censored survival data. *Ann. Statist.* **20**, 1002–1021.

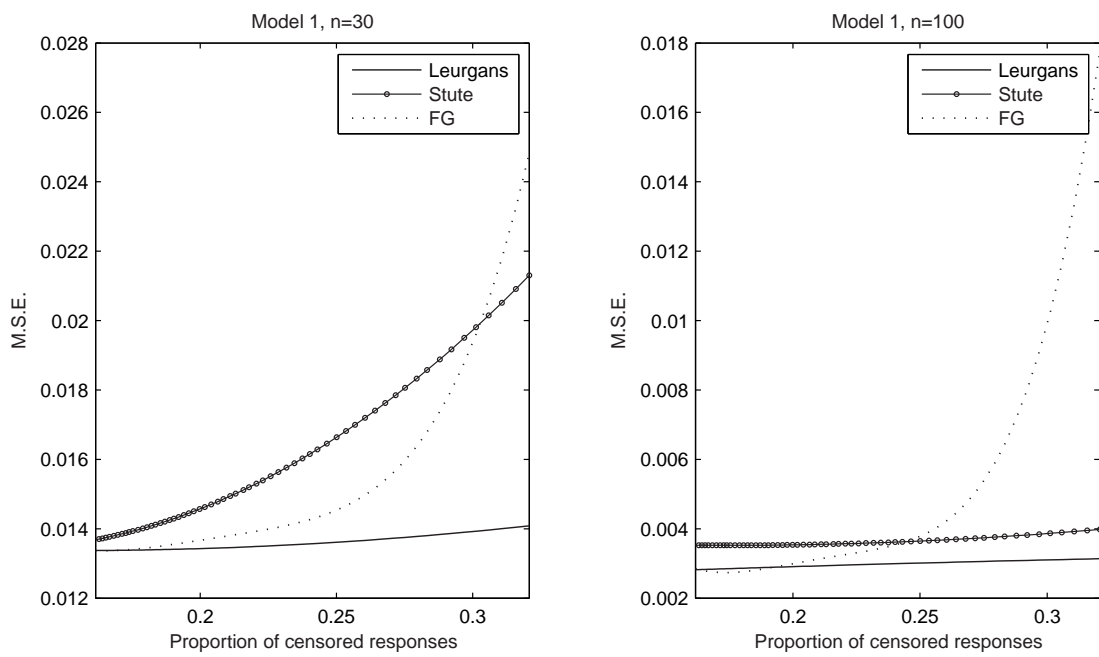


Figure 1: Mean squared error of Stute, Leurgans and FG estimators with homoscedastic errors.

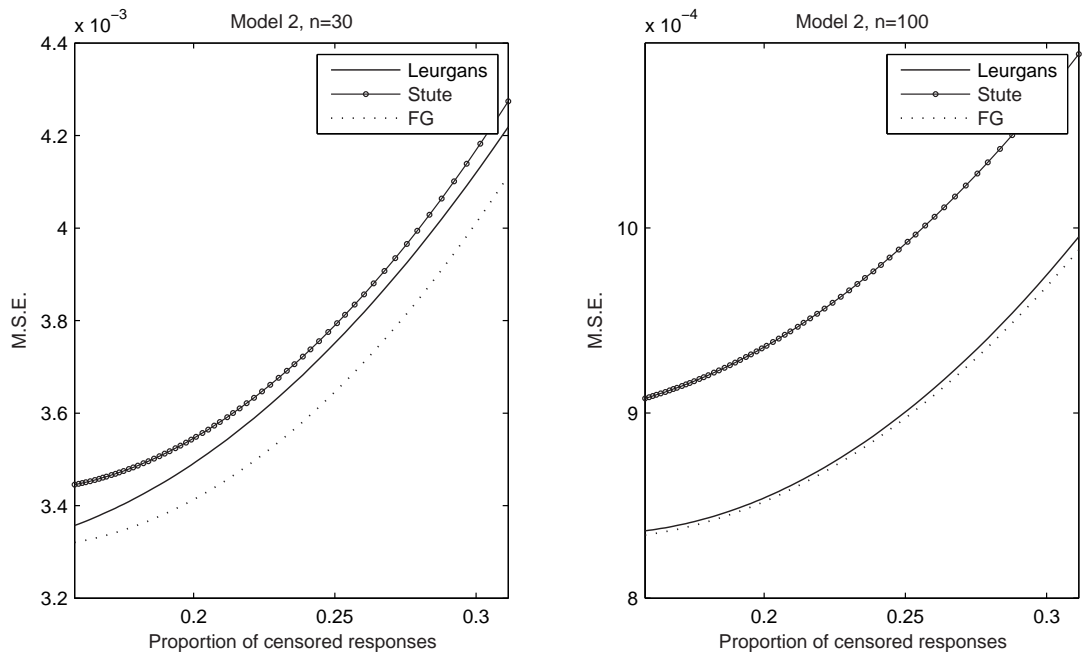


Figure 2: Mean squared error of Stute, Leurgans and FG estimators with heteroscedastic errors.