Median regression and the missing information principle

Ian W. McKeague
Department of Statistics
Florida State University

Sundarraman Subramanian
Department of Mathematics & Statistics
University of Maine

Yanqing Sun
Department of Mathematics
University of North Carolina at Charlotte

October 20, 1999

Abstract

Median regression analysis has robustness properties which make it an attractive alternative to regression based on the mean. In this paper, the missing information principle is applied to a right-censored version of the median regression model, leading to a new estimator for the regression parameters. Our approach adapts Efron's derivation of self-consistency for the Kaplan–Meier estimator to the context of median regression; we replace the least absolute deviation estimating function by its (estimated) conditional expectation given the data. The new estimator is shown to be asymptotically equivalent to an *ad hoc* estimator introduced by Ying, Jung and Wei, and to have improved moderate-sample performance in simulation studies.

Key words: Least absolute deviation, martingale, heteroscedasticity, counting processes, kernel conditional Kaplan–Meier estimator, Cox proportional hazards.

1 Introduction

In survival analysis it is frequently of interest to estimate median life length at given covariate levels. In this paper we consider a right-censored version of the median regression model in which the median is specified as a linear function of covariates.

Median regression offers an appealing alternative to the popular Cox proportional hazards (PH) and accelerated failure time (AFT) approaches to survival analysis, with robustness and ease of interpretation being the principal benefits. Extracting information on median survival from PH and AFT analyses is complicated; for PH models, bootstrap techniques are needed (Burr and Doss, 1993); for AFT models, it is not possible to estimate the intercept parameter effectively in the presence of right censoring (Meier, 1975). In median regression, on the other hand, information about median life length is directly available once the regression parameters are estimated.

We study the problem of fitting a heteroscedastic median regression model from right-censored data. Let T be the response of interest, Z a $p+1$ dimensional covariate whose first component is taken to be 1, and suppose

$$T = \beta' Z + \epsilon, \tag{1.1}$$

where β is an unknown vector of regression parameters. The joint distribution of the observation error ϵ and the covariate Z is unknown, but the conditional median of ϵ given Z is known to be 0. The statistical problem is to estimate the true value of β based on a random sample of observations from the distribution of (X, δ, Z), where $X = T \wedge C$, C is a censoring variable, and $\delta = I(X = T)$. It is assumed that T and C are conditionally independent given Z.

This median regression model for censored data was developed by Ying, Jung and Wei (1995), henceforth YJW, who proposed an *ad hoc* estimator for the regression parameter β under the assumption that C is independent of Z, or that Z takes discrete values. In earlier work, Newey and Powell (1990) found an efficient estimator for β in the special case in which C is always observed. Robins (1996) developed an extension of the model to allow surrogate marker information, and constructed locally efficent estimators at a parametric submodel. For uncensored data, Jung (1996) obtained an efficient estimating function for the median regression parameters based on quasi-likelihood. Recently, Yang (1999) considered the case in which the observation error ϵ is independent of Z, and constructed relatively simple estimators in this case.

We propose a new estimator for the regression parameter β for the general case in which the censoring C and the observation error ϵ can depend nonparametrically on the covariate. The idea is to replace the least absolute deviation (LAD) estimating function, which is not available in the censored data context, by its (estimated) conditional expectation given the observed data. Our approach is an application of the missing information principle of Orchard and Woodbury (1972), see also Laird (1985). The missing information principle (MIP) gives a heuristic strategy for constructing estimators in missing data problems: replace a full-data estimating function or estimator by its estimated conditional expectation given the observed data. When applied to a parametric score function, the MIP amounts to a single iteration of the EM-algorithm (Dempster et al., 1977).

For estimation of a survival function from right censored data, application of the MIP produces an asymptotically efficient nonparametric maximum likelihood estimator (NPMLE)—the Kaplan–Meier estimator. Indeed, the conditional expectation of the full-data empirical distribution function given the observed data produces Efron's (1967) self-consistency equation which has the Kaplan–Meier estimator as a unique solution; iterations of the self-consistency equation form an EM-algorithm. For general missing data problems, every NPMLE of the full-data distribution is self-consistent, and, conversely, there exists a self-consistent estimator which is a NPMLE, see Tsai and Crowley (1985) and Gill (1989). A self-consistency equation may have a wide class of solutions, some of which may be inconsistent; in some cases iteration is needed to solve the equation, e.g., for interval censored data, see Groeneboom and Wellner (1992). The MIP has been discussed in the context of the accelerated failure time model by Lai and Ying (1994).

In the present setting, application of the MIP leads to an estimating equation which involves the conditional distribution function of T given Z, and in general this function would need to be estimated nonparametrically. To avoid the curse of dimensionality, we shall restrict attention to discrete covariates, or assume a Cox PH model for the conditional hazard function of T given Z. For discrete covariates, we show that our estimator is consistent, asymptotically normal, and asymptotically equivalent to YJW's estimator. We compare the moderate-sample performance of the two estimators via simulation, and show that our estimator offers considerable improvements.

The paper is organized as follows. In Section 2 we introduce the proposed estimator of the median regression parameter β. In Section 3, we present our numerical results. The asymptotic properties are described in Section 4. In Section 4.1, we state the theorem on the asymptotic consistency and asymptotic normality of the proposed estimator. A proof of the theorem is given in the Appendix following Section 4. In Section 4.2, we show that the ad hoc estimator of YJW is asymptotically equivalent to our estimator based on the MIP.

2 Proposed estimating equation

In this Section, we recall YJW's estimating function and introduce the proposed alternative based on an application of the MIP. Throughout the paper, $(X_i, \delta_i, Z_i), i = 1, \ldots, n$ denotes a random sample of observations from (X, δ, Z) under the model (1.1).

For uncensored data, the LAD estimator of β is obtained by minimizing $\sum_{i=1}^n |T_i - \beta' Z_i|$, or by solving the LAD estimating equation

$$\sum_{i=1}^n \left(I(T_i \geq \beta' Z_i) - \frac{1}{2} \right) Z_i \approx 0. \tag{2.1}$$

A "root" of this estimating equation is a minimizer of the Euclidean norm of the estimating function. For censored data, under the assumption that C and Z are independent, YJW proposed the following adjusted estimating equation:

$$\sum_{i=1}^n \left(\frac{I(X_i \geq \beta' Z_i)}{\hat{G}(\beta' Z_i)} - \frac{1}{2} \right) Z_i \quad \approx \quad 0, \tag{2.2}$$

where $\hat{G}$ is the Kaplan–Meier estimator of G, the survival function of C. The rationale given by YJW for this estimating equation is that $I(X \geq \beta_0' Z)$ has expectation $G(\beta_0' Z)/2$, where β_0 is the true value of β. When the censoring is dependent on the covariate Z, and Z takes only finitely many values, YJW instead proposed the estimating function

$$S_n^*(\beta) = \sum_{i=1}^n \left(\frac{I(X_i \geq \beta' Z_i)}{\hat{G}(\beta' Z_i, Z_i)} - \frac{1}{2} \right) Z_i, \tag{2.3}$$

where $\hat{G}(t,z)$ is the local Kaplan–Meier estimator of $G(t,z) = P(C > t|Z = z)$, the conditional survival function of C given Z.

Our approach is to apply the MIP to the uncensored LAD estimating equation (2.1). The idea is to replace the unobservable $I(T_i \geq \beta' Z_i)$ by an estimate of its conditional expectation given the data. It can be shown that (cf. Efron (1967), p. 840, equation (7.4))

$$E_i \equiv E(I(T_i \geq \beta' Z_i)|(Z_i, \delta_i, X_i)) = I(X_i \geq \beta' Z_i) + (1-\delta_i) I(X_i < \beta' Z_i) \frac{F(\beta' Z_i, Z_i)}{F(X_i, Z_i)}, \tag{2.4}$$

where $F(t,z) = P(T > t|Z = z)$ is the conditional survival function of T given Z.

The new estimating equation has the form

$$S_n(\beta) = \sum_{i=1}^{n} (\hat{E}_i - \frac{1}{2}) Z_i \approx 0,$$

where $\hat{E}_i$ is an estimate of E_i. The function F in (2.4) is unknown and needs to be estimated nonparametrically or semiparametrically by an appropriate estimator $\hat{F}$. If Z takes only finitely many values, we suggest estimating $F(t,z)$ by the local Kaplan-Meier estimator; this case is considered in detail in Section 4. For a one-dimensional covariate ($p = 1$), Dabrowska's (1989) kernel conditional Kaplan–Meier estimator can be used to estimate $F(t,z)$ nonparametrically. For $p \geq 3$, we recommend a Cox PH model based estimate of F, see Andersen et al. (1993, p.509); this estimator of F is used in the numerical examples in the next section. We arrive at the estimating function

$$S_n(\beta) = \sum_{i=1}^{n} \left(\left\{ I(X_i \geq \beta' Z_i) + \frac{I(X_i < \beta' Z_i, \delta_i = 0) \hat{F}(\beta' Z_i, Z_i)}{\hat{F}(X_i, Z_i)} \right\} - \frac{1}{2} \right) Z_i$$

and define $\hat{\beta}$ to be a minimizer of the Euclidean norm of $S_n(\beta)$.

It is difficult to estimate F nonparametrically because of the curse of dimensionality. This is a common problem with estimation in high dimensions—the higher the dimension the more spread apart are the data points, and the larger the data set required for a sensible analysis. For that reason we recommend using a Cox PH model to estimate F when $p \geq 3$. This creates a potential conflict with the median regression model because the conditional median under the Cox PH model is not necessarily a linear function of Z, so $\hat{\beta}$ may be inconsistent if the Cox PH model is misspecified. Nevertheless, the conditional median is a smooth function of $\beta' Z$, where β is the PH model regression parameter (under the mild condition that the baseline hazard function of the PH model is positive everywhere). The median regression model therefore appears as the first step in a Taylor series approximation to the median of the Cox PH model. In Section 3 we find that this approximation is adequate in simulations; use of a Cox PH model for F apparently has no detrimental effect on the performance of $\hat{\beta}$.

In general, YJW's estimator requires the conditional survival function G of C given Z to be estimated nonparametrically, so the curse of dimensionality is unavoidable there as well. YJW focused on the case that the censoring is independent of the covariate, but this assumption is too strong for most survival analysis applications. A less restrictive approach is to use a Cox PH model based estimate for G; this is the version of YJW's estimator we use for comparison with the proposed estimator in the next section.

Theoretical results obtained in Section 4 show that the two estimators are asymptotically equivalent in the case of discrete covariates. The choice of which estimator to use should therefore primarily be based on a comparison of their moderate-sample performance. In practice, when dealing with high-dimensional covariates, the adequacy of the Cox PH model based estimates of F and G should also be taken into consideration.

Robins (1996) addressed the curse of dimensionality by assuming (1) a parametric model for the conditional distribution of ϵ given Z, and (2) a Cox PH model for the conditional distribution of C given Z. His estimator is asymptotically efficient under (1) and (2), remains consistent if the model in (1) is misspecified, but is inconsistent if the model in (2) is misspecified. Robins was able to use high-dimensional surrogate marker information to further increase estimator efficiency, but his approach appears difficult to implement in practice. Yang (1999) avoided the curse of dimensionality by restricting attention to the homoscedastic case in which the observation error ϵ is independent of Z (so the covariate only influences the survival time though its median), and found relatively simple ad hoc estimators of the regression parameters in that case.

3 Numerical results

In this section we report some simulation results comparing the moderate-sample performance of the proposed estimator to that of YJW's estimator. Throughout, Cox PH models are used to estimate both F and G, as discussed above. We consider three scenarios for the simulation model:

1. The conditional hazard function of T given Z follows a Cox PH model, but the conditional hazard of C given Z is misspecified, i.e., departs significantly from a Cox PH model.

2. The conditional hazard function of T given Z is misspecified but the conditional hazard of C given Z follows a Cox PH model.

3. The conditional hazard functions of T and C given Z are both misspecified.

In each simulation example we used $Z = (1, Z_2)'$, with Z_2 uniformly distributed on $[1, 2]$. The means and root mean squared errors (RMSE) of the estimators are based on 10,000 Monte Carlo replications at a given sample size. A grid search over the rectangle $(-2, 2) \times (-1, 3)$ was used to locate the solution of each estimating equation.

For scenario 1, the conditional distribution of T was taken to be exponential with parameter $\log(2)/z_2$. This gives a median regression model (the conditional median of T given Z is $\beta_0' z$ where $\beta_0 = (0, 1)'$), and also a Cox PH model (with covariate $\log z_2$, regression parameter

-1, and baseline hazard $\log(2)$). The conditional hazard function of C given Z was specified as $\lambda_c(t|z_2) = 0.25m(z_2)$ where $m(x) = \min(x, 3-x)$, which is a strong departure from a Cox PH model. The factor 0.25 was used to calibrate the censoring rate at about 40%. The Cox PH model based-estimates for F and G both used $\log z_2$ as the covariate. From the results in Table 1, we see that the bias in the YJW estimator is large. The proposed estimator has negligible bias and lower RMSE.

Table 1. Comparison between the proposed estimator and YJW's estimator. $T|Z \sim$ exponential$(\log(2)/z_2)$, and $\lambda_c(t|z_2) = 0.25m(z_2)$. The true value of the intercept is 0, and the true value of the slope is 1. Sample size is 100 and censoring rate is about 40%.

Estimator	*Intercept*		*Slope*	
	Mean	RMSE	Mean	RMSE
Proposed	0.0184	1.0146	1.0038	0.7265
Ying et al.	-0.1137	1.0800	1.1098	0.7898

For scenario 2, the conditional distribution of the failure time T was taken as either $N(z_2, 0.5)$, or $z_2 + U$, where U is uniform on the interval $[-0.5, 0.5]$. In each case we have a median regression model with true value of $\beta_0 = (0, 1)'$, but departure from a Cox PH model. A Cox PH model was used for the censoring: $\lambda_c(t|z_2) = 0.25z_2$, which is a Cox PH model with covariate $\log z_2$, regression parameter 1, and baseline hazard 0.25. As before, the factor 0.25 was used to calibrate the censoring rate at about 40%. The results in Table 2 show that, despite misspecification, the proposed estimator outperforms the YJW estimator in terms of having lower RMSE. The bias in the proposed estimator is lower in one case and slightly more in the other case.

Table 2. Comparison between the proposed estimator and YJW's estimator. The distribution of T given Z is misspecified, but distribution of C given Z is specified correctly by a Cox PH model. The true value of the intercept is 0, and the true value of the slope is 1. Sample size is 100 and censoring rate about 40%.

Simulation model	*Estimator*	*Intercept*		*Slope*	
		Mean	RMSE	Mean	RMSE
$T|Z \sim N(z_2, 0.5)$	Proposed	0.0015	0.5441	1.0038	0.3644
$\lambda_c(t|z_2) = 0.25z_2$	Ying et al.	-0.0375	0.6643	1.0402	0.4538
$T|Z \sim z_2 + U$	Proposed	0.0035	0.2990	0.9793	0.2004
$\lambda_c(t|z_2) = 0.25z_2$	Ying et al.	-0.0108	0.3732	1.0135	0.2518

We also considered the examples of Table 2 with increased sample size, $n = 200$, see Table 3. The proposed estimator still has lower RMSE than the YJW estimator, but its bias is considerably more than YJW's in the second example.

Table 3. Comparison between the proposed estimator and YJW's estimator. Distribution of T given Z misspecified, but distribution of C given Z specified correctly by a Cox PH model. The true value of the intercept is 0, and the true value of the slope is 1. Sample size is 200 and censoring rate about 40%.

Simulation model	*Estimator type*	*Intercept*		*Slope*	
		Mean	RMSE	Mean	RMSE
$T\|Z \sim N(z_2, 0.5)$	Proposed	-0.0256	0.4001	1.0248	0.2707
$\lambda_c(t\|z_2) = 0.25z_2$	Ying et al.	-0.0303	0.4724	1.0312	0.3229
$T\|Z \sim z_2 + U$	Proposed	-0.0528	0.2486	1.0618	0.1827
$\lambda_c(t\|z_2) = 0.25z_2$	Ying et al.	-0.0096	0.2788	1.0103	0.1874

We next considered five different examples for scenario 3, see Table 4. The failure time was taken as uniform or normal. The censoring time was taken as uniform, normal, or with hazard $m(z_2)/4$; Cox PH models are misspecified in each case. The proposed estimator outperformed the YJW estimator in all these examples.

Table 4. Comparison between the proposed estimator and YJW's estimator. Conditional distributions of T given Z and of C given Z are both misspecified. The true value of the intercept is 0, and the true value of the slope is 1. Sample size is 100 and the censoring rate is about 40%.

Simulation model	*Estimator*	*Intercept*		*Slope*	
		Mean	RMSE	Mean	RMSE
$T\|Z \sim z_2 + U$	Proposed	-0.0158	0.2992	1.0189	0.2065
$\lambda_c(t\|z_2) = 0.25m(z_2)$	Ying et al.	-0.0110	0.3434	1.0087	0.2264
$T\|Z \sim z_2(2U + 1)$	Proposed	-0.0518	0.7385	1.0342	0.5249
$\lambda_c(t\|z_2) = 0.25m(z_2)$	Ying et al.	-0.0659	0.8563	1.0519	0.6146
$T\|Z \sim z_2(2U + 1)$	Proposed	-0.0012	0.9309	1.0301	0.6943
$C\|Z \sim z_2 + U$	Ying et al.	-0.2070	1.0325	1.1444	0.7464
$T\|Z \sim z_2(2U + 1)$	Proposed	0.0395	0.7982	0.9692	0.5539
$C\|Z \sim N(z_2, 0.5)$	Ying et al.	0.0505	0.4245	0.9853	0.2686
$T\|Z \sim z_2 + U$	Proposed	-0.0328	0.3057	1.0066	0.2075
$C\|Z \sim N(z_2, 0.5)$	Ying et al.	0.0505	0.4245	0.9853	0.2686

Finally we considered examples 1 and 2 of Table 4, but with higher censoring rates, see Table 5. The first example has censoring rate about 55%, the second about 75%. In both cases, the YJW estimator has lower RMSE than the proposed estimator. Note, however, that the YJW estimator has a much higher bias than the proposed estimator in the second example. It appears that neither estimator performs well under high censoring rates.

Table 5. Comparison between the proposed estimator and YJW's estimator. Conditional distributions of T given Z and of C given Z are both misspecified. The true value of the intercept is 0, and the true value of the slope is 1. Sample size is 100. Censoring rate is about 55% for example 1, and about 75% for example 2.

Simulation model	*Estimator*	*Intercept*		*Slope*	
		Mean	RMSE	Mean	RMSE
$T\|Z \sim z_2 + U$	Proposed	-0.0432	0.3655	1.0576	0.2828
$\lambda_c(t\|z_2) = 0.5m(z_2)$	Ying et al.	-0.0152	0.4146	1.0113	0.2768
$T\|Z \sim z_2(2U+1)$	Proposed	0.3905	1.2056	0.7579	0.8841
$\lambda_c(t\|z_2) = 1.33m(z_2)$	Ying et al.	0.4058	1.1534	0.5793	0.8571

4 Large sample results

We shall assume that the true value β_0 of β is in the interior of a bounded convex region D. In addition, we will need the following conditions (cf. YJW, Appendix A):

1. The covariate vector Z is bounded, say $||Z|| \leq L$, where $||\cdot||$ is Euclidean norm.

2. For $\beta \in D$, there exists a constant t_1 such that $P(X \geq t_1|Z) > 0$ and $\beta' Z \leq t_1$, with probability 1.

3. The derivatives $f(t,z)$ and $g(t,z)$ respectively, of $-F(t,z)$ and $-G(t,z)$, with respect to t are uniformly bounded in $(t,z) \in (-\infty, t_1] \times [-L, L]$. The conditional survival function of T, namely $F(t,z)$, is bounded away from zero over the region $(-\infty, t_1] \times [-L, L]$.

4. The matrix $E[ZZ'f(0|Z)]$ is positive definite, where $f(t|z) = f(t + \beta_0' z, z)$ is the conditional density of $\epsilon = T - \beta_0' Z$ given $Z = z$.

For simplicity we only consider the large sample properties of $\hat{\beta}$ in the case that Z takes finitely many values. Denote the possible values of Z by z_k, $k = 1, \ldots, K$, and assume each occurs with positive probability. Let n_k denote the number of $Z_j, j = 1, \ldots, n$, taking the value z_k. Rewrite the sample $(X_i, \delta_i, Z_i), i = 1, \ldots, n$ as $(X_{j,k}, \delta_{j,k}), j = 1, \ldots, n_k$, for $k = 1, \ldots, K$, where $(X_{j,k}, \delta_{j,k})$ corresponds to (X_i, δ_i) with covariate Z_i having the value z_k. Let $\hat{F}(t, z_k)$ be the (local) Kaplan–Meier estimator based on the pairs $(X_{j,k}, \delta_{j,k})$, $j = 1, \ldots, n_k$.

We will find it convenient to introduce some counting process notation. Let $N_{jk}^u(t) = I(X_{jk} \leq t, \delta_{jk} = 1), N_{jk}^c(t) = I(X_{jk} \leq t, \delta_{jk} = 0)$, $Y_{jk}(t) = I(X_{jk} \geq t)$, $M_{jk}^u(t) = N_{jk}^u(t) - \int_0^t \lambda_k^u(s) Y_{jk}(s)\, ds$, and $M_{jk}^c(t) = N_{jk}^c(t) - \int_0^t \lambda_k^c(s) Y_{jk}(s)\, ds$, where $\lambda_k^u(s)$ and $\lambda_k^c(s)$ are the hazard functions of T_{jk} and C_{jk}, respectively. Let $M_{\cdot k}^u = \sum_{j=1}^{n_k} M_{jk}^u$, $M_{\cdot k}^c = \sum_{j=1}^{n_k} M_{jk}^c$, $Y_{\cdot k} = \sum_{j=1}^{n_k} Y_{jk}$. Note that $Y_{\cdot k}(t)/n_k \xrightarrow{P} y_k(t) \equiv P(X_i \geq t | Z_i = z_k)$, uniformly in $t \geq 0$. Also, denote $H(t, z) = 1 - G(t, z)$.

4.1 Consistency and asymptotic normality

Theorem 1 *Under conditions 1–4, the estimator $\hat{\beta}$ is consistent and asymptotically normal: $n^{1/2}\left(\hat{\beta} - \beta_0\right) \xrightarrow{D} N\left(0, A^{-1}\Gamma(A^{-1})\right)$, where $A = -E[ZZ'f(0|Z)]$ and Γ is the covariance matrix of τ_i given in (A.14).*

The proof of Theorem 1, given in the Appendix, is based on YJW's approach in proving asymptotic normality of their estimator. Note that the estimating function $S_n(\beta)$ is discontinuous and non-monotone. As such, neither the usual Taylor expansion method, often used in analyzing a maximum likelihood estimator, nor a contiguity argument can be used. Therefore, for proving the asymptotic normality, we first establish local linearity of S_n in the $n^{-1/3}$ neighborhood of β_0. Next we show asymptotic normality of $n^{-1/2}S_n(\beta_0)$. The asymptotic normality of $\hat{\beta}$ then follows from these two results.

4.2 Asymptotic equivalence of the proposed and YJW estimators

From Theorem 1, we see that the asymptotic covariance matrix of $\sqrt{n}(\hat{\beta} - \beta_0)$ is $A^{-1}\Gamma A^{-1}$. In this Section, we further show that YJW's estimator $\hat{\beta}^*$, defined as a minimizer of $||S_n^*(\beta)||$, is asymptotically equivalent to our estimator based on the MIP. We state this result in the following theorem.

Theorem 2 *Under conditions 1–4, $\hat{\beta}$ and $\hat{\beta}^*$ are asymptotically equivalent: $||\hat{\beta} - \hat{\beta}^*|| = o_p(n^{-1/2})$.*

Proof. Using the local linearity of $n^{-1/2}S_n(\beta_0)$ (cf. Appendix A.2) and $n^{-1/2}S_n^*(\beta_0)$ (cf. Appendix C of YJW), and the facts $n^{-1/2}S_n(\hat{\beta}) = o_p(1)$, and $n^{-1/2}S_n^*(\hat{\beta}^*) = o_p(1)$, we have

$$\sqrt{n}A(\hat{\beta} - \beta_0) = -n^{-1/2}S_n(\beta_0) + o_p(1)$$
$$\sqrt{n}A(\hat{\beta}^* - \beta_0) = -n^{-1/2}S_n^*(\beta_0) + o_p(1),$$

where the matrix $A = -E[ZZ'f(0|Z)]$ is the same under both estimation procedures. It is sufficient to show that $n^{-1/2}(S_n(\beta_0) - S_n^*(\beta_0)) \xrightarrow{P} 0$ as $n \to \infty$.

First, we give some useful identities for completing the proof. Note that $\lambda_k^u + \lambda_k^c$ is the hazard function of $X_{jk} = \min(T_{jk}, C_{jk})$. It follows that

$$\int_{-\infty}^{\beta_0' z_k} \frac{\lambda_k^u(s)}{y_k(s)}\, ds + \int_{-\infty}^{\beta_0' z_k} \frac{\lambda_k^c(s)}{y_k(s)}\, ds = \frac{1 - G(\beta_0' z_k, z_k)/2}{G(\beta_0' z_k, z_k)/2}. \tag{4.1}$$

From

$$\int_{-\infty}^{\beta_0' z_k} \frac{dM_{jk}^c(s)}{y_k(s)} = \int_{-\infty}^{\beta_0' z_k} \frac{dN_{jk}^c(s)}{y_k(s)} - \int_{-\infty}^{\beta_0' z_k} \frac{\lambda_k^c(s) I(X_{jk} \geq s)}{y_k(s)} ds$$
$$= \frac{I(X_{jk} < \beta_0' z_k, \delta_{jk} = 0)}{y_k(X_{jk})} - \int_{-\infty}^{\beta_0' z_k \wedge X_{jk}} \frac{\lambda_k^c(s)}{y_k(s)} ds,$$

we have,

$$\int_{-\infty}^{\beta_0' z_k} \frac{dM_{jk}^u(s) + dM_{jk}^c(s)}{y_k(s)} = \frac{I(X_{jk} < \beta_0' z_k)}{y_k(X_{jk})} - \int_{-\infty}^{\beta_0' z_k \wedge X_{jk}} \frac{\lambda_k^u(s) + \lambda_k^c(s)}{y_k(s)} ds$$
$$= \frac{I(X_{jk} < \beta_0' z_k)}{y_k(X_{jk})} - \frac{1 - y_k(\beta_0' z_k \wedge X_{jk})}{y_k(\beta_0' z_k \wedge X_{jk})}$$
$$= \begin{cases} 1 & \text{if } X_{jk} < \beta_0' z_k \\ -\frac{1 - G(\beta_0' z_k, z_k)/2}{G(\beta_0' z_k, z_k)/2} & \text{if } X_{jk} \geq \beta_0' z_k \end{cases}$$
$$= 1 - \frac{I(X_{jk} \geq \beta_0' z_k)}{G(\beta_0' z_k, z_k)/2}. \tag{4.2}$$

From (2.3), we note that

$$S_n^*(\beta_0) = \sum_{k=1}^{K} \sum_{j=1}^{n_k} z_k \Big[\frac{I(X_{jk} \geq \beta_0' z_k)}{G(\beta_0' z_k, z_k)} - \frac{1}{2} - \frac{\hat{G}(\beta_0' z_k, z_k) - G(\beta_0' z_k, z_k)}{\hat{G}(\beta_0' z_k, z_k) G(\beta_0' z_k, z_k)} P(X_{jk} \geq \beta_0' z_k)$$
$$+ \frac{\hat{G}(\beta_0' z_k, z_k) - G(\beta_0' z_k, z_k)}{\hat{G}(\beta_0' z_k, z_k) G(\beta_0' z_k, z_k)} (P(X_{jk} \geq \beta_0' z_k) - I(X_{jk} \geq \beta_0' z_k)) \Big],$$

and as $n \to \infty$,

$$n^{-1/2} \sum_{k=1}^{K} z_k \frac{\hat{G}(\beta_0' z_k, z_k) - G(\beta_0' z_k, z_k)}{\hat{G}(\beta_0' z_k, z_k) G(\beta_0' z_k, z_k)} \sum_{j=1}^{n_k} (P(X_{jk} \geq \beta_0' z_k) - I(X_{jk} \geq \beta_0' z_k)) \xrightarrow{P} 0.$$

Using a martingale representation for $(\hat{G} - G)/\hat{G}$ (cf. Gill (1980)), we have

$$S_n^*(\beta_0) = \sum_{k=1}^{K} \sum_{j=1}^{n_k} z_k \Big[\frac{I(X_{jk} \geq \beta_0' z_k)}{G(\beta_0' z_k, z_k)} - \frac{1}{2} - \frac{\hat{G}(\beta_0' z_k, z_k) - G(\beta_0' z_k, z_k)}{\hat{G}(\beta_0' z_k, z_k) G(\beta_0' z_k, z_k)} P(X_{jk} \geq \beta_0' z_k) \Big] + o_p(n^{1/2})$$
$$= \sum_{k=1}^{K} \Big\{ \sum_{j=1}^{n_k} z_k \Big[\frac{I(X_{jk} \geq \beta_0' z_k)}{G(\beta_0' z_k, z_k)} - \frac{1}{2} \Big] + \frac{1}{2} z_k \int_{-\infty}^{\beta_0' z_k} \frac{dM_{\cdot k}^c(s)}{y_k(s)} \Big\} + o_p(n^{1/2})$$
$$= -\sum_{k=1}^{K} \frac{1}{2} z_k \int_{-\infty}^{\beta_0' z_k} \frac{dM_{\cdot k}^u(s)}{y_k(s)} + o_p(n^{1/2}), \tag{4.3}$$

where the last equality is obtained by using (4.2).

Now, from (4.3) and (A.13), we have

$$n^{-1/2}(S_n(\beta_0) - S_n^*(\beta_0)) = n^{-1/2}\sum_{k=1}^{K}\sum_{j=1}^{n_k}\Big\{\Big[I(X_{j,k} \ge \beta_0' z_k) + \frac{I(X_{j,k} < \beta_0' z_k, \delta_{j,k} = 0)}{2F(X_{j,k}, z_k)} - \frac{1}{2}\Big] z_k + \frac{1}{2} z_k \int_{-\infty}^{\beta_0' z_k} \frac{G(s, z_k)}{y_k(s)}\, dM_{jk}^u(s)\Big\} + o_p(1).$$

Noting that $G(s,z_k)/y_k(s) = 1/F(s,z_k)$, we have

$$\begin{aligned}
\int_{-\infty}^{\beta_0' z_k} \frac{G(s,z_k)}{y_k(s)}\, dM_{jk}^u(s) &= \frac{I(X_{j,k} < \beta_0' z_k, \delta_{j,k} = 1)}{F(X_{j,k}, z_k)} - \int_{-\infty}^{\beta_0' z_k \wedge X_{j,k}} \frac{\lambda_k^u(s)}{F(s,z_k)}\, ds \\
&= \frac{I(X_{j,k} < \beta_0' z_k, \delta_{j,k} = 1)}{F(X_{j,k}, z_k)} - \Big(\frac{1}{F((\beta_0' z_k) \wedge X_{j,k}, z_k)} - 1\Big) \\
&= \begin{cases} -2 + 1 & \text{if } X_{jk} \ge \beta_0' z_k \\ \frac{I(X_{j,k} < \beta_0' z_k, \delta_{j,k} = 1)}{F(X_{j,k}, z_k)} - \frac{1}{F(X_{j,k}, z_k)} + 1 & \text{if } X_{jk} < \beta_0' z_k \end{cases} \\
&= \begin{cases} -1 & \text{if } X_{jk} \ge \beta_0' z_k \\ -\frac{I(X_{j,k} < \beta_0' z_k, \delta_{j,k} = 0)}{F(X_{j,k}, z_k)} + 1 & \text{if } X_{jk} < \beta_0' z_k \end{cases}.
\end{aligned}$$

Thus, $n^{-1/2}(S_n(\beta_0) - S_n^*(\beta_0)) = o_p(1)$ and $\hat\beta - \hat\beta^* = o_p(n^{-1/2})$. The two estimators are asymptotically equivalent. □

4.3 Some remarks

For the discrete case, let $a_k = P(Z = z_k) = \lim_{n\to\infty}(n_k/n)$. Then, $A = -\sum_{k=1}^K a_k z_k z_k' f(0|z_k)$. From (4.3), the asymptotic covariance matrix of $n^{-1/2} S_n^*(\beta_0)$ is

$$\sum_{k=1}^{K} a_k \frac{z_k z_k'}{4} \int_{-\infty}^{\beta_0' z_k} \frac{\lambda_k^u(s)}{y_k(s)}\, ds. \tag{4.4}$$

Thus, the asymptotic covariance matrix of $\sqrt{n}(\hat\beta^* - \beta_0)$ is given by

$$\Big(\sum_{k=1}^{K} a_k z_k z_k' f(0|z_k)\Big)^{-1}\Big(\sum_{k=1}^{K} a_k \frac{z_k z_k'}{4}\int_{-\infty}^{\beta_0' z_k} \frac{\lambda_k^u(s)}{y_k(s)}\, ds\Big)\Big(\sum_{k=1}^{K} a_k z_k z_k' f(0|z_k)\Big)^{-1}, \tag{4.5}$$

which, by the asymptotic equivalence of $\hat\beta$ to $\hat\beta^*$, is also the asymptotic covariance matrix of $\sqrt{n}(\hat\beta - \beta_0)$.

In the one sample case, i.e., $K = 1$, the expression given by (4.4) is equal to

$$\frac{z_1^2}{4}\int_{-\infty}^{\beta_0 z_1} \frac{\lambda_1^u(s)}{y(s)}\, ds.$$

Thus, the asymptotic variance of $\sqrt{n}(\hat{\beta}-\beta_0)$ and $\sqrt{n}(\hat{\beta}^*-\beta_0)$ is equal to

$$\frac{\Gamma}{A^2} = \frac{1}{4z_1^2 f^2(0|z_1)} \int_{-\infty}^{\beta_0 z_1} \frac{\lambda_1^u(s)}{y(s)}\, ds,$$

which is the well-known result on the asymptotic variance of the median divided by z_1^2; see for example, Doss and Gill (1992).

Finally, for the independent case where censoring distribution does not depend on covariates, we point out that YJW's formula (A.3) and the first display following it have a wrong sign. We now make this point for the discrete case. Let $Y_i(t) = I(X_i \geq t)$, $M^c(t) = \sum_{i=1}^n (I(X_i \leq t, \delta_i = 0) - \int_0^t \lambda^c(s) Y_i(s)\, ds)$, $Y(t) = \sum_{i=1}^n Y_i(t)$ and $y(t) = \lim_{n\to\infty} Y(t)/n$, and let $\hat{G}, G, Q$ and q be defined as in YJW's paper. Note that the following martingale representation for $\hat{G}-G$ holds:

$$\hat{G}(t) - G(t) = -G(t) \int_{-\infty}^{t} \frac{dM^c(s)}{Y(s)} + o_p(n^{-1/2}).$$

We have

$$\begin{aligned} n\int_{-\infty}^{\infty} \frac{\hat{G}(t)-G(t)}{\hat{G}(t)G(t)}\, dQ(t) &= -n\int_{-\infty}^{\infty} \frac{1}{\hat{G}(t)} \int_{-\infty}^{t} \frac{dM^c(s)}{Y(s)}\, dQ(t) + o_p(n^{1/2}) \\ &= -n\int_{-\infty}^{\infty} \frac{1}{Y(s)} \int_{s}^{\infty} \frac{dQ(t)}{\hat{G}(t)}\, dM^c(s) + o_p(n^{1/2}). \end{aligned}$$

Since

$$\begin{aligned} \int_s^{\infty} \frac{dQ(t)}{\hat{G}(t)} &= n^{-1}\sum_{i=1}^n \frac{I(s \leq \beta_0 Z_i \leq X_i) Z_i}{\hat{G}(\beta_0 Z_i)} = n^{-1}\sum_{i=1}^n \frac{I(s \leq \beta_0 Z_i \leq X_i) Z_i}{G(\beta_0 Z_i)} + o_p(1) \\ &\xrightarrow{P} \frac{1}{2} E[I(\beta_0 Z_i \geq s) Z_i] \equiv \frac{1}{2} q(s), \end{aligned}$$

we have

$$n\int_{-\infty}^{\infty} \frac{\hat{G}(t)-G(t)}{\hat{G}(t)G(t)}\, dQ(t) = -\frac{1}{2}\int_{-\infty}^{\infty} \frac{q(s)}{y(s)}\, dM^c(s) + o_p(n^{1/2}),$$

and

$$n^{-1/2} S_n(\beta_0) = n^{-1/2} \sum_{i=1}^n \left[\frac{I(X_i \geq \beta_0 Z_i)}{G(\beta_0 Z_i)} - \frac{1}{2}\right] Z_i + \frac{1}{2} n^{-1/2} \int_{-\infty}^{\infty} \frac{q(s)}{y(s)}\, dM^c(s) + o_p(1).$$

Appendix

A.1 Consistency of $\hat{\beta}$

Define

$$\tilde{S}_n(\beta) = \sum_{i=1}^n \left(F(\beta' Z_i, Z_i) - 1/2\right) Z_i. \tag{A.1}$$

It is easy to see that

$$\tilde{S}_n(\beta) = \sum_{i=1}^{n} \Big\{ P(X_i \geq \beta' Z_i | Z_i) + P(C_i < \beta' Z_i | Z_i) F(\beta' Z_i, Z_i) - 1/2 \Big\} Z_i.$$

Following the arguments in YJW (Appendix A), we first establish that $\sup_{\beta \in D} ||n^{-1} S_n(\beta) - n^{-1}\tilde{S}_n(\beta)|| \xrightarrow{\text{a.s.}} 0$. We may write

$$\begin{aligned} S_n(\beta) - \tilde{S}_n(\beta) &= \sum_{i=1}^{n} \left(I\left(X_i \geq \beta' Z_i\right) - P(X_i \geq \beta' Z_i | Z_i) \right) Z_i \\ &\quad + \sum_{i=1}^{n} \Big(I(X_i < \beta' Z_i, \delta_i = 0) \frac{\hat{F}(\beta' Z_i, Z_i)}{\hat{F}(X_i, Z_i)} \\ &\quad - P(C_i < \beta' Z_i | Z_i) F(\beta' Z_i, Z_i) \Big) Z_i. \end{aligned}$$

By some elementary probability arguments, the first summation above is $o(n^{1/2+\epsilon})$ a.s., as $n \to \infty$, uniformly in $\beta \in D$, for every $\epsilon > 0$.

With probability 1, the second summation is equal to

$$\begin{aligned} &\sum_{i=1}^{n} \frac{I(X_i < \beta' Z_i, \delta_i = 0)}{\hat{F}(X_i, Z_i)} (\hat{F}(\beta' Z_i, Z_i) - F(\beta' Z_i, Z_i)) Z_i \\ &+ \sum_{i=1}^{n} F(\beta' Z_i, Z_i) \Big[\frac{I(X_i < \beta' Z_i, \delta_i = 0)}{\hat{F}(X_i, Z_i)} - P(C_i < \beta' Z_i | Z_i) \Big] Z_i \\ &= \sum_{i=1}^{n} \frac{I(X_i < \beta' Z_i, \delta_i = 0)}{\hat{F}(X_i, Z_i)} (\hat{F}(\beta' Z_i, Z_i) - F(\beta' Z_i, Z_i)) Z_i \\ &+ \sum_{i=1}^{n} F(\beta' Z_i, Z_i) I(X_i < \beta' Z_i, \delta_i = 0) \Big[\frac{1}{\hat{F}(X_i, Z_i)} - \frac{1}{F(X_i, Z_i)} \Big] Z_i \\ &+ \sum_{i=1}^{n} F(\beta' Z_i, Z_i) \Big[\frac{I(X_i < \beta' Z_i, \delta_i = 0)}{F(X_i, Z_i)} - P(C_i < \beta' Z_i | Z_i) \Big] Z_i. \end{aligned}$$

The first two summations are both $o(n^{1/2+\epsilon})$ a.s., as $n \to \infty$, uniformly in $\beta \in D$. For the third term, note that

$$\begin{aligned} &E\Big\{ \frac{I(X < \beta' Z, \delta = 0)}{F(X, Z)} \Big| Z \Big\} = E\Big\{ \frac{I(C < \beta' Z, \delta = 0)}{F(C, Z)} \Big| Z \Big\} \\ &= E\Big\{ \frac{I(C < \beta' Z, C < T)}{F(C, Z)} \Big| Z \Big\} = \int_{-\infty}^{\beta' Z} \frac{P(T > c | Z)}{F(c, Z)} \, dH(c, Z) \\ &= P(C < \beta' Z | Z), \end{aligned} \tag{A.2}$$

where $H(t, z) = P(C \leq t | z)$ and the intergration is with respect to c. Thus, the third term is $o(n^{1/2+\epsilon})$ a.s., as $n \to \infty$, uniformly for $\beta \in D$. Hence,

$$\sup_{\beta \in D} ||n^{-1} S_n(\beta) - n^{-1} \tilde{S}_n(\beta)|| \overset{\text{a.s.}}{=} o(n^{-1/2+\epsilon}) \xrightarrow{\text{a.s.}} 0. \tag{A.3}$$

Since the matrix $E[ZZ'f(0|Z)]$ is positive definite, we have

$$A_n(\beta) \equiv n^{-1}\frac{\partial}{\partial\beta}\tilde{S}_n(\beta) = -n^{-1}\sum_{i=1}^{n} f(\beta' Z_i, Z_i)Z_iZ_i' = -n^{-1}\sum_{i=1}^{n} f((\beta-\beta_0)'Z_i|Z_i)Z_iZ_i' \quad \text{(A.4)}$$

is nonpositive definite and, with probability 1, $A_n(\beta_0) \to -E[ZZ'f(0|Z)]$, which is negative definite. Because $\tilde{S}_n(\beta_0) = 0$, it follows that the sequence $n^{-1}\tilde{S}_n(\beta)$ is bounded away from zero a.s. for any $\beta \neq \beta_0$. From (A.3), we have $n^{-1}\tilde{S}_n(\hat{\beta}) \xrightarrow{\text{a.s.}} 0$. Hence, $\hat{\beta} \xrightarrow{\text{a.s.}} \beta_0$ as $n \to \infty$. Furthermore, since $n^{-1}\|\tilde{S}_n(\hat{\beta})\| = o(n^{-1/2+\epsilon})$, from the expansion $n^{-1}\tilde{S}_n(\hat{\beta}) - n^{-1}\tilde{S}_n(\beta_0) = A_n(\tilde{\beta})(\hat{\beta}-\beta_0)$, where $\tilde{\beta}$ is on the line segment between $\hat{\beta}$ and β_0, we have $\|\hat{\beta}-\beta_0\| = o(n^{-1/2+\epsilon})$. □

A.2 Local linearity of $S_n(\beta)$

Let $A = -E[ZZ'f(0|Z)]$. As we have seen in the previous subsection, A is the limit of $A_n(\beta_0)$. Local linearity for $S_n(\beta)$ means that for any fixed constant c and all β in $\|\beta-\beta_0\| < cn^{-1/3}$,

$$S_n(\beta) = S_n(\beta_0) + nA(\beta-\beta_0) + o_p\left(\max(n^{1/2}, n\|\beta-\beta_0\|)\right). \quad \text{(A.5)}$$

Asymptotic normality of $\hat{\beta}$ in any $n^{-1/3}$-neighborhood of β_0 follows from (A.5) and asymptotic normality of $n^{-1/2}S_n(\beta_0)$ (shown in the next subsection).

$$\begin{aligned} S_n(\beta) - S_n(\beta_0) &= \sum_{i=1}^{n}\left(I\left(X_i \geq \beta' Z_i\right) - I\left(X_i \geq \beta_0' Z_i\right)\right)Z_i \\ &+ \sum_{i=1}^{n}\left[I(X_i < \beta' Z_i, \delta_i = 0)\frac{\hat{F}(\beta' Z_i, Z_i)}{\hat{F}(X_i, Z_i)} - I(X_i < \beta_0' Z_i, \delta_i = 0)\frac{\hat{F}(\beta_0' Z_i, Z_i)}{\hat{F}(X_i, Z_i)}\right] Z_i. \quad \text{(A.6)} \end{aligned}$$

The second summation in (A.6) is equal to

$$\begin{aligned} &\sum_{i=1}^{n}\frac{Z_i}{\hat{F}(X_i, Z_i)}\Big\{I(X_i < \beta' Z_i, \delta_i = 0)[\hat{F}(\beta' Z_i, Z_i) - F(\beta' Z_i, Z_i)] \\ &\qquad -I(X_i < \beta_0' Z_i, \delta_i = 0)[\hat{F}(\beta_0' Z_i, Z_i) - F(\beta_0' Z_i, Z_i)] \\ &\qquad +I(X_i < \beta' Z_i, \delta_i = 0)F(\beta' Z_i, Z_i) - I(X_i < \beta_0' Z_i, \delta_i = 0)F(\beta_0' Z_i, Z_i)\Big\} \\ = &\sum_{i=1}^{n}\Big\{\frac{Z_i}{\hat{F}(X_i, Z_i)}I(X_i < \beta' Z_i, \delta_i = 0) \\ &\times\left[(\hat{F}(\beta' Z_i, Z_i) - F(\beta' Z_i, Z_i)) - (\hat{F}(\beta_0' Z_i, Z_i) - F(\beta_0' Z_i, Z_i))\right]\Big\} \quad \text{(A.7)} \\ &+\sum_{i=1}^{n}\Big\{\frac{Z_i}{\hat{F}(X_i, Z_i)}[I(X_i < \beta' Z_i, \delta_i = 0) - I(X_i < \beta_0' Z_i, \delta_i = 0)] \\ &\times\left[\hat{F}(\beta_0' Z_i, Z_i) - F(\beta_0' Z_i, Z_i)\right]\Big\} \quad \text{(A.8)} \end{aligned}$$

$$+\sum_{i=1}^{n}\Big\{\frac{Z_i}{\hat{F}(X_i,Z_i)}$$

$$\times\left[I(X_i<\beta' Z_i,\delta_i=0)F(\beta' Z_i,Z_i)-I(X_i<\beta_0' Z_i,\delta_i=0)F(\beta_0' Z_i,Z_i)\right]\Big\}. \tag{A.9}$$

By Lemma 1 of YJW, under conditions 1–3, the summation (A.7) is equal to $o_p(n^{1/2})$, for $||\beta-\beta_0||<cn^{-1/3}$.

The summation (A.8) is bounded by

$$\begin{aligned}&o_p(n^{-1/2+\epsilon})\sum_{i=1}^{n}|I(X_i<\beta' Z_i,\delta_i=0)-I(X_i<\beta_0' Z_i,\delta_i=0)|\\&=o_p(n^{-1/2+\epsilon})O_p(n^{2/3}),\ \text{for } ||\beta-\beta_0||<cn^{-1/3} \text{ by Lemma 2 of YJW}\\&=o_p(n^{1/2}).\end{aligned}$$

The summation (A.9) is equal to

$$\begin{aligned}&\sum_{i=1}^{n}Z_i\Big\{\frac{1}{\hat{F}(X_i,Z_i)}-\frac{1}{F(X_i,Z_i)}\Big\}\\&\Big[I(X_i<\beta' Z_i,\delta_i=0)F(\beta' Z_i,Z_i)-I(X_i<\beta_0' Z_i,\delta_i=0)F(\beta_0' Z_i,Z_i)\Big]\\&+\sum_{i=1}^{n}\frac{Z_i}{F(X_i,Z_i)}\Big[I(X_i<\beta' Z_i,\delta_i=0)F(\beta' Z_i,Z_i)-I(X_i<\beta_0' Z_i,\delta_i=0)F(\beta_0' Z_i,Z_i)\Big].\end{aligned}$$

Here, the first term is $o_p(n^{1/2})$ by conditions 1–3 and Lemma 2 of YJW. Hence,

$$\begin{aligned}&S_n(\beta)-S_n(\beta_0)=\sum_{i=1}^{n}[I\,(X_i\ge\beta' Z_i)-I\,(X_i\ge\beta_0' Z_i)]Z_i+o_p(n^{1/2})\\&+\sum_{i=1}^{n}\frac{Z_i}{F(X_i,Z_i)}[I(X_i<\beta' Z_i,\delta_i=0)F(\beta' Z_i,Z_i)-I(X_i<\beta_0' Z_i,\delta_i=0)F(\beta_0' Z_i,Z_i)]\\=&\sum_{i=1}^{n}\Big\{I\,(X_i\ge\beta' Z_i)+\frac{1}{F(X_i,Z_i)}I(X_i<\beta' Z_i,\delta_i=0)F(\beta' Z_i,Z_i)-\frac{1}{2}\Big\}Z_i \qquad (A.10)\\&-\sum_{i=1}^{n}\Big\{I\,(X_i\ge\beta_0' Z_i)+\frac{1}{F(X_i,Z_i)}I(X_i<\beta_0' Z_i,\delta_i=0)F(\beta_0' Z_i,Z_i)-\frac{1}{2}\Big\}Z_i+o_p(n^{1/2}).\end{aligned}$$

By (A.2), the conditional expectation of the term in the second summation in the decomposition (A.10) given Z_i is equal to

$$\begin{aligned}&\Big\{P\,(X_i\ge\beta_0' Z_i|Z_i)+F(\beta_0' Z_i,Z_i)E\Big[\frac{I(X_i<\beta_0' Z_i,\delta_i=0)}{F(X_i,Z_i)}|Z_i\Big]-\frac{1}{2}\Big\}Z_i\\&=\Big\{P\,(X_i\ge\beta_0' Z_i|Z_i)+F(\beta_0' Z_i,Z_i)P(C_i<\beta_0' Z_i|Z_i)-\frac{1}{2}\Big\}Z_i=0.\end{aligned}$$

Subtracting and adding the conditional expectation of the term in the first summation in the decomposition (A.10) given Z_i and using (A.2), we have

$$S_n(\beta) - S_n(\beta_0) = \sum_{i=1}^{n} \Big\{ [I(X_i \geq \beta' Z_i) - P(X_i \geq \beta' Z_i | Z_i)]$$

$$+ F(\beta' Z_i, Z_i) \Big[\frac{I(X_i < \beta' Z_i, \delta_i = 0)}{F(X_i, Z_i)} - P(C_i < \beta' Z_i | Z_i) \Big] \Big\} Z_i$$

$$- \sum_{i=1}^{n} \Big\{ [I(X_i \geq \beta_0' Z_i) - P(X_i \geq \beta_0' Z_i | Z_i)]$$

$$+ F(\beta_0' Z_i, Z_i) \Big[\frac{I(X_i < \beta_0' Z_i, \delta_i = 0)}{F(X_i, Z_i)} - P(C_i < \beta_0' Z_i | Z_i) \Big] \Big\} Z_i$$

$$+ \sum_{i=1}^{n} \Big\{ P(X_i \geq \beta' Z_i | Z_i) + F(\beta' Z_i, Z_i) P(C_i < \beta' Z_i | Z_i) - \frac{1}{2} \Big\} Z_i.$$

By conditions 1–3 and Lemma 2 of YJW, the first two summations are equal to $o_p(n^{1/2})$, for $||\beta - \beta_0|| < cn^{-1/3}$. So

$$S_n(\beta) - S_n(\beta_0) = \tilde{S}_n(\beta) + o_p(n^{1/2}). \qquad \text{(A.11)}$$

Let $A_n(\beta)$ be as given in (A.4). Taking Taylor's expansion of $\tilde{S}_n(\beta)$ at β_0 and applying the strong law of large numbers, we have

$$\tilde{S}_n(\beta) = nA_n(\beta_0)(\beta - \beta_0) + o_p(n||\beta - \beta_0||) = nA(\beta - \beta_0) + o_p(n||\beta - \beta_0||),$$

uniformly for $||\beta - \beta_0|| < cn^{-1/3}$. This, together with (A.11), implies (A.5). □

A.3 Asymptotic Normality of $\hat{\beta}$

We first show the asymptotic normality of $n^{-1/2} S_n(\beta_0)$. We may write $S_n(\beta_0) = \sum_{k=1}^{K} S_{n_k}(\beta_0)$, where

$$S_{n_k}(\beta_0) = \sum_{j=1}^{n_k} \left(\left\{ I(X_{j,k} \geq \beta_0' z_k) + \frac{I(X_{j,k} < \beta_0' z_k, \delta_{j,k} = 0) \hat{F}(\beta_0' z_k, z_k)}{\hat{F}(X_{j,k}, z_k)} \right\} - \frac{1}{2} \right) z_k.$$

Note that $G(t, z) = P(C > t | Z = z)$ and $H(t, z) = 1 - G(t, z)$. We give the following decomposition for $S_{n_k}(\beta_0)$:

$$S_{n_k}(\beta_0) = \sum_{j=1}^{n_k} \Big\{ (I(X_{j,k} \geq \beta_0' z_k) - G(\beta_0' z_k, z_k)/2) z_k$$

$$+ \Big(\frac{I(X_{j,k} < \beta_0' z_k, \delta_{j,k} = 0)}{\hat{F}(X_{j,k}, z_k)} - P(C_{j,k} < \beta_0' z_k | z_k) \Big) \hat{F}(\beta_0' z_k, z_k) z_k$$

$$+ (\hat{F}(\beta_0' z_k, z_k) - F(\beta_0' z_k, z_k)) P(C_{j,k} < \beta_0' z_k | z_k) z_k$$

$$
\begin{aligned}
= & \sum_{j=1}^{n_k}\Big\{(I(X_{j,k}\ge\beta_0'z_k)-G(\beta_0'z_k,z_k)/2)z_k \\
& +\Big(\frac{I(X_{j,k}<\beta_0'z_k,\delta_{j,k}=0)}{F(X_{j,k},z_k)}-P(C_{j,k}<\beta_0'z_k|z_k)\Big)\hat F(\beta_0'z_k,z_k)z_k \\
& +(\hat F(\beta_0'z_k,z_k)-F(\beta_0'z_k,z_k))P(C_{j,k}<\beta_0'z_k|z_k)z_k \\
& +\Big(\frac{1}{\hat F(X_{j,k},z_k)}-\frac{1}{F(X_{j,k},z_k)}\Big)I(X_{j,k}<\beta_0'z_k,\delta_{j,k}=0)\hat F(\beta_0'z_k,z_k)z_k\Big\}.
\end{aligned}
$$

By the well-known martingale representation for Kaplan-Meier estimator, we have

$$
\sqrt{n_k}(\hat F(t,z_k)-F(t,z_k))=-F(t,z_k)\int_{-\infty}^{t}{n_k}^{-1/2}\frac{dM_{\cdot k}^u(s)}{y_k(s)}+o_p(1). \qquad \text{(A.12)}
$$

Thus,

$$
\begin{aligned}
& \sum_{j=1}^{n_k}\Big(\frac{1}{\hat F(X_{j,k},z_k)}-\frac{1}{F(X_{j,k},z_k)}\Big)I(X_{j,k}<\beta_0'z_k,\delta_{j,k}=0) \\
& =-\sum_{j=1}^{n_k}\frac{\hat F(X_{j,k},z_k)-F(X_{j,k},z_k)}{\hat F(X_{j,k},z_k)F(X_{j,k},z_k)}I(X_{j,k}<\beta_0'z_k,\delta_{j,k}=0) \\
& =\sum_{j=1}^{n_k}\int_{-\infty}^{X_{j,k}}{n_k}^{-1}\frac{dM_{\cdot k}^u(s)}{y_k(s)}\frac{I(X_{j,k}<\beta_0'z_k,\delta_{j,k}=0)}{F(X_{j,k},z_k)}+o_p(n^{1/2}) \\
& =\sum_{j=1}^{n_k}\int_{-\infty}^{\infty}{n_k}^{-1}\frac{I(s\le X_{j,k})}{y_k(s)}\frac{I(X_{j,k}<\beta_0'z_k,\delta_{j,k}=0)}{F(X_{j,k},z_k)}\,dM_{\cdot k}^u(s)+o_p(n^{1/2}) \\
& =\int_{-\infty}^{\infty}{n_k}^{-1}\frac{1}{y_k(s)}\sum_{j=1}^{n_k}\frac{I(s\le X_{j,k})I(X_{j,k}<\beta_0'z_k,\delta_{j,k}=0)}{F(X_{j,k},z_k)}\,dM_{\cdot k}^u(s)+o_p(n^{1/2}) \\
& =\int_{-\infty}^{\beta_0'z_k}\frac{1}{y_k(s)}r(s,z_k)\,dM_{\cdot k}^u(s)+o_p(n^{1/2}),
\end{aligned}
$$

where, conditional on $Z=z_k$,

$$
\begin{aligned}
r(s,z_k) &= \lim_{n_k\to\infty}{n_k}^{-1}\sum_{j=1}^{n_k}\frac{I(s\le X_{j,k}<\beta_0'z_k,\delta_{j,k}=0)}{F(X_{j,k},z_k)} \\
&= E\Big(\frac{I(s\le X_{j,k}<\beta_0'z_k,\delta_{j,k}=0)}{F(X_{j,k},z_k)}\Big|z_k\Big) \\
&= E\Big(\frac{I(s\le C_{j,k}<\beta_0'z_k,C_{j,k}<T_{j,k})}{F(C_{j,k},z_k)}\Big|z_k\Big) \\
&= \int_s^{\beta_0'z_k}\frac{P(T_{j,k}>c|z_k)}{F(c,z_k)}\,dH(c,z_k) \\
&= H(\beta_0'z_k,z_k)-H(s,z_k).
\end{aligned}
$$

Now, applying the martingale representation to the third term of the decomposition for $S_{n_k}(\beta_0)$ and using the definition $H(\beta_0' z_k, z_k) = P(C_{j,k} < \beta_0' z_k | z_k)$, we have

$$\begin{aligned} S_{n_k}(\beta_0) &= \sum_{j=1}^{n_k} \Big\{ (I(X_{j,k} \geq \beta_0' z_k) - G(\beta_0' z_k, z_k)/2) z_k \\ &+ \Big(\frac{I(X_{j,k} < \beta_0' z_k, \delta_{j,k} = 0)}{F(X_{j,k}, z_k)} - H(\beta_0' z_k, z_k) \Big) F(\beta_0' z_k, z_k) z_k \Big\} \\ &- F(\beta_0' z_k, z_k) H(\beta_0' z_k, z_k) z_k \int_{-\infty}^{\beta_0' z_k} \frac{dM_{\cdot k}^u(s)}{y_k(s)} \\ &+ F(\beta_0' z_k, z_k) z_k \int_{-\infty}^{\beta_0' z_k} \frac{r(s, z_k)}{y_k(s)} dM_{\cdot k}^u(s) + o_p(n^{1/2}). \end{aligned}$$

Note that $F(\beta_0' z_k, z_k) = \frac{1}{2}$. We have

$$\begin{aligned} S_{n_k}(\beta_0) &= \sum_{j=1}^{n_k} \Big\{ \Big[I(X_{j,k} \geq \beta_0' z_k) + \frac{I(X_{j,k} < \beta_0' z_k, \delta_{j,k} = 0)}{2F(X_{j,k}, z_k)} - \frac{1}{2} \Big] z_k \\ &+ \frac{1}{2} z_k \int_{-\infty}^{\beta_0' z_k} \Big(\frac{r(s, z_k)}{y_k(s)} - \frac{H(\beta_0' z_k, z_k)}{y_k(s)} \Big) dM_{jk}^u(s) \Big\} + o_p(n^{1/2}) \\ &= \sum_{j=1}^{n_k} \Big\{ \Big[I(X_{j,k} \geq \beta_0' z_k) + \frac{I(X_{j,k} < \beta_0' z_k, \delta_{j,k} = 0)}{2F(X_{j,k}, z_k)} - \frac{1}{2} \Big] z_k \\ &- \frac{1}{2} z_k \int_{-\infty}^{\beta_0' z_k} \frac{H(s, z_k)}{y_k(s)} dM_{jk}^u(s) \Big\} + o_p(n^{1/2}). \end{aligned} \tag{A.13}$$

Let

$$\begin{aligned} \tau_i &= \Big[I(X_i \geq \beta_0' Z_i) + \frac{I(X_i < \beta_0' Z_i, \delta_i = 0)}{2F(X_i, Z_i)} - \frac{1}{2} \Big] Z_i \\ &- \frac{1}{2} Z_i \int_{-\infty}^{\beta_0' Z_i} \frac{H(s, Z_i)}{P(X_i \geq s | Z_i)} dM_i^u(s). \end{aligned} \tag{A.14}$$

Then

$$S_n(\beta_0) = \sum_{i=1}^{n} \tau_i + o_p(n^{1/2}). \tag{A.15}$$

This implies that $n^{-1/2} S_n(\beta_0)$ is asymptotic normal with mean zero and asymptotic covariance matrix $\Gamma = E(\tau_i \tau_i')$.

It follows from the local linearity property for $S_n(\beta)$ that $\sqrt{n}(\hat{\beta} - \beta_0)$ is asymptotically normal with mean zero and covariance matrix $A^{-1} \Gamma A^{-1}$. □

References

Andersen, P. K., Borgan, Ø., Gill, D. R., and Keiding, N. (1993). *Statistical Models Based on Counting Processes.* Springer-Verlag, New York.

Burr, D. and Doss, H. (1993). Confidence bands for the median survival time as a function of the covariates in the Cox model. *J. Amer. Statist. Assoc.* **88** 1330–1340.

Dabrowska, D. M. (1989). Uniform consistency of the kernel conditional Kaplan–Meier estimate. *Ann. Statist.* **17** 1157–1167.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. Ser. B* **39** 1–38.

Doss, H. and Gill, R. (1992). An elementary approach to weak convergence for quantile processes, with applications to censored survival data. *J. Amer. Statist. Assoc.* **87** 869–877.

Efron, B. (1967). Two sample problem with censored data. In: *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability* **IV** 831–853. Prentice-Hall, New York.

Gill, R. D. (1980). *Censoring and Stochastic Integrals.* Mathematisch Centrum, Amsterdam.

Gill, R. D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method (Part 1). *Scand. J. Statist.* **16** 97–128.

Groeneboom, P. and Wellner, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation.* Birkhauser Verlag, Basel.

Jung, S. (1996). Quasilikelihood for median regression models. *J. Amer. Statist. Assoc.* **91** 251–257.

Lai, T. L. and Ying, Z. (1994). A missing information principle and M-estimators in regression analysis with censored and truncated data. *Ann. Statist.* **22** 1222–1255.

Laird, N. M. (1985). Missing information principle. *Encycl. Statist. Sci.* **5** 548–552.

Meier, P. (1975). Estimation of a distribution function from incomplete observations. In: *Perspectives in Probability and Statistics*, pp. 67–87 (J. Gani, editor). Applied Probability Trust, Sheffield.

Newey, W. K. and Powell, J. L. (1990). Efficient estimation of linear and type I censored regression models under conditional quantile restrictions. *Econometric Theory* **6** 295–317.

Orchard, T. and Woodbury, M. A. (1972). A missing information principle: Theory and applications. *Proc. 6th Berkeley Symp. Math. Statist. Prob.* **1** 697–715.

Robins, J. M. (1996). Locally efficient median regression with random censoring and surrogate markers. *Lifetime Data: Models in Reliability and Survival Analysis*, 263–274.

Tsai, W. and Crowley, J (1985). A Large sample study of generalized maximum likelihood estimators from incomplete data via self-consistency (Corr:V18 P470). *Ann. Statist.* **13** 1317–1334.

Yang, S. (1999). Censored median regression using weighted empirical survival and hazard functions. *J. Amer. Statist. Assoc.* **94** 137–145.

Ying, Z., Jung, S. and Wei, L. J. (1995). Survival analysis with median regression models. *J. Amer. Statist. Assoc.* **90** 178–184.