Electronic Journal of Statistics
Vol. 11 (2017) 2773–2799
ISSN: 1935-7524
DOI: 10.1214/17-EJS1293

Parametrically guided local quasi-likelihood with censored data

Majda Talamakrouni*, **Anouar El Ghouch***

Institute of Statistics, Biostatistics and Actuarial Sciences
Université catholique de Louvain
Louvain-la-Neuve, Belgium
e-mail: majda.talamakrouni@uclouvain.be; anouar.elghouch@uclouvain.be

and

Ingrid Van Keilegom*,†

ORSTAT, KU Leuven, Leuven, Belgium
Institute of Statistics, Biostatistics and Actuarial Sciences,
Université catholique de Louvain
Louvain-la-Neuve, Belgium
e-mail: ingrid.vankeilegom@kuleuven.be

Abstract: It is widely pointed out in the literature that misspecification of a parametric model can lead to inconsistent estimators and wrong inference. However, even a misspecified model can provide some valuable information about the phenomena under study. This is the main idea behind the development of an approach known, in the literature, as parametrically guided nonparametric estimation. Due to its promising bias reduction property, this approach has been investigated in different frameworks such as density estimation, least squares regression and local quasi-likelihood. Our contribution is concerned with parametrically guided local quasi-likelihood estimation adapted to randomly right censored data. The generalization to censored data involves synthetic data and local linear fitting. The asymptotic properties of the guided estimator as well as its finite sample performance are studied and compared with the unguided local quasi-likelihood estimator. The results confirm the bias reduction property and show that, using an appropriate guide and an appropriate bandwidth, the proposed estimator outperforms the classical local quasi-likelihood estimator.

Keywords and phrases: Beran's estimator, generalized linear model, local linear smoothing, parametric guide, quasi-likelihood method, right censoring, synthetic data.

Received March 2016.

1. Introduction

The concept of quasi-likelihood estimation was proposed by [43] as a flexible extension of the maximum likelihood estimation method for generalized linear

*M. Talamakrouni, A. El Ghouch and I. Van Keilegom all acknowledge financial support from IAP research network P7/06 of the Belgian Government (Belgian Science Policy).

†I. Van Keilegom also acknowledges support from the European Research Council (2016-2021, Horizon 2020/ERC grant agreement No. 694409).

models (GLMs). The latter, as introduced by [32], relies on strong parametric assumption about the distribution of the data that can be hard to verify in practice. In such situations, the quasi-likelihood estimation may be a suitable alternative, since it relies only on assumptions about the first two moments. Moreover, the quasi-likelihood function has similar properties as the classical full log-likelihood function; see [31] for more details.

Likelihood and quasi-likelihood provide consistent and powerful estimators if the required assumptions imposed on the data are met. However, a misspecified model can create an important bias in the estimation of the underlying target function. For this reason, nonparametric techniques, that are more robust, have been investigated by many studies. This include [20], [34], [37], [23], and [36], to cite just a few examples. More recent contributions include the work of [14] and [11], who investigated local polynomial fitting for likelihood and quasi-likelihood in the context of GLMs, [4], who studied local quasi-likelihood for missing data.

Even when the proposed model is misspecified, parametric estimation can provide a useful information about the target function. This information can be injected into a nonparametric estimator in order to improve its performance in terms of bias and mean-squared error (MSE). In the literature, there exists an attractive method that allows for that, namely the parametrically guided nonparametric estimation. In contrast to a traditional semi-parametric method, a parametrically guided estimator is fully nonparametric in the sense that no global parametric structure is imposed on the data. In the complete data case, considerable attention has recently been paid to this approach in the literature. First, [22] introduced the parametric guided kernel scheme for density estimation. Then, [17], [18] and [30] investigated this method for mean regression function. Later, the same approach has been extended to GLMs and local quasi-likelihood by [16]. Very recently, [15] applied the guided estimation to generalized additive models and [7] studied the guided estimation for varying coefficient models. These papers noticed and showed the interesting property of bias reduction for their guided nonparametric estimators compared with the unguided ones without any increase in the variance.

There exist three different schemes allowing to guide parametrically a nonparametric estimator. The first scheme has been developed by [22] using a multiplicative correction that requires a nonzero value for the parametric part which is not always respected in practice. In the second scheme, the correction is carried out in an additive rather than a multiplicative scale. Such guided scheme has been introduced in kernel regression by [35] and used later in different frameworks. Finally, the last guided scheme combines both the additive and the multiplicative scheme in a unified family indexed by a calibration parameter that controls the balance between the two corrections. The unified family has the advantage of being more general than the two other corrections. However, the additional calibration parameter needs to be selected, which is not an easy task. In the context of local quasi-likelihood, [16] studied in detail the three different schemes. For the sake of simplicity, we first restrict our attention to the additive scheme, and next we give an extension of our results to the unified family of corrections. In the following, we give a brief description of the guided additive scheme in the context of local quasi-likelihood, more details are provided in Section 2.

Suppose for the moment that we have completely observed and i.i.d. data $(Y_i, X_i), i = 1, \ldots, n$, and let $m(x) = E(Y|X = x)$ be the true mean regression function. In classical parametric GLMs, $m(x)$ is modeled linearly using a known link function $g(\cdot)$, that is $g(m(x)) = \eta(x)$, with $\eta(x) = \theta_0 + \theta_1 x$. The parameter of interest $\theta = (\theta_0, \theta_1)^T$ can be estimated via the likelihood or the quasi-likelihood. However, in practice, the linearity assumption is not met in many situations. In such cases, local quasi-likelihood is more appropriate since it allows the estimation of $\eta(x)$ without any explicit specification of its form. In between these two "extreme" approaches, [16] proposed a guided local quasi-likelihood estimator with the objective of combining the advantages of both parametric and local quasi-likelihood estimators. As stated before, we focus on the additively guided local quasi-likelihood estimator. The additive scheme starts with a parametric quasi-likelihood estimator which is not necessarily correctly specified. Then, in a second step, this crude parametric approximation is adjusted using a local quasi-likelihood estimator. More formally, let $\eta(x, \widehat{\theta})$ be a "naive" quasi-likelihood estimator of $\eta(x, \theta)$, a given, possibly misspecified, parametric model for $\eta(x)$. [16] proposed to estimate the error term $r_{\widehat{\theta}}(x) := \eta(x) - \eta(x, \widehat{\theta})$ using a nonparametric weighted local quasi-likelihood (LQL) estimator that we denote by $\widehat{r}_{\widehat{\theta}}(x)$. The additive parametrically guided local quasi-likelihood (GLQL) estimator is defined by $\widehat{\eta}(x) = \eta(x, \widehat{\theta}) + \widehat{r}_{\widehat{\theta}}(x)$. When the parametric model is properly chosen, $r_{\widehat{\theta}}$ may be flatter and easier to adjust non-parametrically than the original function η. In this case, the guided local quasi-likelihood estimator should be of smaller MSE than the classical LQL estimator. Otherwise, the nonparametric correction is expected to correct for the misspecification and there should not be much loss in accuracy for the resulting GLQL estimator compared to the classical LQL estimator.

Regression problems in which the response is subject to censoring have been widely studied in the literature. Many investigations have been devoted to parametric regression, among them, [3], [25], [26] and [8]. An extensive field of research has been developed for nonparametric regression, see for example [2], [12], [21], [10] and [28], among others. However, only few papers extending parametric quasi-likelihood to censored data exist in the literature. The first extension of quasi-likelihood to the right censored data case has been established in the framework of partially linear single-index models by [29]. In the generalized linear model, [44] adapted the parametric quasi-likelihood to censored data. Recently, [45], [47] and [46] proposed different semi-parametric quasi-likelihood estimators in the framework of accelerated failure time models. Note that, none of the papers mentioned above has considered a fully nonparametric quasi-likelihood. Thus, one of the main objectives of this paper is to extend the local quasi-likelihood of [14] to the censored data case.

Regarding the parametrically guided nonparametric estimation, as far as we know, except the recent work of [39], [40], the guided nonparametric estimation has never been studied in the context of censored data. A well known challenge in the presence of censoring is that the response is not always available. Consequently, the parametrically guided local quasi-likelihood method can not be directly applied. In order to address this problem, we first need to transform the data before applying the GLQL. Several transformations have been proposed in the literature. In this work we investigate the transformation proposed by [25] since it does not require any iterative procedure.

The paper is organized as follows. Section 2 explains in detail the different steps of the proposed methodology. Section 3 provides some asymptotic results for the proposed method, while Section 4 illustrates the performance of the proposed estimator via simulation studies. Finally, some general conclusions are drawn in Section 5. The proofs are given in the Appendix.

2. Model and methodology

Regression techniques are commonly used to describe a relationship between a variable of interest $Y \in \Re$ and a covariate $X \in \Re$. In a right censored regression framework, the response Y is not directly available. Indeed, in the presence of a censoring variable C one can only observe an i.i.d. random sample $(X_i, T_i, \delta_i), i = 1, \dots, n$, from (X, T, δ), where $T = \min(Y, C)$ and $\delta = I(Y \le C)$. In the following we suppose that given the covariate X, the censoring variable C is independent of the variable of interest Y. Set $F(y|x) = P(Y \le y|X = x)$ and $G(y|x) = P(C \le y|X = x)$ the conditional distribution function of Y and C given $X = x$, respectively. Suppose that there exists a known positive function $V(\cdot)$ that relates the conditional mean and the conditional variance of $\phi(Y)$ given X as follows:

$$m(x) = E(\phi(Y)|X = x) \text{ and } Var(\phi(Y)|X = x) = V(m(x)),$$

where ϕ is a known function used to cover various parameters of interest. For example, when $\phi(y) = y\mathbf{1}_{\{y \le \tau\}}$, for some known τ, we get the truncated mean $m(x) = \int_{-\infty}^{\tau} y dF(y|x)$. Our main objective is to estimate $\eta(x) = g(m(x))$, where $g(\cdot)$ is a known link function. Since only the relationship between the conditional mean and the conditional variance is known, the likelihood estimation method can not be used. In the following, we first introduce the guided local quasi-likelihood for complete data, and then we adapt the method to handle censoring.

2.1. Guided local quasi-likelihood for complete data

[43] defined the quasi-log-likelihood function as any function $Q(\mu, y)$ satisfying

$$\frac{\partial}{\partial \mu} Q(\mu, y) = \frac{y - \mu}{V(\mu)}.$$

Assuming that $\eta(x) = \theta_0 + \theta_1 x$, the parameters θ_0 and θ_1 can be estimated via maximizing the parametric quasi-likelihood $\sum_{i=1}^{n} Q\big(g^{-1}(\theta_0 + \theta_1 X_i), Y_i\big)$, that plays the role of the log-likelihood in the classical GLM model. Because the assumption of linearity does not hold in many situations, [14] proposed a local polynomial quasi-likelihood method to estimate $\eta(\cdot)$ without assuming any specific form for it. The maximum local quasi-likelihood estimator of order p ($p \geq 1$) of $\eta(x)$ is $\widehat{\eta}(x) = \widehat{\beta}_0$, where $(\widehat{\beta}_0, \ldots, \widehat{\beta}_p)$ is the maximizer of

$$\sum_{i=1}^{n} Q\big(g^{-1}(\beta_0 + \ldots + \beta_p(X_i - x)^p), Y_i\big) K_h\big(X_i - x\big),$$

where $K(\cdot)$ is a kernel density, $h \equiv h_n$ is a smoothing bandwidth, and $K_h(\cdot) = h^{-1}K(\cdot/h)$.

Let $\eta(x, \theta)$ be a parametric model which belongs to some family of parameterized functions $\{\eta(x, \theta) : \theta \in \Theta \subset \Re^d\}$ and define the parametric maximum quasi-likelihood estimator of θ as

$$\widehat{\theta} = \arg\max_{\theta \in \Theta} \sum_{i=1}^{n} Q\big(g^{-1}(\eta(X_i, \theta)), Y_i\big). \tag{2.1}$$

As discussed in the introduction, the parametric estimator given in (2.1) provides some useful and interesting information about the target function $\eta(\cdot)$ that may help us to improve the local quasi-likelihood estimator. To simplify the presentation, we focus on the local linear case ($p = 1$). Under the additive scheme, [16] proposed to estimate the error term $r_{\widehat{\theta}}(x) = \eta(x) - \eta(x, \widehat{\theta})$ by $\widehat{r}_{\widehat{\theta}}(x) = \widehat{\beta}_0$, where $(\widehat{\beta}_0, \widehat{\beta}_1)$ is the maximizer, with respect to (β_0, β_1), of the local quasi-likelihood function $\sum_{i=1}^{n} Q\big(g^{-1}(\beta_0 + \beta_1(X_i - x) + \eta(X_i, \widehat{\theta})), Y_i\big) K_h\big(X_i - x\big)$. $\eta(x)$ can then be estimated by $\widehat{\eta}(x) = \eta(x, \widehat{\theta}) + \widehat{r}_{\widehat{\theta}}(x)$. The latter can, equivalently and directly, be derived by maximizing

$$\sum_{i=1}^{n} Q\big(g^{-1}(\beta_0 + \beta_1(X_i - x) + \eta(X_i, \widehat{\theta}) - \eta(x, \widehat{\theta})), Y_i\big) K_h\big(X_i - x\big),$$

and taking $\widehat{\eta}(x) = \widehat{\beta}_0$, where $(\widehat{\beta}_0, \widehat{\beta}_1)$ maximizes the above function with respect to (β_0, β_1).

2.2. Guided local quasi-likelihood and censoring

In the presence of censoring, as $E(\phi(T)|X = x) \neq m(x)$, one cannot directly use the observed data to estimate $\eta(x) = g(m(x))$. In order to overcome this problem, we will use the synthetic data approach. In this approach, the observed response T is substituted by a synthetic response Y^*, such that, under the conditional independence of Y and C given X, $E(Y^*|X = x) = m(x)$. Different transformations satisfying this equality exist in the literature, see for instance [27] and [48], among others. We limit ourselves to the transformation of [25] defined by

$$Y^* = \frac{\delta\phi(T)}{1 - G(T^-|X)}. \tag{2.2}$$

This transformation is not directly applicable in practice, since it depends on

$G(y|x)$, the conditional distribution of C given $X = x$, which is unknown. An estimator of this function was proposed by [2] and is given by

$$\hat{G}(y|x) = 1 - \prod_{i=1}^{n} \left(1 - \frac{(1-\delta_i)\mathbf{1}_{\{T_i \le y\}} w_i(x)}{\sum_{j=1}^{n} \mathbf{1}_{\{T_i \le T_j\}} w_i(x)}\right),$$

where $w_i(x) = K_0((X_i - x)/h_0)/\sum_{j=1}^{n} K_0((X_j - x)/h_0)$, are the Nadaraya-Watson weights K_0 is a kernel density function and h_0 is a bandwidth parameter. Note that if $w_i(x) = n^{-1}, i = 1, \ldots, n$, then $\hat{G}$ reduces to the well known Kaplan-Meier estimator. Beran's estimator was studied by many authors, among them we cite [9], [6], [19] and [42]. We define the synthetic response $\hat{Y}^*$ by plugging Beran's estimator into the transformation (2.2) as follows:

$$\hat{Y}^* = \frac{\delta\phi(T)}{1 - \hat{G}(T^-|X)}. \tag{2.3}$$

Following [16], we define our parametrically guided local quasi-likelihood estimator of η, based on the synthetic sample $(\hat{Y}_i^*, X_i),\ i = 1, \ldots, n$, as $\widehat{\eta}_{\hat{G},\widehat{\theta}}(x) = \widehat{\beta}_0$, where $\widehat{\beta} = (\widehat{\beta}_0, \widehat{\beta}_1)^T$ is the maximizer of

$$\sum_{i=1}^{n} Q\big(g^{-1}(\beta_0 + \beta_1(X_i - x) + \eta(X_i, \widehat{\theta}) - \eta(x, \widehat{\theta})), \hat{Y}_i^*\big) K_h\big(X_i - x\big), \tag{2.4}$$

with respect to $\beta = (\beta_0, \beta_1)^T$, and $\widehat{\theta}$ is a pseudo parametric quasi-likelihood estimator of θ adapted to censored data. The estimation approach that we adopt will be discussed in detail in Section 3.2. Note that the parametrically guided local quasi-likelihood given in (2.4) raises new challenges when compared to the equivalent estimator with completely observed data since the synthetic observations $\hat{Y}_i^*, i = 1, \ldots, n$ defined by (2.3) are estimated using the whole sample.

Remark 2.1. *We didn't find any results in the literature concerning the estimation of a general misspecified parametric model using quasi-likelihood under censoring. We also note that using a linear guide reduces the estimator to the classical local quasi-likelihood estimator of η, which means that our GLQL estimator $\widehat{\eta}_{\hat{G},\widehat{\theta}}(x)$ is a generalization of the classical LQL estimator that can be obtained by maximizing (2.4) with $\eta = 0$.*

3. Theoretical properties

In order to show the bias reduction property of our new estimator, we investigate in this section the asymptotic distribution of $\widehat{\eta}_{\hat{G},\widehat{\theta}}(\cdot)$. First of all, we derive in Theorem 3.1 the asymptotic properties of $\widetilde{\eta}_{\hat{G}}(\cdot)$ an estimator of $\eta(\cdot)$ guided by a given non stochastic approximation $\widetilde{\eta}(\cdot)$. Then, in Theorem 3.2 we generalize the results to cover the case of a data-driven parametric guide.

3.1. The model with non-random guide

Let $\widetilde{\eta}(x)$ be a non stochastic guide that approximates the true function $\eta(x)$ and let $\widetilde{\beta} = (\widetilde{\beta}_0, \widetilde{\beta}_1)$ maximize the following function:

$$\sum_{i=1}^{n} Q\big(g^{-1}(\beta_0 + \beta_1(X_i - x) + \widetilde{\eta}(X_i) - \widetilde{\eta}(x)), \hat{Y}_i^*\big) K_h\big(X_i - x\big). \tag{3.1}$$

Define the corresponding GLQL estimator as $\widetilde{\eta}_{\hat{G}}(x) = \widetilde{\beta}_0$. In the following, we provide the assumptions required for the main results.

Assumption 3.1. *A1. i. X has a compact support $S_X \subset \Re$.*
ii. $f_X(.)$, the marginal density of X, is twice continuously differentiable and $\inf_{x\in S_X} f_X(x) > 0$.
A2. The function ϕ is bounded and vanishes outside the interval $[0, \tau]$ for some $\tau < \inf_{x\in S_X} \tau_x$ with $\tau_x = \sup\{y : H(y|x) < 1\}$, that is, the right endpoint of the support of $H(y|x) = P(T \le y|X = x)$.
A3. The functions $H_j(y|x) = P(T \le y, \delta = j|X = x)$, $j = 0, 1$, have four derivatives with respect to x. Furthermore, the derivatives are bounded uniformly for all $y \le \tau$ and $x \in S_X$.
A4. $E(\phi(Y)^2) < \infty$.
A5. i. K is a symmetric probability density function with compact support, say $S_K = [-1, 1]$.
ii. K_0 is a symmetric, twice continuously differentiable probability density function with compact support S_{K_0}.
iii. $\int x^2K(x)dx = \mu_2^K < \infty$, $\int x^2K_0(x)dx = \mu_2^{K_0} < \infty$ and $\int x^jK^2(x)dx = \nu_j^K < \infty$ for $j = 0, 1, 2$.
A6. $nh^5 = O(1)$, $nh_0^5/\log n = O(1)$, $nh \to \infty$ and $nh_0 \to \infty$ as $n \to \infty$.
A7. $\eta(\cdot)$, $V(\cdot)$, $\widetilde{\eta}(\cdot)$ and $g'(\cdot)$ are twice continuously differentiable.

Assumptions A1 and A7 are regularity assumptions needed for the consistency and the asymptotic normality of the guided estimator. Assumptions A2 and A3 are usual assumptions in nonparametric regression with censored data allowing to avoid the problem of inconsistency of Beran's estimator on the right tail of the distribution. Assumption A4 insures a finite variance for the guided estimator. Finally, assumption A5 and A6 concerns the supports of the kernels and the sequence of bandwidths, respectively. The supports are supposed to be compact to control the bias and the rate of convergence of both Beran's estimator and the guided estimator.

Let $q_l(x, y) = \frac{\partial^l}{\partial x^l} Q(g^{-1}(x), y)$ and $\rho_l(x) = \big(g'(g^{-1}(x))^l V(g^{-1}(x))\big)^{-1}$, $l = 1, 2$. Note that q_l is linear in y for a fixed x, $q_1(\eta(x), m(x)) = 0$ and $q_2(\eta(x), m(x)) = -\rho_2(\eta(x))$. The following additional assumptions are also required.

Assumption 3.2. *B1. The function $q_2(x, y) < 0$ for all $x \in S_X$ and $y \le \tau$.*
B2. The function $\sigma_*^2(x) = Var(Y^*|X = x)$ is continuous on S_X.
B3. For all $x \in S_X$, $\rho_2(x) \ne 0$, $\sigma_*^2(x) \ne 0$ and $g'(m(x)) \ne 0$.

Assumption B1 implies the concavity of the quasi-likelihood function (expression 3.5) on β and so the uniqueness of the guided maximum local quasi-likelihood. Assumptions B2 and B3 are needed to ensure a bounded and non-zero asymptotic variance for the guided estimator. These assumption are similar to the assumptions in [14] and [16] for uncensored case. The following Theorem provides the asymptotic distribution of $\widetilde{\eta}_{\hat{G}}(\cdot)$.

Theorem 3.1. *Suppose that Assumptions 3.1 and 3.2 hold. Then,*

$$(nh)^{1/2}\left\{\widetilde{\eta}_{\hat{G}}(x)-\eta(x)-\widetilde{B}(x)+O_p\left(\left(\frac{\log n}{nh_0}\right)^{1/2}\right)\right\} \xrightarrow{d} \mathcal{N}\left(0,\ \sigma_*^2(x)g'(m(x))^2 f_X^{-1}(x)\nu_0^K\right),$$

with

$$\widetilde{B}(x)=\frac{1}{2}h^2\mu_2^K\big(\eta''(x)-\widetilde{\eta}''(x)\big)\{1+o(1)\}.$$

Remark 3.1. *The bias produced by Beran's estimator is bounded by $\big(\frac{\log n}{nh_0}\big)^{1/2}$. This extra term vanishes when the bandwidths are chosen such that $\frac{h_0}{h\log n}\to\infty$. Therefore, there is no loss of accuracy when one replaces the response by synthetic data, provided that the bandwidth for Beran's estimator is asymptotically larger than the bandwidth used in the local linear fit. This fact has also been pointed out by [39] in the context of guided nonparametric regression with censored data. The bias term $\widetilde{B}(x)$ is similar to the fully observed data case and reveals the effect of the parametric guide. If the guide is chosen such that $|\eta''(x)-\widetilde{\eta}''(x)|\le|\eta''(x)|$, then the bias of the GLQL estimator will be smaller compared with that of the classical LQL estimator. If the second derivatives of the parametric guide and the true function are equal, then the bias term $\widetilde{B}(x)$ vanishes. Regarding the variance, there is no difference compared with the classical LQL under censorship. The only difference appears when one compares the variance term of the GLQL estimator in the presence and the absence of censoring. In fact, the term $\sigma_*^2(x)=\sigma^2(x)+E\big[\phi(Y)^2G(Y^-|X)/(1-G(Y^-|X))|X=x\big]$ replaces $\sigma^2(x)=Var(\phi(Y)|X=x)$ and this is due to the synthetic data. Note that, if the parametric guide is chosen to be constant, then the GLQL estimator reduces to the classical LQL estimator. Therefore, the result of our Theorem 3.1 is a generalization of Theorem 1.a (for $p=1,r=0$) in [14] to right censored data. Finally, we note that Theorem 1 in [39] is a special case of Theorem 3.1 using an identity link function $g=I$ and a constant variance function V.*

3.2. The model with an estimated guide

In the previous section, Theorem 3.1 investigated the simple case of a fixed guide. However, in practice, the guide needs to be estimated. In the following, we consider the case where the parametric guide $\eta(x,\widehat{\theta})$ is obtained from a first stage estimation procedure. Following [16], we denote by

$$f(x,y)=f_X(x)\exp(Q(g^{-1}(\eta(x)),y)),$$

the true unknown joint density of (X, Y) and by

$$f(x, y; \theta) = f_X(x) \exp(Q(g^{-1}(\eta(x, \theta)), y)),$$

the proposed parametric joint density. Define $\theta_* \in \Theta \subset \Re^d$, the value of θ which maximizes the following function:

$$\iint_\Delta Q(g^{-1}(\eta(x, \theta)), y) dF(x, y), \tag{3.2}$$

where $F(x, y)$ is the joint distribution function of (X, Y) and $\Delta = S_X \times (-\infty, \tau]$ is needed because the right tail of the distribution $F(x, y)$ cannot be estimated consistently when the response Y is censored. θ_* is the parameter value that minimizes the Kullback-Leibler distance between the true joint density $f(x, y)$ and the parametric joint density $f(x, y; \theta)$, that is, $\theta_* = \arg\min_{\theta \in \Theta} E_\Delta\big[\log\big(f(X, Y)/f(X, Y; \theta)\big)\big]$ with $E_\Delta(A) = E(A.\mathbb{1}_{\{(X,Y)\in\Delta\}})$. If the parametric model is correct, i.e. there exists $\theta_0 \in \Theta$ such that $f(x, y) = f(x, y; \theta_0)$, then $\theta_0 = \theta_*$.

In the spirit of [38], we estimate θ by $\widehat{\theta}$, the maximizer of a suitable analogue of (2.1) that we define as

$$\iint_\Delta Q(g^{-1}(\eta(x, \theta)), y) d\widehat{F}(x, y), \tag{3.3}$$

where $\widehat{F}$ is an estimator of F satisfying the following assumptions:

Assumption 3.3. *1.*

$$\sup_{\theta \in \Theta} \left| \iint_\Delta \nabla_\theta^r \log f(x, y; \theta) d(\widehat{F} - F)(x, y) \right| = o_p(1), \quad \textit{for } r = 0, 2.$$

2.

$$\sqrt{n} \iint_\Delta \nabla_\theta^1 \log f(x, y; \theta_*) d(\widehat{F} - F)(x, y) \xrightarrow{d} \mathcal{N}(0, \Sigma),$$

where Σ is a nonnegative-definite matrix and $\nabla_\theta^r \Phi(x, y; \theta) = \partial^r \Phi(x, y; \theta)/\partial \theta^r$ for a twice differentiable function $\theta \to \Phi(x, y; \theta)$ and $r = 0, 1, 2$.

Note that, the first assumption is the uniform convergence condition (in probability) required for the proof of the first point of Proposition 3.1. The second condition is needed for verifying asymptotic normality of the parametric estimator.

When the data are completely observed, the estimator $\widehat{F}$ may be replaced by the usual bivariate empirical distribution function $F_n(x, y) = \frac{1}{n}\sum_{i=1}^n 1_{X_i \le x, Y_i \le y}$. In this case, the pseudo quasi-likelihood defined by (3.3) reduces to (2.1), meaning that our approach is more general. In the censored data framework, there have been few proposals for estimating $F(x, y)$ in the literature. For example, [28] has developed an estimator of $F(x, y)$ satisfying Assumption 3.3 (see Theorem 3.1 and Theorem 3.6 in [28]) and given by the following expression

$$\widehat{F}_L(x, y) = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i 1_{X_i \le x, T_i \le y}}{1 - \hat{G}(T_i^- | X_i)}. \tag{3.4}$$

Another different and interesting approach was introduced by [41]. Their estimator is constructed through an integrated version of Beran's estimator as follows

$$\widehat{F}_{\mathrm{VA}}(x,y) = \int_{-\infty}^{x} \hat{F}(y|u)dF_n(u), \tag{3.5}$$

where $F_n(x)$ is the empirical distribution function of X and $\hat{F}(y|x)$ is Beran's estimator of $F(y|x) = P(Y \leq y|X = x)$. We note that both estimators can be used in practice. However, to the best of our knowledge, Assumption 3.3 has not yet been investigated for $\widehat{F}_{\mathrm{VA}}$. Therefore, for sake of consistency, we only investigate the estimator of [28] in our simulation studies. Next, we give additional conditions that are also needed.

Assumption 3.4. *D1. $\eta(x,\theta)$ belongs to a parametrically indexed class of functions defined by the following characteristics:*
1. $\theta \in \Theta$, Θ is a compact subset of $\Re^d$.
2. The function $(x,\theta) \mapsto \eta(x,\theta)$ is twice continuously differentiable with respect to x and θ.

D2. The function $\log f(x,y;\theta)$ is twice continuously differentiable with respect to θ.

D3. $E_\Delta\big(\log f(X,Y)\big)$ exists and there exists a function $\ell(x,y)$ such that $|\log f(x,y;\theta)| \leq \ell(x,y)$ for any $\theta \in \Theta$ and $E_\Delta \ell(X,Y) < \infty$.

D4. $|\partial^2 \log f(x,y;\theta)/\partial\theta_i\theta_j|$ and $|\partial \log f(x,y;\theta)/\partial\theta_i \times \partial \log f(x,y;\theta)/\partial\theta_j|$, for $i,j = 1,\dots,d$, are dominated by integrable functions with respect to $F(.,.)$ for all (x,y) in Δ and all θ in Θ.

D5. $\theta_* = \arg\min_{\theta\in\Theta} E_\Delta\big(\log(f(X,Y)/f(X,Y;\theta))\big)$ is unique.

D6. The matrix of second derivatives $\nabla^2_\theta \log f(x,y;\theta_*)$ is nonsingular.

Conditions D1.1, D2 and D5 are respectively, the compactness of the parameter set, the continuity condition, and the condition for the limiting objective function to have a unique maximum. These three conditions are needed for the consistency of the parametric guide. Conditions D2, D3 and D4 are classical conditions in the uncensored case that allow to take derivatives under the integrals. The following proposition provides the weak consistency and the asymptotic normality of the estimator $\widehat{\theta}$.

Proposition 3.1. *Under Assumptions 3.3 and 3.4, we have*

1. *$\widehat{\theta}$ converges to θ_* in probability as $n \to \infty$.*
2. *$\sqrt{n}(\widehat{\theta} - \theta_*) \xrightarrow{d} \mathcal{N}(0, \Omega^{-1}\Sigma\Omega^{-1})$, with $\Omega \equiv \Omega(\theta_*)$ and $\Omega(\theta) = E_\Delta\big[\nabla^2_\theta \log f(X,Y;\theta)\big]$.*

Note that, the results of Proposition 3.1 reveal the $\sqrt{n}$-consistency of the estimator $\widehat{\theta}$, that is $\sqrt{n}(\widehat{\theta} - \theta_*) = O_p(1)$. Now, given this result and some additional conditions, the next Theorem states that there is no loss in accuracy when the parametric guide is estimated.

Theorem 3.2. *Suppose that Assumptions 3.3 and 3.4 hold. Then, under assumptions of Theorem 3.1, we have*

$$(nh)^{1/2}\left\{\widehat{\eta}_{\hat{G},\widehat{\theta}}(x)-\eta(x)-B(x,\theta_*)+O_p\left(\left(\frac{\log n}{nh_0}\right)^{1/2}\right)\right\}$$
$$\xrightarrow{d}\mathcal{N}\left(0,\ \sigma_*^2(x)g'(m(x))^2 f_X^{-1}(x)\nu_0^K\right),$$

with

$$B(x,\theta_*)=\frac{1}{2}h^2\mu_2^K\big(\eta''(x)-\eta''(x,\theta_*)\big)\{1+o(1)\}.$$

Comparing this last result with the result of Theorem 3.1, we notice that the estimation of the parameter θ_* does not affect the asymptotic bias and the asymptotic variance. A crucial issue that arises in any nonparametric method is the choice of the bandwidth parameters. From Theorem 3.2, the asymptotic mean integrated squared error is given by

$$\frac{1}{4}h^4(\mu_2^K)^2\int_{S_X}\big(\eta''(x)-\eta''(x,\theta_*)\big)^2dx+\frac{\nu_0^K}{nh}\int_{S_X}\sigma_*^2(x)g'(m(x))^2f_X^{-1}(x)dx. \quad (3.6)$$

If $\eta''(x)-\eta''(x,\theta_*)=0$, then $B(x,\theta_*)=0$. In such a case, one can choose an arbitrary large bandwidth so that the variance is reduced to its minimum possible value, which is impossible in a fully nonparametric framework (except for a linear η). If $\eta''(x)-\eta''(x,\theta_*)\neq 0$, then minimizing (3.6) with respect to h gives the following theoretical optimal bandwidth:

$$h_{opt}=\left(\frac{\nu_0^K\int_{S_X}\sigma_*^2(x)g'(m(x))^2f_X^{-1}(x)dx}{(\mu_2^K)^2\int_{S_X}(\eta''(x)-\eta''(x,\theta_*))^2dx}\right)^{1/5}n^{-1/5}. \quad (3.7)$$

This last expression indicates that, if the parametric guide is chosen so that its second derivatives $\eta''(x,\theta_*)$ is close to the second derivative of the true function $\eta''(x)$, then the optimal bandwidth for the GLQL estimator will be larger than the optimal bandwidth of the classical LQL estimator. This allows to reduce also the variance compared with the classical LQL estimator. This fact is widely noticed in our simulation studies. In practice, expression (3.7) cannot be used directly since it depends on a number of unknown quantities. [13] (see Section 4.9) and [14] proposed some guidelines for the selection of the bandwidth h based on the bias-variance tradeoff. Their procedures can be easily extended to censored data framework by simply substituting the censored response Y_i by the synthetic data $\hat{Y}_i^*$. Finally, the bandwidth for Beran's estimator can be chosen using for example the plug-in method (see [6]) or the bootstrap method investigated by [42].

3.3. Extension to unified family of corrections

As mentioned in the introduction, we investigate the additive correction in order to simplify our presentation. However, in addition to the additive scheme,

[30] proposed a unified family of corrections in the uncensored data case. In the following we give some guidelines allowing to generalize their proposal to our framework. Starting from a parametric model $\eta(x,\widehat{\theta})$, the basic idea of the guided estimation can be generalized using the following more general identity:

$$\eta(x) = \eta(x,\widehat{\theta}) + r_{\widehat{\theta},\alpha}(x)\eta(x,\widehat{\theta})^{\alpha},$$

where $r_{\widehat{\theta},\alpha}(x) = [\eta(x) - \eta(x,\widehat{\theta})]/\eta(x,\widehat{\theta})^{\alpha}$ and $\alpha \geq 0$. We propose to estimate the correction factor $r_{\widehat{\theta},\alpha}(x)$ by $\widehat{r}_{\widehat{\theta},\alpha}(x) = \widehat{\beta_0}$, where $(\widehat{\beta_0}, \widehat{\beta}_1)$ is the maximizer of

$$\sum_{i=1}^{n} Q\big(g^{-1}(\eta(X_i,\widehat{\theta}) + (\beta_0 + (X_i - x)\beta_1)\eta(X_i,\widehat{\theta})^{\alpha}), \hat{Y}_i^{*}\big)K_h(X_i - x).$$

Therefore, the extended guided local quasi-likelihood estimator is given by $\widehat{\eta}_{\hat{G},\widehat{\theta},\alpha}(x) = \eta(x,\widehat{\theta}) + \widehat{r}_{\widehat{\theta},\alpha}(x)\eta(x,\widehat{\theta})^{\alpha}$. Similarly as in Section 2.1, the extended guided estimator $\widehat{\eta}_{\hat{G},\widehat{\theta},\alpha}(x)$ can be defined directly as the first component of the maximizer of

$$\sum_{i=1}^{n} Q\big(g^{-1}(\eta(X_i,\widehat{\theta}) + (\beta_0 + (X_i - x)\beta_1 - \eta(x,\widehat{\theta}))\eta(X_i,\widehat{\theta})^{\alpha}/\eta(x,\widehat{\theta})^{\alpha}), \hat{Y}_i^{*}\big)K_h(X_i - x)$$

with respect to $\beta = (\beta_0, \beta_1)$. All the results established before can be generalized to the guided estimator based on the unified family of corrections, the generalization of the proof is straightforward and is omitted here. Theorem 3.3 generalizes the result of Theorem 3.2.

Theorem 3.3. *Suppose that Assumption 3.3 and 3.4 hold and $\eta(x,\theta_*) \neq 0$. Then, under the Assumptions of Theorem 3.1, we have*

$$(nh)^{1/2}\left\{\widehat{\eta}_{\hat{G},\widehat{\theta},\alpha}(x) - \eta(x) - B(x,\theta_*,\alpha) + O_p\left(\left(\frac{\log n}{nh_0}\right)^{1/2}\right)\right\} \xrightarrow{d} \mathcal{N}\left(0,\ \sigma_*^2(x)(g'(m(x))^2 f_X^{-1}(x)\nu_0^K\right),$$

with $B(x,\theta_*,\alpha) = \frac{1}{2}h^2\mu_2^K\eta(x,\theta_*)^{\alpha}r''_{\theta_*,\alpha}(x)\{1 + o(1)\}$ and $r_{\theta_*,\alpha}(x) = [\eta(x) - \eta(x,\theta_*)]/\eta(x,\theta_*)^{\alpha}$.

Note that, the additive correction is a special case of the unified family for $\alpha = 0$. The choice of the parameter α was investigated by [16]. However, using the best α does not enhance the performance considerably compared with the additive correction. Therefore, to simplify our simulation studies we investigate the additive correction.

4. Simulation results

This section is concerned with the evaluation of the finite sample performance of the GLQL estimator. To this end, we conduct two Monte Carlo simulation studies. In the first study, a Poisson model is investigated under right censoring. Such model is widely used in studies dealing with quasi-likelihood and discrete responses, see for example [16] and [7]. Then, in a second time, an exponential model is considered to cover the case of continuous responses. Our target function is $\eta(x) = g\big(\int_0^\tau y dF(y|x)\big)$ where g is the canonical link, $\tau = \inf_x\{\tau_x\}$ and τ_x is the 99.99% upper quantile of $H(y|x) = P(T \leq y|X = x)$. The parametric guides are estimated via maximizing the pseudo QL given in (3.3) combined with the estimator (3.4) proposed by [28]. Along the simulations we use local linear fitting and the Epanechnikov kernel for both K_0 and K. To reduce our calculation time, we first selected the value of the bandwidths h_0 and h by minimizing the average mean squared error (MSE) using a small number of simulations. Then, we applied both guided and traditional LQL to 1000 other simulated data sets using the selected "optimal" bandwidths for each method.

4.1. Poisson model

In this model, the covariate X is generated from a uniform distribution over the interval $[-2, 2]$. On the other hand, given $X = x$, the response Y is generated from a Poisson distribution with mean $\exp(\Lambda(x))$ and $\Lambda(x) = 6+3\sin(\frac{\pi}{4}x - \frac{\pi}{2})$. Given $X = x$ and independently from Y, the censoring variable C is also drawn from a Poisson distribution with mean $\exp(\Lambda(x)+\lambda)$. The parameter λ allows us to control the rate of censoring. The values $\lambda = 0.22, 0.135$, and 0.078 correspond to a fixed censoring rate of 10%, 20% and 30%, respectively. Following [16], three different parametric guides are investigated. The first two guides are misspecified and are given by $\eta_1(x, \theta) = \theta_0+\theta_1x+\theta_2x^2$ and $\eta_2(x, \theta) = \theta_0+\theta_1x+\theta_2x^2+\theta_3x^3$, respectively. The third parametric guide is correctly specified and is given by the following sinusoidal function $\eta_3(x, \theta) = \theta_0+\theta_1\sin(\frac{\pi}{4}x-\frac{\pi}{2})$. As a quasi-likelihood function we used $Q(\mu, y) = y\log|\mu|-\mu$. We investigate the performance of both the GLQL estimator and the LQL estimator at ten equidistant data points in the interval $[-2, 2]$ using three sample sizes $n = 100, 250$ and 500.

As stated before, to select the bandwidths we repeat the simulation 200 times. Figure 1 shows how the squared bias, the variance and the MSE change with the bandwidth h, for sample size $n = 250$ and a censoring rate of 20%. As established in the asymptotic results, the bias is substantially reduced for the three guided estimators compared with the unguided estimator, while the variance remains unchanged or is slightly reduced especially when a large bandwidth is used. We also note that when the appropriate guide (sinusoidal) is used, the bias of the GLQL estimator is almost zero. This allows us to choose a larger bandwidth and so to reduce the variance substantially.

Now, using the selected bandwidths, we compute the different estimators 1000 times. The average squared bias ($Bias^2 \times 10^3$), the average variance ($Var \times 10^3$), the average mean squared error AMSE ($\times 10^3$) as well as the selected bandwidths are given in Table 1 for each setting. Generally speaking, the results show that the GLQL estimators have lower MSE compared to the classical LQL even if the parametric guide is not completely correct. As expected, the

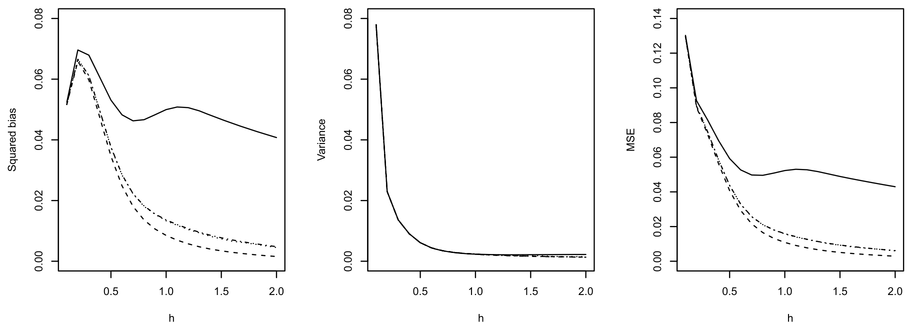

Fig 1. *Average squared bias, average variance and average MSE as a function of h, for the LQL estimator (solid curve), GLQL1 estimator (dot-dashed curve), GLQL2 (dotted curve), GLQL3 (dashed curve). The sample size is* $n = 250$, *the proportion of censoring is* $p = 0.20$.

best results are obtained when the guide is correctly specified, namely with the sinusoidal guide. Overall, we can say that the GLQL estimator considerably outperforms the classical LQL estimator. As expected, increasing the sample size improves the quality of all the estimators, in terms of AMSE, but increasing the censoring rate affects negatively the results. A comparison between the censored and uncensored ($p = 0\%$) data cases shows that the AMSE of the GLQL is less affected by the presence of censoring compared to that of the LQL. Therefore, using a parametric guide allows to reduce the negative effect of the censoring on the efficiency of the local quasi-likelihood estimator. As expected, the selected bandwidths under censoring are larger compared with those selected with fully observed data.

Finally, we investigated the selection of the parameter α for the generalized guided local quasi-likelihood estimator. Simulations not given here show that using the optimal α (which minimizes the AMSE) does not enhance the performance considerably compared with the additive correction. Moreover, choosing an additional parameter is highly time-consuming under censoring. Therefore, we would recommend to use the additive correction which is less time-consuming and significantly improves the efficiency of the local quasi-likelihood estimator.

4.2. Exponential model

This section addresses the case of a continuous response. Given $X = x$, the response Y is generated from an exponential distribution with parameter $\Lambda(x) = (0.5x^2+1)+a(\sin(2\pi x))^2$, where $a = 0, 0.1, 0.3, 0.5$, while the covariate X is uniformly distributed on $[0, 4]$. The censoring variable C is independent of Y given $X = x$ and is also generated from an exponential distribution with parameter $\Lambda(x)/2$ which leads to almost 33.4% rate of censoring. Regarding the parametric guide, we consider a second order polynomial guide $\eta(x, \theta) = \theta_0 + \theta_1 x + \theta_2 x^2$. The parameter a allows to control the difference between the true function and the parametric guide. Figure 2 gives the shapes of different target functions.

Table 1
Average squared bias ($\times 10^3$), average variance ($\times 10^3$), average MSE ($\times 10^3$), the optimal bandwidth h, four censoring rates $p = (0\%, 10\%, 20\%, 30\%)$, three sample sizes $n = (100, 250, 500)$ and $N = 1000$ replications.

p	0%				10%				20%				30%			
Method	Bias^2	Var	MSE	h	Bias^2	Var	MSE	h	Bias^2	Var	MSE	h	Bias^2	Var	MSE	h
							n= 100									
GLQL1	0.391	1.296	1.687	0.8	1.454	2.578	4.032	2	4.005	3.679	7.684	2	6.630	4.930	11.56	2
GLQL2	0.361	1.363	1.724	0.8	1.387	2.626	4.013	2	3.738	3.835	7.573	2	6.333	5.034	11.37	2
GLQL3	0.011	1.995	2.006	0.7	0.099	2.604	2.703	2	1.224	3.632	4.856	2	3.011	4.770	7.781	2
LQL	0.703	3.027	3.730	0.6	30.36	4.740	35.10	2	37.23	6.660	43.89	2	40.90	8.650	49.55	2
							n= 250									
GLQL1	0.158	0.715	0.873	0.6	1.688	1.120	2.807	2	4.662	1.493	6.155	2	7.953	1.887	9.840	2
GLQL2	0.151	0.731	0.882	0.6	1.582	1.281	2.863	2	4.425	1.712	6.137	2	7.589	2.017	9.606	2
GLQL3	9×10^{-5}	0.832	0.832	0.6	0.127	1.136	1.263	2	1.463	1.486	2.949	2	3.616	1.843	5.459	2
LQL	4.394	0.756	5.150	0.3	32.42	1.900	34.32	2	40.76	2.530	43.29	2	46.53	3.220	49.75	2
							n= 500									
GLQL1	0.142	0.335	0.477	0.6	1.795	0.488	2.283	2	4.640	0.700	5.340	2	8.177	0.876	9.053	2
GLQL2	0.066	0.420	0.486	0.5	1.711	0.596	2.307	2	4.471	0.803	5.274	2	7.824	1.038	8.862	2
GLQL3	10^{-4}	0.400	0.400	0.6	0.138	0.503	0.641	2	1.399	0.692	2.091	2	3.696	0.853	4.549	2
LQL	0.264	0.729	0.993	0.3	33.62	0.850	34.47	2	41.55	1.220	42.77	2	47.55	1.520	49.07	2

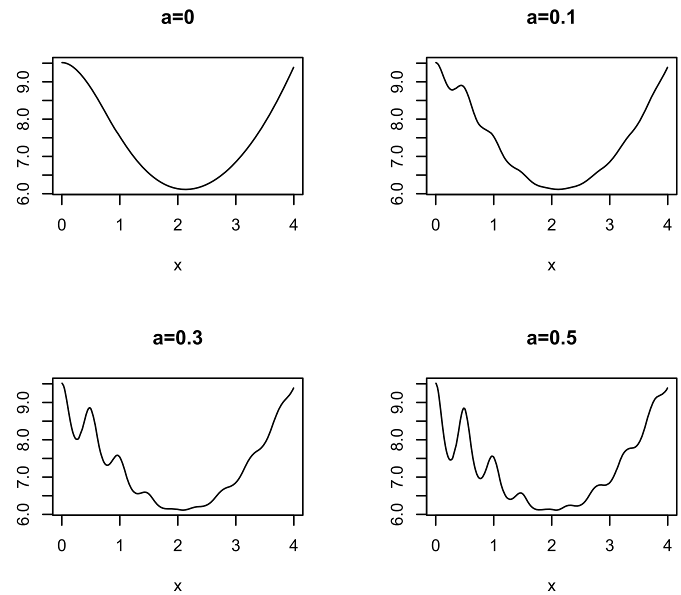

FIG 2. *Target function $\eta(x) = g\big(\int_0^\tau ydF(y|x)\big)$ for different values of a.*

The case $a = 0$ is the only situation where the guide is correct. Similarly to the Poisson model, the bandwidths are selected using 200 simulations. Note that the four settings are not comparable to each other since the target function changes for each value of a. Therefore, one can only compare the GLQL and the LQL estimators within each setting. As a quasi-likelihood function we choose $Q(\mu, y) = -\frac{y}{\mu} - \log|\mu|$. Using $N = 1000$ replications and samples of size $n = 400$, we calculate the estimators at ten equidistant data points in the interval $[0, 4]$. For a data point $x_i, i = 1, ..., 10$, we calculate the empirical bias by $b_i = N^{-1}\sum_{j=1}^{N}\left[\widehat{\eta}^{j}_{\hat{G},\widehat{\theta}}(x_i) - \eta(x_i)\right]$ and the empirical variance by $v_i^2 = N^{-1}\sum_{j=1}^{N}\left[\widehat{\eta}^{j}_{\hat{G},\widehat{\theta}}(x_i) - N^{-1}\sum_{j=1}^{N}\widehat{\eta}^{j}_{\hat{G},\widehat{\theta}}(x_i)\right]^2$, where $\widehat{\eta}^{j}_{\hat{G},\widehat{\theta}}(x_i)$ is the GLQL estimator for the j^{th} replication. Then we calculate $B^2 = 10^{-1}\sum_{i=1}^{10} b_i^2$, the average squared bias, $V = 10^{-1}\sum_{i=1}^{10} v_i^2$, the average variance and $MSE = B^2 + V$, the average mean squared error. The obtained results are summarized in Table 2. When the guide is correct $(a = 0)$ the GLQL estimator clearly outperforms the LQL estimator. In fact, in this case, the average squared bias is approximately reduced by half. For $a = 0.1, 0.3, 0.5$, even if the parametric guide is not correctly specified, the GLQL estimator behaves better than the classical LQL estimator. Regarding the variance, the guided estimator has generally smaller variance, except for the case $a = 0.5$ where we observe a slightly larger variance for the GLQL estimator. Finally, as noticed in the first example, the bandwidth selected for the GLQL method is generally larger than the one selected for the classical LQL method.

Table 2

Average squared bias (×10), average variance (×10), average MSE (×10) and the optimal bandwidth h of the estimators for different conditional mean functions ($a = 0, 0.1, 0.3, 0.5$) computed for ten equidistant data points. The samples are of size $n = 400$ with a censoring rate of 33.4%, and $N = 1000$ replications.

a	Method	Bias^2	Var	AMSE	h
0	GLQL	0.269	1.298	1.567	1.612
	LQL	0.489	1.369	1.858	1.263
0.1	GLQL	0.268	1.319	1.587	1.574
	LQL	0.493	1.391	1.884	1.264
0.3	GLQL	0.287	1.320	1.608	1.573
	LQL	0.549	1.369	1.918	1.263
0.5	GLQL	0.204	1.472	1.676	1.186
	LQL	0.518	1.440	1.958	1.147

5. Conclusions

Thanks to its bias reduction property, parametrically guided nonparametric estimation is more and more investigated in different areas of statistics. The application of the guided nonparametric method to density estimation, non-parametric regression, local quasi-likelihood, additive models and very recently varying coefficient models has revealed an improved performance for the guided estimator compared with the classical nonparametric estimator. However, most of these investigations are based on completely observed data.

In this paper, we focused on the adaptation of the parametrically guided local quasi-likelihood estimation to the censored data case. To deal with censoring, we considered the synthetic data approach. We investigated the simplest guided scheme which is based on the additive correction. We also generalized the asymptotic results to an unified family of additive-multiplicative corrections. Our results provide a generalization to the censored data case of both the results of [14] and [16]. The asymptotic results confirm the bias reduction property of the guided local quasi-likelihood estimator in the presence of censoring. The results also show that when an optimal bandwidth and an appropriate parametric guide are used the variance can also be reduced. Our finite sample simulation investigated both the case of discrete and continuous responses. The simulation results corresponded quite closely to the theoretical results and proved that the guided local quasi-likelihood estimator outperforms the unguided local-quasi-likelihood estimator in terms of bias and mean squared error.

Appendix: Proofs

We start this section with some notations. Set $\psi(x) = \eta(x) - \widetilde{\eta}(x)$, $\bar{\eta}(x, u) = \psi(x) + \psi'(x)(u - x)$ and $\mathbf{X}_k = (1, (X_k - x)/h)^T$ for $k = 1, \ldots, n$. Let S and S^* be the (2×2) matrices given by $S = (\mu^K_{i+j-2})_{1 \le i,j \le 2}$, $S^* = (\nu^K_{i+j-2})_{1 \le i,j \le 2}$, and set $U^K = (\mu^K_2, \mu^K_3)^T$ and $\mu_l = \int x^l K(x)dx$ for $l = 0, \cdots, 3$. Let $\widetilde{\beta}^*$ be a normalized estimator defined as follows:

$$\widetilde{\beta}^* = (nh)^{1/2}\big(\widetilde{\beta}_0 - \eta(x), h\{\widetilde{\beta}_1 - \psi'(x)\}\big)^T.$$

If $\widetilde{\beta} = (\widetilde{\beta}_0, \widetilde{\beta}_1)$ maximizes (3.1), then $\widetilde{\beta}^*$ maximizes

$$\sum_{k=1}^{n} Q\big(g^{-1}(\widetilde{\eta}(X_k) + \bar{\eta}(x, X_k) + (nh)^{-1/2}\beta^{*T}\mathbf{X}_k), \hat{Y}_k^*\big) K\left(\frac{X_k - x}{h}\right),$$

with respect to β^*. Also define the following quantities:

$$\mathbf{V}_{n,\hat{G}} = (nh)^{-1/2} \sum_{k=1}^{n} q_1\big(\widetilde{\eta}(X_k) + \bar{\eta}(x, X_k), \hat{Y}_k^*\big)\mathbf{X}_k K\left(\frac{X_k - x}{h}\right),$$

$$\mathbf{B}_{n,\hat{G}} = (nh)^{-1} \sum_{k=1}^{n} q_2\big(\widetilde{\eta}(X_k) + \bar{\eta}(x, X_k), \hat{Y}_k^*\big)\mathbf{X}_k \mathbf{X}_k^T K\left(\frac{X_k - x}{h}\right),$$

where $q_1(x, y) = (y - g^{-1}(x))\rho_1(x)$ and $q_2(x, y) = (y - g^{-1}(x))\rho_1'(x) - \rho_2(x)$. In order to prove Theorem 3.1, the following lemmas are needed.

Lemma A.1. *Under the assumptions of Theorem 3.1 we have,*

$$\mathbf{B}_{n,\hat{G}} = -\rho_2(\eta(x)) f_X(x) S + o_p(1) \equiv \mathbf{B} + o_p(1).$$

Proof. Set $\mathbf{B}_{n,G} = (nh)^{-1} \sum_{k=1}^{n} q_2\big(\widetilde{\eta}(X_k) + \bar{\eta}(x, X_k), Y_k^*\big)\mathbf{X}_k \mathbf{X}_k^T K\big(\frac{X_k - x}{h}\big)$. In view of conditions A1 and A7, for $1 \le i, j \le 2$, we have

$$\begin{aligned}
&\big|\mathbf{B}_{n,\hat{G}} - \mathbf{B}_{n,G}\big|_{ij} \\
&= (nh)^{-1}\Bigg| \sum_{k=1}^{n} \frac{\delta_k \phi(T_k)[\hat{G}(T_k^-|X_k) - G(T^-|X_k)]}{(1 - \hat{G}(T_k^-|X_k))(1 - G(T^-|X_k))} \rho_1'(\widetilde{\eta}(X_k) + \bar{\eta}(x, X_k)) \\
&\quad \times K\left(\frac{X_k - x}{h}\right)\left(\frac{X_k - x}{h}\right)^{i+j-2}\Bigg| \\
&\le O_p(1) \times \sup_{t \le \tau, x \in S_X} \big|\hat{G}(t^-|x) - G(t^-|x)\big| \\
&\quad \times (nh)^{-1} \sum_{k=1}^{n} K\left(\frac{X_k - x}{h}\right)\left|\frac{X_k - x}{h}\right|^{i+j-2}.
\end{aligned}$$

The above supremum tends to zero in probability by Proposition 4.3 in [41] and the empirical sum is bounded in probability by assumptions A1 and A5. Hence,

$$\mathbf{B}_{n,\hat{G}} - \mathbf{B}_{n,G} = o_p(1). \tag{A.1}$$

Now, note that $(\mathbf{B}_{n,G})_{ij} = (E\mathbf{B}_{n,G})_{ij} + O_p\big(Var\{(\mathbf{B}_{n,G})_{ij}\}^{1/2}\big)$. Since q_2 is linear in y and using A2, A5 and A7, we obtain that

$$\begin{aligned}
&(E\mathbf{B}_{n,G})_{ij} \\
&= h^{-1} E\left[q_2\big(\widetilde{\eta}(X_1) + \bar{\eta}(x, X_1), m(X_1)\big) K\left(\frac{X_1 - x}{h}\right)\left(\frac{X_1 - x}{h}\right)^{i+j-2}\right]
\end{aligned}$$

$$= h^{-1}\int q_2\big(\widetilde{\eta}(u)+\bar{\eta}(x,u),m(u)\big)f_X(u)K\left(\frac{u-x}{h}\right)\left(\frac{u-x}{h}\right)^{i+j-2}du$$
$$= \int q_2\big(\widetilde{\eta}(x+vh)+\bar{\eta}(x,x+vh),m(x+vh)\big)f_X(x+vh)K(v)v^{i+j-2}dv$$
$$\longrightarrow -\rho_2(\eta(x))f_X(x)\mu^K_{i+j-2} \text{ as } n\to\infty.$$

In view of Assumption 3.1, we have

$$(nh)Var(\mathbf{B}_{n,G})_{ij}$$
$$\leq h^{-1}E\left[q_2\big(\widetilde{\eta}(X_1)+\bar{\eta}(x,X_1),Y_1^*\big)^2K^2\left(\frac{X_1-x}{h}\right)\left(\frac{X_1-x}{h}\right)^{2(i+j-2)}\right]$$
$$\leq \int E\big[q_2\big(\widetilde{\eta}(X_1)+\bar{\eta}(x,X_1),Y_1^*\big)^2|X_1=x+vh\big]f_X(x+vh)K^2(v)v^{2(i+j-2)}dv$$
$$= O(1).$$

Therefore,

$$\mathbf{B}_{n,G} = -\rho_2(x)f_X(x)S+o_p(1). \tag{A.2}$$

The result of Lemma A.1 is now a direct consequence of (A.1) and (A.2). □

Lemma A.2. *Suppose that the assumptions of Theorem 3.1 hold. Then,*

$$V_{n,G}-EV_{n,G}\xrightarrow{d}\mathcal{N}\left(0,\sigma_*^2(x)f_X(x)\rho_1^2(\eta(x))S^*\right),$$

where $V_{n,G}=(nh)^{-1/2}\sum_{k=1}^n q_1\big(\widetilde{\eta}(X_k)+\bar{\eta}(x,X_k),Y_k^*\big)\mathbf{X}_kK\big(\frac{X_k-x}{h}\big)$.

Proof. Using Taylor's expansion of $\psi(\cdot)$, we have

$$\widetilde{\eta}(x+vh)+\bar{\eta}(x,x+vh)=\eta(x+vh)-\frac{\psi''(x)}{2}(vh)^2+o(h^2). \tag{A.3}$$

Since $q_1\big(\eta(\cdot),m(\cdot)\big)=0$ and using Taylor expansion and (A.3), we obtain

$$q_1\big(\widetilde{\eta}(x+vh)+\bar{\eta}(x,x+vh),m(x+vh)\big)=\frac{\psi''(x)}{2}(vh)^2\rho_2(\eta(x+vh))+o(h^2). \tag{A.4}$$

Note that $q_1(\cdot,y)$ is linear in y. For $1\leq i\leq 2$, we have

$$E(V_{n,G})_i \tag{A.5}$$
$$= (n/h)^{1/2}E\left[q_1\big(\widetilde{\eta}(X_1)+\bar{\eta}(x,X_1),m(X_1)\big)K\left(\frac{X_1-x}{h}\right)\left(\frac{X_1-x}{h}\right)^{i-1}\right]$$
$$= (nh)^{1/2}\int q_1\big(\widetilde{\eta}(x+vh)+\bar{\eta}(x,x+vh),m(x+vh)\big)f_X(x+vh)K(v)v^{i-1}dv$$
$$= (nh)^{1/2}\rho_2(\eta(x))f_X(x)h^2\frac{\psi''(x)}{2}\mu^K_{i+1}\{1+o(1)\}. \tag{A.6}$$

Thus,

$$E(V_{n,G})=(nh)^{1/2}\rho_2(\eta(x))f_X(x)h^2\frac{\psi''(x)}{2}U^K\{1+o(1)\}.$$

Now, from Assumption 3.1, we obtain

$$
\begin{aligned}
Var(V_{n,G}) &= h^{-1}Var\bigg[q_1\big(\widetilde{\eta}(X_1)+\bar{\eta}(x,X_1),Y_1^*\big)\mathbf{X}_1K\bigg(\frac{X_1-x}{h}\bigg)\bigg]\\
&= h^{-1}E\bigg[Var\bigg\{q_1\big(\widetilde{\eta}(X_1)+\bar{\eta}(x,X_1),Y_1^*\big)\mathbf{X}_1K\bigg(\frac{X_1-x}{h}\bigg)\bigg|X_1\bigg\}\bigg]\\
&\quad + h^{-1}Var\bigg[E\bigg\{q_1\big(\widetilde{\eta}(X_1)+\bar{\eta}(x,X_1),Y_1^*\big)\mathbf{X}_1K\bigg(\frac{X_1-x}{h}\bigg)\bigg|X_1\bigg\}\bigg]\\
&= W_{1n,G}+W_{2n,G}.
\end{aligned}
$$

For $1\le i\le 2$, we have

$$
\begin{aligned}
&(W_{1n,G})_{ij}\\
&\quad= h^{-1}E\bigg\{Var(Y_1^*|X_1)\rho_1^2(\widetilde{\eta}(X_1)+\bar{\eta}(x,X_1))\bigg(\frac{X_1-x}{h}\bigg)^{i+j-2}K^2\bigg(\frac{X_1-x}{h}\bigg)\bigg\}\\
&\quad= \int\sigma_*^2(x+vh)f_X(x+vh)\rho_1^2(\widetilde{\eta}(x+vh)+\bar{\eta}(x,x+vh))K^2(v)v^{i+j-2}dv\\
&\quad= \sigma_*^2(x)f_X(x)\rho_1^2(\eta(x))\nu_{i+j-2}^K+O(h).
\end{aligned}
$$

The second term $W_{2n,G}$ can be bounded as follows:

$$
W_{2n,G}\le h^{-1}E\bigg[q_1^2\big(\widetilde{\eta}(X_1)+\bar{\eta}(x,X_1),m(X_1)\big)\mathbf{X}_1\mathbf{X}_1^TK^2\bigg(\frac{X_1-x}{h}\bigg)\bigg].
$$

In view of expression (A.4) and conditions A1 and A7, we get

$$
\begin{aligned}
(W_{2n,G})_{ij} &\le \int q_1^2\big(\widetilde{\eta}(x+vh)+\bar{\eta}(x,x+vh),m(x+vh)\big)v^{i+j-2}K^2(v)dv\\
&= O(h^4), \qquad \text{for } 1\le i,j\le 2.
\end{aligned}
$$

Thus,

$$
Var(V_{n,G})=\sigma_*^2(x)f_X(x)\rho_1^2(\eta(x))S^*+o(1).
$$

Finally, it suffices to check the Lyapunov condition. Let $c\in\Re^2$, based on similar arguments to those used to develop (A.5), we can easily show that $\{c^TVar(V_{n,G})c\}^{-3/2}\sum_{k=1}^n|c^Tv_{G,k}-Ec^Tv_{G,k}|=O_p\big((nh)^{-1/2}\big)$, where $v_{G,k}=q_1\big(\widetilde{\eta}(X_k)+\bar{\eta}(x,X_k),Y_k^*\big)\mathbf{X}_kK\big(\frac{X_k-x}{h}\big)$. The result of Lemma A.2 is now a direct consequence of the Cramér-Wold device. □

Proof of Theorem 3.1. Consider $\ell_{n,\hat{G}}(\beta^*)$ the normalized function defined as follows

$$
\begin{aligned}
\ell_{n,\hat{G}}(\beta^*) = \sum_{k=1}^n\bigg\{&Q\big(g^{-1}(\widetilde{\eta}(X_k)+\bar{\eta}(x,X_k)+(nh)^{-1/2}\beta^{*T}\mathbf{X}_k),\hat{Y}_k^*\big)\\
&-Q\big(g^{-1}(\widetilde{\eta}(X_k)+\bar{\eta}(x,X_k)),\hat{Y}_k^*\big)\bigg\}K\bigg(\frac{X_k-x}{h}\bigg).
\end{aligned}
$$

Then, $\widetilde{\beta}^*$ maximizes $\ell_{n,\hat{G}}(\beta^*)$. Using a Taylor expansion of $Q\{g^{-1}(\cdot, \hat{Y}_i^*)\}$, we have

$$\ell_{n,\hat{G}}(\beta^*) = \mathbf{V}_{n,\hat{G}}^T \beta^* + \frac{1}{2}\beta^{*T}\mathbf{B}_{n,\hat{G}}\beta^*\{1 + o_p(1)\}.$$

Now by Lemma A.1 and the quadratic approximation lemma of [13], we obtain

$$\widetilde{\beta}^* = \mathbf{B}^{-1}\mathbf{V}_{n,\hat{G}} + o_p(1). \tag{A.7}$$

Next, write $\mathbf{V}_{n,\hat{G}} = (\mathbf{V}_{n,\hat{G}} - \mathbf{V}_{n,G}) + (\mathbf{V}_{n,G} - E\mathbf{V}_{n,G}) + E\mathbf{V}_{n,G}$. For $i = 1, 2$, we have

$$\begin{aligned}
&\big|\mathbf{V}_{n,\hat{G}} - \mathbf{V}_{n,G}\big|_i \\
&\quad = (nh)^{-1/2}\bigg|\sum_{k=1}^n \frac{\delta_k \phi(T_k)[\hat{G}(T_k^-|X_k) - G(T^-|X_k)]}{(1 - \hat{G}(T_k^-|X_k))(1 - G(T^-|X_k))}\rho_1(\widetilde{\eta}(X_k) + \bar{\eta}(x, X_k)) \\
&\qquad \times K\left(\frac{X_k - x}{h}\right)\left(\frac{X_k - x}{h}\right)^{i-1}\bigg| \\
&\quad \le (nh)^{1/2} O_p(1) \times \sup_{t \le \tau, x \in S_X} \big|\hat{G}(t^-|x) - G(t^-|x)\big| \\
&\qquad \times (nh)^{-1}\sum_{k=1}^n K\left(\frac{X_k - x}{h}\right)\left|\frac{X_k - x}{h}\right|^{i-1}.
\end{aligned}$$

From Proposition 4.3 in [41], it follows that if $\frac{nh_0^5}{\log n} = O(1)$, then,

$$\sup_{t \le \tau, x \in S_X} |\hat{G}(t^-|x) - G(t^-|x)| = O_p((nh_0)^{-1/2}(\log n)^{1/2}).$$

Since $(nh)^{-1}\sum_{k=1}^n K\big(\frac{X_k - x}{h}\big)\big|\frac{X_k - x}{h}\big|^{i-1} = O_p(1)$, we get

$$\mathbf{V}_{n,\hat{G}} - \mathbf{V}_{n,G} = O_p\bigg(\Big(\frac{h\log n}{h_0}\Big)^{1/2}\bigg). \tag{A.8}$$

Finally, from Lemma A.2 and equation (A.8), we obtain

$$\begin{aligned}
\widetilde{\beta}^* - (nh)^{1/2}\frac{1}{2}h^2\big(\eta''(x) - \widetilde{\eta}''(x)\big)S^{-1}U^K\{1 + o(1)\} + O_p\bigg(\Big(\frac{h\log n}{h_0}\Big)^{1/2}\bigg) \\
\xrightarrow{d} \mathcal{N}\big(0, \sigma_*^2(x) f_X(x)\rho_1^2(\eta(x))\mathbf{B}^{-1}S^*\mathbf{B}^{-1}\big).
\end{aligned} \tag{A.9}$$

The result of Theorem 3.1 is a special case of (A.9). □

Proof of Proposition 3.1.

1. Define

$$\begin{aligned}
\mathbb{L}(\theta) &= E_\Delta \log f(X, Y; \theta), \\
\widehat{\mathbb{L}}(\theta) &= \iint_\Delta \log f(x, y; \theta) d\widehat{F}(x, y).
\end{aligned}$$

Then, it is obvious that $\theta_* = \arg\max_{\theta\in\Theta} \mathbb{L}(\theta)$ and $\widehat{\theta} = \arg\max_{\theta\in\Theta} \widehat{\mathbb{L}}(\theta)$. The proof of the first point is a direct consequence of Theorem 2.1 in [33], under conditions D1.1, D2, D3, D5 and the first condition in Assumption 3.3.

2. In view of Corollary 5.8 in [1], conditions D2, D3 and D4 ensure an interchange of differentiation and integration. Since $\Omega = \Omega(\theta^*)$ is non-singular by condition D6 and using Assumption 3.3, we get

$$\sup_{\theta\in\Theta} || \nabla^2_\theta \widehat{\mathbb{L}}(\theta) - \Omega(\theta) ||= o_p(1),$$

and

$$\sqrt{n}\nabla^1_\theta \widehat{\mathbb{L}}(\theta_*) \xrightarrow{d} \mathcal{N}(0,\Sigma).$$

Therefore, the second point results directly from the first point together with Theorem 3.1 in [33]. □

Proof of Thereom 3.2. Similarly to the proof of Theorem 3.1, we now define $\widehat{\beta}^*$ and $\widetilde{\beta}^*$ the normalized guided estimators of β^* based on the estimated guide $\eta(.,\widehat{\theta})$ and the fixed guide $\eta(.,\theta_*)$, respectively. Define $\psi(x,\theta) = \eta(x) - \eta(x,\theta)$, $\bar{\eta}(x,u;\theta) = \psi(x,\theta) + \psi'(x,\theta)(u-x)$ and

$$B_{n,\hat{G}}(\theta) = (nh)^{-1}\sum_{k=1}^n q_2\big(\eta(X_k,\theta) + \bar{\eta}(x,X_k;\theta), \hat{Y}^*_k\big)\mathbf{X}_k\mathbf{X}_k^T K\left(\frac{X_k-x}{h}\right),$$
$$V_{n,\hat{G}}(\theta) = (nh)^{-1/2}\sum_{k=1}^n q_1\big(\eta(X_k,\theta) + \bar{\eta}(x,X_k;\theta), \hat{Y}^*_k\big)\mathbf{X}_k K\left(\frac{X_k-x}{h}\right).$$

Write $B_{n,\hat{G}}(\hat{\theta}) = [B_{n,\hat{G}}(\hat{\theta}) - B_{n,G}(\hat{\theta})] + [B_{n,G}(\hat{\theta}) - B_{n,G}(\theta_*)] + B_{n,G}(\theta_*)$, where $B_{n,G}(\theta) = (nh)^{-1}\sum_{k=1}^n q_2\big(\eta(X_k,\theta)+\bar{\eta}(x,X_k;\theta), Y^*_k\big)\mathbf{X}_k\mathbf{X}_k^T K\big(\frac{X_k-x}{h}\big)$. Using a Taylor expansion, for $i,j=1,2$, we have

$$\begin{aligned}
&(B_{n,G}(\hat{\theta}) - B_{n,G}(\theta_*))_{i,j}\\
&\quad = (nh)^{-1}\sum_{k=1}^n \nabla^1_\theta q_2\big(\eta(X_k,\tilde{\theta}) + \bar{\eta}(x,X_k;\tilde{\theta}), Y^*_k\big)(\widehat{\theta}-\theta_*)\\
&\qquad \times K\left(\frac{X_k-x}{h}\right)\left(\frac{X_k-x}{h}\right)^{i+j-2},
\end{aligned}$$

for $\tilde{\theta}$ between $\widehat{\theta}$ and θ_*. By assumptions A1, A7 and D1, there exists a constant $c>0$ such that

$$|B_{n,G}(\hat{\theta}) - B_{n,G}(\theta_*)|_{i,j} \le c\big\|\widehat{\theta}-\theta_*\big\| \times (nh)^{-1}\sum_{k=1}^n K\left(\frac{X_k-x}{h}\right)\left|\frac{X_k-x}{h}\right|^{i+j-2}.$$

Note that $\hat{\theta}-\theta_*$ converges to zero in probability by Proposition 3.1 and the empirical sum is bounded in probability by assumptions A1 and A5. Thus, $B_{n,G}(\hat{\theta}) - B_{n,G}(\theta_*) = o_p(1)$. Now, we have

$$|B_{n,\hat{G}}(\hat{\theta}) - B_{n,G}(\hat{\theta})| \le |D_{1n}| + |D_{2n}|,$$

where

$$
\begin{aligned}
&(D_{1n})_{i,j} \\
&\quad = (nh)^{-1}\sum_{k=1}^{n}\frac{\delta_k\phi(T_k)[\hat{G}(T_k^-|X_k)-G(T^-|X_k)]}{(1-\hat{G}(T_k^-|X_k))(1-G(T^-|X_k))}K\left(\frac{X_k-x}{h}\right)\left(\frac{X_k-x}{h}\right)^{i+j-2} \\
&\qquad \times [\rho_1(\eta(X_k,\hat{\theta})+\bar{\eta}(x,X_k;\hat{\theta}))-\rho_1(\eta(X_k,\theta_*)+\bar{\eta}(x,X_k;\theta_*))],
\end{aligned}
$$

and

$$
\begin{aligned}
&(D_{2n})_{i,j} \\
&\quad = (nh)^{-1}\sum_{k=1}^{n}\frac{\delta_k\phi(T_k)[\hat{G}(T_k^-|X_k)-G(T^-|X_k)]}{(1-\hat{G}(T_k^-|X_k))(1-G(T^-|X_k))}\rho_1(\eta(X_k,\theta_*)+\bar{\eta}(x,X_k;\theta_*)) \\
&\qquad \times K\left(\frac{X_k-x}{h}\right)\left(\frac{X_k-x}{h}\right)^{i+j-2}.
\end{aligned}
$$

By a Taylor expression and assumptions A1, A7 and D1, we have

$$
\begin{aligned}
|(D_{1n})_{i,j}| \le O_p(1)\times &\sup_{t\le\tau,x\in S_X}\left|\hat{G}(t^-|x)-G(t^-|x)\right|\left\|\widehat{\theta}-\theta_*\right\| \\
&\times (nh)^{-1}\sum_{k=1}^{n}K\left(\frac{X_k-x}{h}\right)\left|\frac{X_k-x}{h}\right|^{i+j-2}.
\end{aligned}
$$

From Assumptions 3.1, 3.4, Proposition 4.3 in [41] and Proposition 3.1, we get $D_{1n}=o_p(1)$. Similar arguments give $D_{2n}=o_p(1)$. Thus, $B_{n,\hat{G}}(\hat{\theta})=B_{n,G}(\theta_*)+o_p(1)$. Therefore, by Lemma A.1, we get

$$
B_{n,\hat{G}}(\hat{\theta})=\mathbf{B}+o_p(1). \tag{A.10}
$$

Using (A.10) and similar arguments to those used to get equation (A.7), we have

$$
\widehat{\beta}^*-\widetilde{\beta}^*=\mathbf{B}^{-1}[V_{n,\hat{G}}(\widehat{\theta})-V_{n,\hat{G}}(\theta_*)]+o_p(1).
$$

Now, write $V_{n,\hat{G}}(\widehat{\theta})-V_{n,\hat{G}}(\theta_*)=[V_{n,G}(\widehat{\theta})-V_{n,G}(\theta_*)]+[V_{n,\hat{G}}(\widehat{\theta})-V_{n,G}(\widehat{\theta})]+[V_{n,G}(\theta_*)-V_{n,\hat{G}}(\theta_*)]$, where
$V_{n,G}(\theta)=(nh)^{-1/2}\sum_{k=1}^{n}q_1\big(\eta(X_k,\theta)+\bar{\eta}(x,X_k;\theta),Y_k^*\big)\mathbf{X}_kK\big(\frac{X_k-x}{h}\big)$. Using a Taylor expansion, for $\dot{\theta}$ between $\widehat{\theta}$ and θ_*, we have

$$
\begin{aligned}
&V_{n,G}(\widehat{\theta})-V_{n,G}(\theta_*) \\
&\quad = (nh)^{-1/2}\sum_{k=1}^{n}\nabla_{\theta}^{1}q_1\big(\eta(X_k,\dot{\theta})+\bar{\eta}(x,X_k;\dot{\theta}),Y_k^*\big)(\widehat{\theta}-\theta_*)\mathbf{X}_kK\left(\frac{X_k-x}{h}\right).
\end{aligned}
$$

By assumptions A1, A7 and D1, there exists a constant $c>0$ such that for $i=1,2$, we get

$$
\begin{aligned}
&|V_{n,G}(\widehat{\theta}) - V_{n,G}(\theta_*)|_i \\
&\quad \le (nh)^{1/2} c \big\|\widehat{\theta} - \theta_*\big\| \times (nh)^{-1} \sum_{k=1}^{n} K\left(\frac{X_k - x}{h}\right) \left|\frac{X_k - x}{h}\right|^{i-1}.
\end{aligned}
$$

Since $\big\|\widehat{\theta} - \theta_*\big\| = O_p(n^{-1/2})$, we have $V_{n,G}(\widehat{\theta}) - V_{n,G}(\theta_*) = O_p(h^{1/2})$. Using expression (A.8), we get $V_{n,G}(\theta_*) - V_{n,\hat{G}}(\theta_*) = O_p\big(\big(\frac{h\log n}{h_0}\big)^{1/2}\big)$. Finally, write

$$
V_{n,\hat{G}}(\widehat{\theta}) - V_{n,G}(\widehat{\theta}) = I_{1n} + I_{2n},
$$

where (for $i = 1, 2$)

$$
\begin{aligned}
&(I_{1n})_i \\
&\quad = (nh)^{-1/2} \sum_{k=1}^{n} \frac{\delta_k \phi(T_k)[\hat{G}(T_k^-|X_k) - G(T^-|X_k)]}{(1 - \hat{G}(T_k^-|X_k))(1 - G(T^-|X_k))} K\left(\frac{X_k - x}{h}\right)\left(\frac{X_k - x}{h}\right)^{i-1} \\
&\qquad \times [\rho_1(\eta(X_k, \hat{\theta}) + \bar{\eta}(x, X_k; \hat{\theta})) - \rho_1(\eta(X_k, \theta_*) + \bar{\eta}(x, X_k; \theta_*))], \\
&(I_{2n})_i \\
&\quad = (nh)^{-1/2} \sum_{k=1}^{n} \frac{\delta_k \phi(T_k)[\hat{G}(T_k^-|X_k) - G(T^-|X_k)]}{(1 - \hat{G}(T_k^-|X_k))(1 - G(T^-|X_k))} K\left(\frac{X_k - x}{h}\right)\left(\frac{X_k - x}{h}\right)^{i-1} \\
&\qquad \times \rho_1(\eta(X_k, \theta_*) + \bar{\eta}(x, X_k; \theta_*)).
\end{aligned}
$$

By a Taylor expression and conditions A1, A7 and D1, we have

$$
\begin{aligned}
&|I_{1n}|_i \\
&\quad \le O_p(1) \times (nh)^{1/2} \times \sup_{t \le \tau, x \in S_X} \big|\hat{G}(t^-|x) - G(t^-|x)\big| \\
&\qquad \times \big\|\widehat{\theta} - \theta_*\big\| \times (nh)^{-1} \sum_{k=1}^{n} K\left(\frac{X_k - x}{h}\right)\left|\frac{X_k - x}{h}\right|^{i-1}.
\end{aligned}
$$

From Proposition 3.4 in [41] and Proposition 3.1, we get $I_{1n} = O_p\big(\big(\frac{h\log n}{nh_0}\big)^{1/2}\big)$. Similar arguments give $I_{2n} = O_p\big(\big(\frac{h\log n}{h_0}\big)^{1/2}\big)$. Hence,

$$
V_{n,\hat{G}}(\widehat{\theta}) - V_{n,\hat{G}}(\theta_*) = O_p\left(\left(\frac{h\log n}{h_0}\right)^{1/2}\right) + o_p(1).
$$

Therefore,

$$
\widehat{\beta}^* - \widetilde{\beta}^* = O_p\left(\left(\frac{h\log n}{h_0}\right)^{1/2}\right) + o_p(1). \tag{A.11}
$$

Finally, write $\widehat{\beta}^* = [\widehat{\beta}^* - \widetilde{\beta}^*] + \widetilde{\beta}^*$. Form expressions (A.9) and (A.11), we get

$$
\begin{aligned}
&\widehat{\beta}^* - (nh)^{1/2}\frac{1}{2}h^2\big(\eta''(x) - \eta''(x, \theta_*)\big)S^{-1}U^K\{1 + o(1)\} + O_p\left(\left(\frac{h\log n}{h_0}\right)^{1/2}\right) \\
&\qquad \xrightarrow{d} \mathcal{N}\big(0, \sigma_*^2(x) f_X(x) \rho_1^2(\eta(x))\mathbf{B}^{-1}S^*\mathbf{B}^{-1}\big).
\end{aligned}
$$

This concludes the proof of Theorem 3.2 which is a special case of this last result. □

Acknowledgements

The authors would like to thank the Associate Editor and two referees for their valuable comments and suggestions on the paper.

References

[1] Bartle, R. (1966). *The Elements of Integration*, John Wiley and Sons, New York. MR0200398

[2] Beran, R. (1981). Nonparametric regression with randomly censored survival data. *Technical report, Univ. California, Berkeley.*

[3] Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika* **66(3)**, 429–436.

[4] Chen, J., Fan, J., Li, K. and Zhou, H. (2006). Local quasi-likelihood estimation with data missing at random. *Statist. Sinica* **16**, 1071–1100. MR2327481

[5] Dabrowska, D. M. (1987). Nonparametric regression with censored survival time data. *Scand. J. Stat.* **14**, 181–197. MR0932943

[6] Dabrowska, D. M. (1992). Variable bandwidth conditional Kaplan-Meier estimate. *Scand. J. Stat.* **19**, 351–361. MR1211789

[7] Davenport, C., Maity, A. and Wu, Y. (2015). Parametrically guided estimation in nonparametric varying coefficient models with quasi-likelihood. *J. Nonparametr. Statist.* **27**, 195–213. MR3345483

[8] Delecroix, M., Lopez, O. and Patilea, V. (2008). Nonlinear censored regression using synthetic data. *Scand. J. Stat.* **35**, 248–265. MR2418739

[9] Doksum, K. A. and Yandell, B. S. (1982). Properties of regression estimates based on censored survival. In Festschrift for Erich L. Lehmann (eds P. J. Bickel, K. A. Doskum, & J. L. Hodges), 140–156. MR0689744

[10] El Ghouch, A. and Van Keilegom, I. (2008). Nonparametric regression with dependent censored data. *Scand. J. Stat.* **35**, 228–247. MR2418738

[11] Fan, J. and Chen, J. (1999). One-step local quasi-likelihood estimation. *J. Roy. Statist. Soc. Ser. B,* **64**, 927–943. MR1722248

[12] Fan, J. and Gijbels, I. (1994). Censored regression: local linear approximations and their applications. *J. Amer. Statist. Assoc.* **89**, 560–570. MR1294083

[13] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and Its Applications.* Chapman & Hall, London. MR1383587

[14] Fan, J., Heckman, N. E. and Wand, M. P. (1995). Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *J. Amer. Statist. Assoc.* **90**, 141–150. MR1325121

[15] Fan, J., Maity, A., Wang, Y. and Wu, Y. (2014). Parametrically guided generalized additive models with application to mergers and acquisitions data. *J. Nonparametr. Statist.* **25**, 109–128. MR3039973

[16] Fan, J., Wu, Y. and Feng, Y. (2009). Local quasi-likelihood with a parametric guide. *Ann. Statist.* **37**, 4153–4183. MR2572456
[17] Glad, I. K. (1998 a). A note on unconditional properties of a parametrically guided Nadaraya-Watson estimator. *Statistics and Probability Letters* **37(1)**, 101–108. MR1622670
[18] Glad, I. K. (1998 b). Parametrically guided nonparametric regression. *Scand. J. Stat.* **25**, 649–668. MR1666776
[19] Gonzalez-Manteiga, W. and Cadarso-Suarez, C. (1994). Asymptotic properties of generalized Kaplan-Meier estimator with some applications. *J. Nonparametr. Statist.* **4**, 65–78. MR1366364
[20] Green, P. J. and Yandell, B. S. (1985). Semi-parametric generalized linear models. In *Lecture Notes in Statistics* **32**, 44–55.
[21] Heuchenne, C. and Van Keilegom, I. (2012). Estimation of a general parametric location in censored regression. In: Van Keilegom, I. and Wilson, P. W. (eds.), *Exploring research frontiers in contemporary statistics and econometrics- Festschrift in honor of L. Simar*, Springer, 177–187. MR3051686
[22] Hjort, N. L. and Glad, I. K. (1995). Nonparametric density estimation with parametric start. *Ann. Statist.* **23**, 882–904. MR1345205
[23] Hunsberger, S. (1994). Semiparametric regression in likelihood-based models. *J. Amer. Statist. Assoc.* **89**, 1354–1365. MR1310226
[24] Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53**, 457–481. MR0093867
[25] Koul, H., Susarla, V. and Van Ryzin, J. (1981). Regression analysis with randomly right censored data. *Ann. Statist.* **9**, 1276–1288. MR0630110
[26] Lai, T. L. and Ying, Z. (1991). Large-sample theory of a modified Buckley-James estimator for regression analysis with censored data. *Ann. Statist.* **19**, 1370–1402. MR1126329
[27] Leurgans, S. (1987). Linear models, random censoring and synthetic data. *Biometrika* **74**, 301–309. MR0903130
[28] Lopez, O. (2011). Nonparametric estimation of the multivariate distribution function in a censored regression model with applications. *Comm. Statist. Theory Methods* **40**, 2639–2660. MR2860770
[29] Lu, X. and Cheng, T. L. (2007). Randomly censored partially linear single-index models. *J. Multivariate Anal.* **98(10)**, 1895–1922. MR2396946
[30] Martins-Filho, C., Mishra, S. and Ullah, A. (2008). A class of improved parametrically guided nonparametric regression estimators. *Econometric Reviews* **27**, 542–573. MR2439867
[31] McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed. Chapman & Hall. MR3223057
[32] Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *J. Roy. Statist. Soc. Ser. A* **135**, 370–384.
[33] Newey, W. K. and McFadden, D. (1999). Large sample estimation and hypothesis testing. In *Handbook of Econometrics*. Vol. **4**. Edited by D. McFadden & R. Engle, Amsterdam, The Netherlands. MR1315971

[34] O'Sullivan, F., Yandell, B. S. and Raynor, W. (1986). Automatic smoothing of regression functions in generalized linear models. *J. Amer. Statist. Assoc.* **81**, 96–103. MR0830570

[35] Rahman, M. and Ullah, A. (2002). Improved combined parametric and nonparametric regression: estimation and hypothesis testing. In *Handbook of Applied Econometrics and Statistical Inference*, Edited by Ullah, A., Wan, A., Chaturvedi, A. Marcel Dekker, New York. MR1893335

[36] Severini, T. A. and Staniswalis, J. G. (1994). Quasi-likelihood estimation in semi-parametric models. *J. Amer. Statist. Assoc.* **89**, 501–511. MR1294076

[37] Staniswalis, J. G. (1989). The kernel estimate of a regression function in likelihood-based models. *J. Amer. Statist. Assoc.* **84**, 276–283. MR0999689

[38] Suzukawa, A., Imai, H. and Sato, Y. (2001). Kullback-Leibler information consistent estimation for censored data. *Ann. Inst. Statist. Math.* **53**, 262–276. MR1841135

[39] Talamakrouni, M., El Ghouch, A. and Van Keilegom, I. (2015). Guided censored regression. *Scand. J. Stat.* **42**, 214–233. MR3318033

[40] Talamakrouni, M., Van Keilegom, I. and El Ghouch, A. (2016). Parametrically guided nonparametric density and hazard estimation with censored data. *Comput. Statist. Data Anal.* **93**, 308–323. MR3406214

[41] Van Keilegom, I. and Akritas, M. G. (1999). Transfer of tail information in censored regression models. *Ann. Statist.* **27**, 1745–1784. MR1742508

[42] Van Keilegom, I. and Veraverbeke, N. (1997). Estimation and bootstrap with censored data in fixed design nonparametric regression. *Ann. Inst. Statist. Math.* **49**, 467–491. MR1482368

[43] Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* **61**, 439–447. MR0375592

[44] Yu, L., Yu, R. and Liu, L. (2009). Quasi-likelihood for right-censored data in the generalized linear model. *Comm. Statist. Theory Methods* **38**, 2187–2200. MR2596906

[45] Yu, L. (2011). Nonparametric quasi-likelihood for right censored data. *Lifetime Data Anal.* **17**, 594–607. MR2838473

[46] Yu, L. and Peace, K. E. (2012). Spline nonparametric quasi-likelihood regression within the frame of the accelerated failure time model. *Comput. Statist. Data Anal.* **56**, 2657–2687. MR2915154

[47] Yu, L., Yu, R., Liu, L. and Chen, D. G. (2012). Extended quasi-likelihood with fractional polynomials in the frame of the accelerated failure time model. *Statistics in Medicine* **31(13)**, 1369–1379. MR2925055

[48] Zheng, Z. (1987). A class of estimators of the parameters in linear regression with censored data. *Acta Mathematicae Applicatae Sinica* **3**, 231–241.