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

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. For discrete covariates 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 [3]; for AFT models, it is not possible to estimate the intercept parameter effectively in the presence of right censoring [4]. 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 [2], henceforth YJW, who proposed an estimator for the regression parameter β under the assumption that C is independent of Z, and an extended version of the estimator that applies to the full model (although restricted to low-dimensional covariates in practice). In earlier work, Newey and Powell [5] found an efficient estimator for β in the special case in which C is always observed. Robins [6] developed an extension of the model to allow surrogate marker information, and constructed locally efficient estimators under the assumptions of (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. For uncensored data, Jung [7] obtained an efficient estimating function for the median regression parameters based on quasi-likelihood. Recently, Yang [8] 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. This is a straightforward application of the well-known missing information principle ([9], [10]) which provides a strategy for constructing estimators in general missing data problems. When applied to a parametric score function, the missing information principle (MIP) amounts to a single iteration of the EM-algorithm [11].

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 [1] 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 ([12], [13]). 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 [14]. The MIP has been discussed in the context of the accelerated failure time model [15].

In the present setting, application of the MIP leads to an estimating equation involving the conditional distribution function of T given Z. Surprizingly, the resulting estimating equation differs from that of YJW, which involves the conditional distribution function of C given Z. Both approaches, however, suffer from the curse of dimensionality due to the difficulty of estimating the conditional distributions nonparametrically for high-dimensional covariates. Indeed, this was to be expected in view of results [16] suggesting that *no* reasonably performing estimator of β exists when Z is high dimensional.

What is the best compromise to avoid the curse of dimensionality? One possibility is to make an independence assumption, as required for consistency of the special estimators of YJW and Yang [8]. However, these independence assumptions are unduly restrictive in typical applications (e.g., independence of C from Z is rare in clinical trials). A more flexible approach is to fit a semiparametric regression model (e.g., Cox PH model) that can provide a relatively good approximation for the unknown conditional distribution in the particular estimating equation. In the case of the MIP estimating equation, a suitable semiparametric model for the conditional distribution of T given Z is often already available, so no extra modeling effort is required. For the YJW estimating equation, however, extra work would be required to find a suitable semiparametric model for the conditional distribution of C

given Z.

Which of the MIP and YJW estimating equations results in better estimator performance? We compare the two approaches via simulation and find that our estimator tends to do better when Cox PH models are used to estimate the conditional distributions. In terms of large sample behavior, we show that the two estimators are asymptotically equivalent in the case of discrete covariates when the fully nonparametric approach is used. The bias induced by the inherent misspecifications of the Cox PH models is difficult to determine theoretically, but our simulation results suggest that it is only a serious problem under high censoring rates.

The paper is organized as follows. Background on the YJW and MIP estimating equations is provided in Section 2. The simulation results are presented in Section 3, and the asymptotic results in Section 4. Most of the proofs are placed in the Appendix.

2 Estimating equations

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:

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

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. [1], p. 840, equation (7.4))

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

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

The MIP 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 [17] kernel conditional Kaplan–Meier estimator can be used to estimate $F(t,z)$ nonparametrically. 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 $S_n(\beta)$.

As discussed in the Introduction, 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. As a compromise, we recommend using a semiparametric regression model to estimate F. Provided the semiparametric model offers a reasonably good fit to the data, the inherent misspecification should not be a serious problem. A plethora of semiparametric models are available for use in this regard, and good-fitting semiparametric models have been found in numerous lifetime data applications [18]. The Cox PH model offers a suitable approximation of this type because its 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 the Taylor expansion for the median of a Cox PH model. In Section 3 we find that the Cox PH model approximation is adequate in various simulations.

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 compromise is to use a semiparametric regression model (e.g., Cox PH model) based estimate for G; we use such a version of YJW's estimator for comparison with the proposed estimator in the next section. The misspecification of the semiparametric model for G, like the misspecification of F in the MIP approach, is the price for avoiding the curse of dimensionality.

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 relative adequacy of the semiparametric model based estimates of F and G should also be taken into consideration.

3 Numerical results

In this section we report some simulation results comparing the moderate-sample performance of the MIP estimator to that of YJW's estimator. Throughout, Cox PH models are used to approximate both F and G.

We consider three scenarios for the simulation model:

1. The conditional hazard function of T given Z follows a Cox PH model, but the con-

ditional 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 $[1, 2]$, $\beta_0 = (0, 1)'$, and a censoring rate of about 40% (except for Table 5 which has censoring rates 55% and 75%). The means and root mean squared errors (RMSE) of the estimators are based on 10,000 Monte Carlo replications in each case. 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 $\lambda_c(t|z_2) = 0.25 m(z_2)$ 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 MIP estimator has negligible bias and lower RMSE.

Table 1. Scenario 1, $T|Z \sim$ exponential$(\log(2)/z_2)$, $\lambda_c(t|z_2) = 0.25m(z_2)$, $n = 100$.

Estimator	*Intercept*		*Slope*	
	Mean	RMSE	Mean	RMSE
MIP	0.0184	1.0146	1.0038	0.7265
YJW	−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 MIP estimator outperforms the YJW estimator in terms of having lower RMSE. The bias in the MIP estimator is lower in one case and slightly higher in the other case.

Table 2. Scenario 2, $n = 100$.

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

We also considered the models of Table 2 with increased sample size, $n = 200$, see Table 3. The MIP estimator still has lower RMSE than YJW, but its bias is considerably higher than that of YJW in the second model.

Table 3. Scenario 2, $n = 200$.

Simulation model	*Estimator*	*Intercept*		*Slope*	
		Mean	RMSE	Mean	RMSE
$T\|Z \sim N(z_2, 0.5)$	MIP	-0.0256	0.4001	1.0248	0.2707
$\lambda_c(t\|z_2) = 0.25z_2$	YJW	-0.0303	0.4724	1.0312	0.3229
$T\|Z \sim z_2 + U$	MIP	-0.0528	0.2486	1.0618	0.1827
$\lambda_c(t\|z_2) = 0.25z_2$	YJW	-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 MIP estimator outperformed the YJW estimator in all these examples.

Table 4. Scenario 3, $n = 100$.

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

Finally we considered the first two simulation models of Table 4 with higher censoring rates, see Table 5. In both models the YJW estimator has the lower RMSE, but the MIP estimator has much lower bias than the YJW estimator in the second model. We see that neither estimator performs well under high censoring rates.

Table 5. Scenario 3, censoring rates 55% and 75%, $n = 100$.

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

4 Large sample results

We assume that the true value β_0 of β is in the interior of a known bounded convex region D. In addition, we will need the following conditions (see [2], 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 $P(X \geq t_1 \mid Z) > 0$ and $\beta' Z \leq t_1$ with probability 1.

3. The partial derivatives $f(t, z)$ and $g(t, z)$ 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 $E[ZZ' f(0 \mid Z)]$ is positive definite, where $f(t \mid z) = f(t + \beta_0' z, z)$ 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. Let $\hat{F}(t, z_k)$ be the (local) Kaplan–Meier estimator based on the pairs (X_i, δ_i), $i = 1, \ldots, n$ with the covariate values equal to z_k.

We will find it convenient to introduce some counting process notation. Let $N_{ik}^u(t) = I(X_i \le t, \delta_i = 1, Z_i = z_k), N_{ik}^c(t) = I(X_i \le t, \delta_i = 0, Z_i = z_k)$, $Y_{ik}(t) = I(X_i \ge t, Z_i = z_k)$, $M_{ik}^u(t) = N_{ik}^u(t) - \int_0^t \lambda_k^u(s) Y_{ik}(s)\, ds$, and $M_{ik}^c(t) = N_{ik}^c(t) - \int_0^t \lambda_k^c(s) Y_{ik}(s)\, ds$, where $\lambda_k^u(s)$ and $\lambda_k^c(s)$ are the conditional hazard functions of T_i and C_i given $Z_i = z_k$, respectively. Let $M_{\cdot k}^u = \sum_{i=1}^n M_{ik}^u$, $M_{\cdot k}^c = \sum_{i=1}^n M_{ik}^c$, $Y_{\cdot k} = \sum_{i=1}^n Y_{ik}$. Define $y_k(t) = P(X_i \ge t | Z_i = z_k)$ and $H(t, z) = 1 - G(t, z)$. Let $\mathcal{F}_t = \sigma\{I(X_i \le t, \delta_i = 1), I(X_i \le t, \delta_i = 0), Z_i, i = 1, \ldots, n\}$ be the filtration representing everything that happens up to time t. Note that $M_{ik}^u, M_{ik}^c, i = 1, ..., n, k = 1, ..., K$ are orthogonal martingales.

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.8).*

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 MIP and YJW estimators

The following result shows that YJW's estimator $\hat{\beta}^*$ is asymptotically equivalent to our estimator $\hat{\beta}$ based on the MIP.

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

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

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

Thus it suffices to show that $n^{-1/2}(S_n(\beta_0) - S_n^*(\beta_0)) \xrightarrow{P} 0$.

First, we give some useful identities for completing the proof. Note that $\lambda_k^u + \lambda_k^c$ is the conditional hazard function of $X_i = \min(T_i, C_i)$ given $Z_i = z_k$. 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

$$\begin{aligned} \int_{-\infty}^{\beta_0' z_k} \frac{dM_{ik}^c(s)}{y_k(s)} &= \int_{-\infty}^{\beta_0' z_k} \frac{dN_{ik}^c(s)}{y_k(s)} - \int_{-\infty}^{\beta_0' z_k} \frac{\lambda_k^c(s) I(X_i \geq s) I(Z_i = z_k)}{y_k(s)}\, ds \\ &= \frac{I(X_i < \beta_0' z_k, \delta_i = 0, Z_i = z_k)}{y_k(X_i)} - \int_{-\infty}^{\beta_0' z_k \wedge X_i} \frac{\lambda_k^c(s)}{y_k(s)}\, ds I(Z_i = z_k), \end{aligned}$$

we have,

$$\int_{-\infty}^{\beta_0' z_k} \frac{dM_{ik}^u(s)+dM_{ik}^c(s)}{y_k(s)} = \Big[\frac{I(X_i<\beta_0' z_k)}{y_k(X_i)} - \int_{-\infty}^{\beta_0' z_k \wedge X_i} \frac{\lambda_k^u(s)+\lambda_k^c(s)}{y_k(s)}\, ds\Big] I(Z_i=z_k)$$
$$= \Big[\frac{I(X_i<\beta_0' z_k)}{y_k(X_i)} - \frac{1-y_k(\beta_0' z_k \wedge X_i)}{y_k(\beta_0' z_k \wedge X_i)}\Big] I(Z_i=z_k)$$
$$= \Big[1-\frac{I(X_i\ge\beta_0' z_k)}{G(\beta_0' z_k, z_k)/2}\Big] I(Z_i=z_k). \tag{4.2}$$

¿From (2.3), we note that

$$S_n^*(\beta_0) = \sum_{k=1}^K \sum_{i=1}^n z_k \Big[\frac{I(X_i\ge\beta_0' z_k)}{G(\beta_0' z_k, z_k)} - \frac{1}{2} - \frac{\hat G(\beta_0' z_k, z_k) - G(\beta_0' z_k, z_k)}{\hat G(\beta_0' z_k, z_k) G(\beta_0' z_k, z_k)} P(X_i \ge \beta_0' z_k)$$
$$+ \frac{\hat G(\beta_0' z_k, z_k) - G(\beta_0' z_k, z_k)}{\hat G(\beta_0' z_k, z_k) G(\beta_0' z_k, z_k)} (P(X_i\ge\beta_0' z_k) - I(X_i \ge \beta_0' z_k))\Big] I(Z_i=z_k)$$

and

$$n^{-1/2}\sum_{k=1}^K z_k \frac{\hat G(\beta_0' z_k, z_k) - G(\beta_0' z_k, z_k)}{\hat G(\beta_0' z_k, z_k) G(\beta_0' z_k, z_k)} \sum_{i=1}^n (P(X_i\ge\beta_0' z_k) - I(X_i\ge\beta_0' z_k)) I(Z_i=z_k) \xrightarrow{P} 0.$$

Using a martingale representation for $(\hat G - G)/\hat G$, see [19], we have

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

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

Now, from (4.3) and (A.8), we have

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

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

$$\begin{aligned}
\int_{-\infty}^{\beta_0' z_k} \frac{G(s,z_k)}{y_k(s)}\, dM_{ik}^u(s) &= \Big[\frac{I(X_i < \beta_0' z_k, \delta_i = 1)}{F(X_i, z_k)} - \int_{-\infty}^{\beta_0' z_k \wedge X_i} \frac{\lambda_k^u(s)}{F(s,z_k)}\, ds\Big] I(Z_i = z_k) \\
&= \Big[\frac{I(X_i < \beta_0' z_k, \delta_i = 1)}{F(X_i, z_k)} - \Big(\frac{1}{F((\beta_0' z_k) \wedge X_i, z_k)} - 1\Big)\Big] I(Z_i = z_k) \\
&= \begin{cases} (-2+1) I(Z_i = z_k) & \text{if } X_i \geq \beta_0' z_k \\ \Big(\frac{I(X_i < \beta_0' z_k, \delta_i = 1)}{F(X_i, z_k)} - \frac{1}{F(X_i, z_k)} + 1\Big) I(Z_i = z_k) & \text{if } X_i < \beta_0' z_k \end{cases} \\
&= \begin{cases} -I(Z_i = z_k) & \text{if } X_i \geq \beta_0' z_k \\ \Big(-\frac{I(X_i < \beta_0' z_k, \delta_i = 0)}{F(X_i, z_k)} + 1\Big) I(Z_i = z_k) & \text{if } X_i < \beta_0' z_k, \end{cases}
\end{aligned}$$

so only the $o_p(1)$ term remains in the previous display, completing the proof. $\square$

4.3 Remarks

Writing $a_k \equiv P(Z = z_k) = \lim_{n\to\infty} n_k/n$ a.s., note that $A = -\sum_{k=1}^K a_k z_k z_k' f(0|z_k)$ and 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 $\hat\beta$ and $\hat\beta^*$, coincides with the asymptotic covariance matrix of $\sqrt{n}(\hat\beta - \beta_0)$. We note in passing that YJW's formula (A.3) and the first display following it need to have some signs corrected; see [21] for further details.

In the one sample case ($K = 1$), the expression (4.4) reduces to

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

so the asymptotic variance of $\sqrt{n}(\hat{\beta} - \beta_0)$ and $\sqrt{n}(\hat{\beta}^* - \beta_0)$ is given by

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

When $z_1 = 1$ this reduces to the well-known formula for the asymptotic variance of the median of the Kaplan–Meier estimator (see, e.g., [13] and [20]).

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

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

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

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

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

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

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

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,

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

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

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

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 $n^{-1} \tilde{S}_n(\hat{\beta}) \xrightarrow{\text{a.s.}} 0$, and thus $\hat{\beta} \xrightarrow{\text{a.s.}} \beta_0$. 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 $S_n(\beta)$

We show that for any fixed constant c and $\|\beta - \beta_0\| < cn^{-1/3}$,

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

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).

Using the lemmas in [2], it can be shown (see [21] for details) that

$$S_n(\beta) - S_n(\beta_0) = \tilde{S}_n(\beta) + o_p(n^{1/2}). \tag{A.6}$$

By a Taylor expansion of $\tilde{S}_n(\beta)$ at β_0 and 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.6), implies (A.5).

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

We may write $S_n(\beta_0) = \sum_{k=1}^{K} S_{n,k}(\beta_0)$, where

$$S_{n,k}(\beta_0) = \sum_{i=1}^{n}\left(\left\{I(X_i \geq \beta_0' z_k) + \frac{I(X_i < \beta_0' z_k, \delta_i = 0)\hat{F}(\beta_0' z_k, z_k)}{\hat{F}(X_i, z_k)}\right\} - \frac{1}{2}\right) z_k I(Z_i = z_k),$$

which we decompose as

$$\begin{aligned}
S_{n,k}(\beta_0) = \sum_{i=1}^{n}\Big\{&(I(X_i \geq \beta_0' z_k) - G(\beta_0' z_k, z_k)/2) z_k \\
&+\left(\frac{I(X_i < \beta_0' z_k, \delta_i = 0)}{F(X_i, z_k)} - P(C < \beta_0' z_k | Z = z_k)\right)\hat{F}(\beta_0' z_k, z_k) z_k \\
&+(\hat{F}(\beta_0' z_k, z_k) - F(\beta_0' z_k, z_k))P(C < \beta_0' z_k | Z = z_k) z_k \\
&+\left(\frac{1}{\hat{F}(X_i, z_k)} - \frac{1}{F(X_i, z_k)}\right) I(X_i < \beta_0' z_k, \delta_i = 0)\hat{F}(\beta_0' z_k, z_k) z_k\Big\} I(Z_i = z_k).
\end{aligned}$$

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

$$\sqrt{n_k}(\hat{F}(t, z_k) - F(t, z_k)) = -F(t, z_k) \int_{-\infty}^{t} {n_k}^{-1/2} \frac{dM^u_{\cdot k}(s)}{y_k(s)} + o_p(1). \tag{A.7}$$

Using (A.7) and (A.2) it follows that

$$\begin{aligned}
&\sum_{i=1}^{n} \left(\frac{1}{\hat{F}(X_i, z_k)} - \frac{1}{F(X_i, z_k)} \right) I(X_i < \beta_0' z_k, \delta_i = 0) I(Z_i = z_k) \\
&= \int_{-\infty}^{\infty} {n_k}^{-1} \frac{1}{y_k(s)} \sum_{i=1}^{n} \frac{I(s \le X_i) I(X_i < \beta_0' z_k, \delta_i = 0, Z_i = z_k)}{F(X_i, z_k)} \, dM^u_{\cdot k}(s) + o_p(n^{1/2}) \\
&= \int_{-\infty}^{\beta_0' z_k} \frac{H(\beta_0' z_k, z_k) - H(s, z_k)}{y_k(s)} \, dM^u_{\cdot k}(s) + o_p(n^{1/2}).
\end{aligned}$$

Now applying the martingale representation to the third term of the decomposition for $S_{n,k}(\beta_0)$, then using the definition $H(\beta_0' z_k, z_k) = P(C < \beta_0' z_k | Z = z_k)$, and noting that $F(\beta_0' z_k, z_k) = \frac{1}{2}$, it can be shown that

$$\begin{aligned}
S_{n,k}(\beta_0) \;=\; & \sum_{i=1}^{n} \Big\{ \Big[I(X_i \ge \beta_0' z_k) + \frac{I(X_i < \beta_0' z_k, \delta_i = 0)}{2F(X_i, z_k)} - \frac{1}{2} \Big] z_k I(Z_i = z_k) \\
& - \frac{1}{2} z_k \int_{-\infty}^{\beta_0' z_k} \frac{H(s, z_k)}{y_k(s)} \, dM^u_{ik}(s) \Big\} + o_p(n^{1/2}).
\end{aligned}$$

It follows that $S_n(\beta_0) = \sum_{i=1}^{n} \tau_i + o_p(n^{1/2})$, where

$$\begin{aligned}
\tau_i \;=\; & \Big[I(X_i \ge \beta_0' Z_i) + \frac{I(X_i < \beta_0' Z_i, \delta_i = 0)}{2F(X_i, Z_i)} - \frac{1}{2} \Big] Z_i \\
& - \frac{1}{2} Z_i \int_{-\infty}^{\beta_0' Z_i} \frac{H(s, Z_i)}{P(X_i \ge s | Z_i)} \, dM^u_{i\cdot}(s),
\end{aligned} \tag{A.8}$$

and $M^u_{i\cdot} = \sum_{k=1}^{K} M^u_{ik}$. This implies that $n^{-1/2} S_n(\beta_0)$ is asymptotically normal with mean zero and 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

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

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

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

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

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

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

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

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

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

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

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

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

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

[14] P. Groeneboom, and J. A. Wellner, *Information Bounds and Nonparametric Maximum Likelihood Estimation*, Birkhauser Verlag, Basel, 1992.

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

[16] J. M. Robins, and Y. Ritov, Y. Toward a curse of dimensionality appropriate (CODA) asymptotic theory for semi-parametric models, *Statist. Medicine* **16**, (1997), 285–319.

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

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

[19] R. D. Gill, *Censoring and Stochastic Integrals.* Mathematisch Centrum, Amsterdam, 1980.

[20] H. Doss, and R. D. Gill, An elementary approach to weak convergence for quantile

processes, with applications to censored survival data, *J. Amer. Statist. Assoc.* **87**, (1992), 869–877.

[21] I. W. McKeague, S. Subramanian, and Y. Sun, Median regression and the missing information principle, unpublished technical report.