Variance estimation in censored quantile regression via induced smoothing

Lei Pang, Wenbin Lu, Huixia Judy Wang

Department of Statistics, North Carolina State University, Raleigh, NC 27606, USA

1. Introduction

Consider a random sample subject to random right censoring: \( \{Y_i = \min(T_i, C_i), \Delta_i, Z_i\}_{i=1}^n \), where \( T_i \) and \( C_i \) denote the failure time and the censoring time (or some monotone transformations thereof), respectively, \( \Delta_i = I(T_i \leq C_i) \) is the censoring indicator, and \( Z_i \) is the \( p \)-dimensional covariate vector with 1 as the first component corresponding to the intercept.

We assume the following quantile regression model at a fixed quantile level \( \tau \in (0, 1) \),

\[
T_i = Z_i' \beta_0(\tau) + \epsilon_i, \quad i = 1, \ldots, n,
\]

where \( \beta_0(\tau) \) is the \( p \)-dimensional quantile coefficient at the \( \tau \)th quantile, and \( \epsilon_i \) is the random error whose \( \tau \)th conditional quantile, given \( Z_i \), equals 0. In the rest of the article, we will suppress \( \tau \) in \( \beta_0(\tau) \) for notational simplicity.

The quantile regression model, first introduced by Koenker and Bassett (1978), is a valuable alternative to the Cox proportional hazards model (Cox, 1972) and the accelerated failure time (AFT) model (Cox and Oaks, 1984) in survival analysis. Compared to the Cox and AFT models, quantile regression offers an automatic and flexible way to capture the heterogeneity in the data. Moreover, the relations between the conditional quantiles of the failure time and the covariates are directly interpretable.

Considerable research efforts have been devoted to the estimation of regression parameters in censored quantile regression. Early works of Powell (1984, 1986) focus on fixed censoring where the censoring variable \( C_i \) is always observable. For quantile regression with random censoring, recent developments include Ying et al. (1995), Bang and Tsiatis (2002), Honoré et al. (2002), Portnoy (2003), Peng and Huang (2008), Wang and Wang (2009), Huang (2010) and Portnoy and Lin (2010). However, relatively fewer methods are available for estimating the variance of censored quantile regression estimators and as such the associated inference becomes more difficult. In general, a plug-in estimator of the variance can be obtained based on the asymptotical normality of the corresponding estimator, while the asymptotical variance–covariance matrix usually involves \( f_i(0|Z_i) \), the unknown conditional error density function \( f_i(\cdot|Z_i) \) evaluated at zero. Nonparametric
approaches can be used to estimate $f_i(0|Z_i)$, but most of them require moderate to large sample size and choosing proper tuning parameters, such as the bandwidth parameter in kernel smoothing. Previous research shows that inference based on nonparametric density estimation is very sensitive to the choice of tuning parameters in finite samples; see Chen and Wei (2005) and Kocherginsky et al. (2005) for comparison of different inference methods in quantile regression without censoring. Some other researchers considered computationally intensive methods for variance estimation, e.g. Portnoy (2003) and Wang and Wang (2009) used bootstrap, and Peng and Huang (2008) employed the resampling approach of Jin et al. (2001). To bypass the variance estimation in the context of censored quantile regression, an alternative inference approach via the inversion of score tests was also considered by Ying et al. (1995) and Bang and Tsiatis (2002). However, the construction of confidence intervals is cumbersome, because one needs to perform hypothesis testing repeatedly on a fine grid in the $p$-dimensional parameter space.

The difficulty in variance estimation for censored quantile regression arises partly from the unsmoothness of the corresponding estimating function that involves indicator functions. The unsmoothness issue can be overcome by approximating the indicator functions with some suitable smooth functions. A common smoothing method is the kernel smoothing; see Chen and Hall (1993), Horowitz (1998), Heller (2007) and Whang (2006). Using kernel smoothing, if the bandwidth parameter converges to zero at a proper rate, the smoothed estimating function is asymptotically equivalent to the original estimating function, and the covariance can be consistently estimated by the sandwich formula derived from the smoothed estimating equation. Though conceptually simple, kernel smoothing usually needs a well-chosen bandwidth parameter to achieve good finite sample performance.

In this paper, we employ the induced smoothing idea of Brown and Wang (2005) and develop a variance estimation procedure for the censored quantile coefficient estimator of Bang and Tsiatis (2002). The induced smoothing method smooths the estimating function by taking its expectation under the distribution of a random perturbation of the unknown parameter. Induced smoothing is practically convenient, as it does not require selecting any tuning parameters. Recently, Wang et al. (2009) demonstrated that induced smoothing provides reliable variance estimation for quantile regression with uncensored data. Other notable applications of induced smoothing in different contexts include Brown and Wang (2006), Johnson and Strawderman (2009), Fu et al. (2010) and Wang and Fu (2010). In this paper, we will show that for censored quantile regression, our proposed variance estimator via induced smoothing is asymptotically consistent. In addition, it has good finite sample performance and is computationally efficient, requiring only a fraction of the computational cost as compared to the bootstrap method.

The rest of this article is organized as follows. In Section 2, we present the variance estimation procedure. In Section 3, we establish the asymptotic property of the variance estimator. The finite sample performance of the proposed procedure is investigated through a simulation study in Section 4, and an analysis of a multiple myeloma data set in Section 5. We conclude the paper with some discussions in Section 6. Theoretical proofs are relegated to the Appendix.

2. Induced smoothing and computational algorithm

When there is no censoring, the quantile coefficient $\beta_0$ in model (1) can be consistently estimated using the solution of the following estimating equation

$$n^{-1} \sum_{i=1}^{n} Z_i I(Y_i - \beta_0' Z_i < 0) - \tau = 0.$$  

In the presence of random right censoring, a number of estimation methods have been proposed. One popular estimator is the inverse-censoring-probability-weighted (ICPW) estimator proposed by Bang and Tsiatis (2002). Specifically, the ICPW estimator $\hat{\beta}$ is defined as the solution of the estimating equation

$$U_n(\beta) = n^{-1} \sum_{i=1}^{n} \frac{Z_i \Delta_i}{G(Y_i)} I(Y_i - \beta_0' Z_i < 0) - \tau = 0, \tag{2}$$

where $\hat{G}(\cdot)$ is the Kaplan–Meier estimate of the survival function of the censoring variable $C_i$. Note that, solving (2) is equivalent to minimizing a weighted objective function

$$n^{-1} \sum_{i=1}^{n} \frac{\Delta_i}{G(Y_i)} \rho_\tau(Y_i - \beta_0' Z_i), \tag{3}$$

where $\rho_\tau(t) = \tau t - \frac{1}{2} t(t < 0)$. The minimization can be solved efficiently by using existing linear programming algorithms, for instance, the rq function in R package quantreg.

In the Bang and Tsiatis method, censoring times are assumed to be independent and identically distributed, and independent of covariates, which we also assume throughout the rest of this paper. Bang and Tsiatis (2002) proved that under some regularity conditions, $\hat{\beta}$ is consistent and asymptotically normal:

$$n^{1/2} (\hat{\beta} - \beta_0) \xrightarrow{d} \mathcal{N}(0, \Gamma),$$

where $\Gamma = \mathbf{A}^{-1} \Sigma \mathbf{A}^{-1}$, $\mathbf{A} = \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} Z_i f_i(0|Z_i)$, and $\Sigma = \lim_{n \to \infty} \text{Var} \{ n^{1/2} U_n(\beta_0) \}$. It is well known that the direct estimation of $\Gamma$ is difficult, since the matrix $\mathbf{A}$ depends on the unknown error densities $f_i(\cdot|Z_i)$, $i = 1, \ldots, n$. Please cite this article in press as: Pang, L., et al., Variance estimation in censored quantile regression via induced smoothing. Computational Statistics and Data Analysis (2010), doi:10.1016/j.csda.2010.01.018
By the asymptotic normality of \( \hat{\beta} \), we can write \( \hat{\beta} = \beta_0 + H^{1/2}V \), where \( H = n^{-1} \Gamma, V \sim N(0, I_p) \), and \( I_p \) is the \( p \times p \) identity matrix. In this way, \( \hat{\beta} \) can be regarded as a random perturbation of \( \beta_0 \). Motivated by the induced smoothing idea of Brown and Wang (2005), we consider the smoothed estimating function \( \tilde{U}_n(\beta, H) = E(U_n(\beta + H^{1/2}V)) \), the expectation of the unsmoothed estimating function with respect to \( V \). We have

\[
\tilde{U}_n(\beta, H) = E_V \left\{ U_n(\beta + H^{1/2}V) \right\} = \frac{1}{n} \sum_{i=1}^{n} \frac{Z_i \Delta_i}{G(Y_i)} \left\{ \phi \left( \frac{\beta'Z_i - Y_i}{\sqrt{Z_i'HWZ_i}} \right) - \tau \right\},
\]

where \( \phi(\cdot) \) is the standard normal cumulative distribution function. When \( H \) is given, the smoothed quantile coefficient estimator \( \hat{\beta} \) can be obtained by solving \( \tilde{U}_n(\hat{\beta}, H) = 0 \). However, since \( H \) is unknown and has to be estimated, we propose to use a positive definite \( p \times p \) matrix \( H_0 = O(n^{-1}) \) as an initial value in the smoothing, and then update through iterations to estimate \( H \) and hence \( \Gamma \). In fact, we will show, as long as the matrix used for initial smoothing is of order \( O(n^{-1}) \) and positive definite, the corresponding smoothed coefficient estimator \( \hat{\beta} \) is asymptotically equivalent to \( \hat{\beta} \).

For any positive definite matrix \( W = O(n^{-1}) \), we define

\[
\tilde{A}_n(\beta, W) = \frac{\partial \tilde{U}_n(\beta, W)}{\partial \beta} = n^{-1} \sum_{i=1}^{n} \frac{\Delta_i}{G(Y_i)} \phi \left( \frac{\beta'Z_i - Y_i}{\sqrt{Z_i'WZ_i}} \right) - \frac{Z_i Z_i'}{\sqrt{Z_i'WZ_i}}.
\]

The estimation algorithm can be summarized as follows.

**Step 1.** Let \( \tilde{\beta}_0 = \hat{\beta} \), the Bang-Tsiatis estimator, and \( \tilde{H}_0 = n^{-1} I_n \).

**Step 2.** Given \( \tilde{\beta}_{k-1} \) and \( \tilde{H}_{k-1} \) from the \((k-1)\)th step, update \( \tilde{\beta}_k \) and \( \tilde{H}_k \) as:

\[
\begin{align*}
\tilde{\beta}_k &= \tilde{\beta}_{k-1} + \left\{ -\tilde{A}_n \left( \tilde{\beta}_{k-1}, \tilde{H}_{k-1} \right) \right\}^{-1} \tilde{U}_n \left( \tilde{\beta}_{k-1}, \tilde{H}_{k-1} \right), \\
\tilde{H}_k &= n^{-1} \tilde{A}_n^{-1} \left( \tilde{\beta}_k, \tilde{H}_{k-1} \right) \hat{\Sigma} \left( \tilde{\beta}_k, \tilde{H}_{k-1} \right) \tilde{A}_n^{-1} \left( \tilde{\beta}_k, \tilde{H}_{k-1} \right),
\end{align*}
\]

where \( \hat{\Sigma}(\tilde{\beta}_k) \) is the consistent estimator of the covariance matrix of \( \sqrt{n}U_n(\tilde{\beta}_k) \), as defined in (10) in the Appendix.

**Step 3.** Repeat step 2 till convergence. Denote the coefficient estimate and the covariance estimate at convergence as \( \hat{\beta} \) and \( \tilde{H} \), respectively, and let \( \hat{\Gamma} = \tilde{H} \).

In addition, we denote the covariance estimator of \( n^{1/2} \tilde{\beta}_k \) at the \( k \)th iteration as \( \hat{\Gamma}_k = n \tilde{H}_k \). Thereafter, we will refer to this procedure as IS (Induced Smoothing).

### 3. Asymptotic results

In this section, we establish the consistency and asymptotic normality of \( \tilde{\beta}_k \), and the consistency of \( \hat{\Gamma}_k \). We impose the following regularity conditions.

**A1.** The conditional error distribution functions, \( f_i(\cdot | Z_i) \), are absolutely continuous with continuous densities \( f_i(\cdot | Z_i) \) uniformly bounded away from 0 and \( \infty \) in a neighborhood of 0, and \( f_i'(\cdot | Z_i) \) exists and is uniformly bounded on the real line.

**A2.** For each \( i = 1, \ldots, n \), \( Z_i \) satisfies the following conditions:

\( n^{-1} \sum_{i=1}^{n} Z_i Z_i' f_i(0 | Z_i) \) converges to a positive definite matrix \( A \);

(b) \( \sup \| Z_i \| < \infty \), where \( \| \cdot \| \) denotes the Euclidean norm.

**A3.** There exists \( L > 0 \) such that \( P(C > L) = 0 \) and \( P(C = L) \geq \nu \), where \( \nu \) is some positive constant.

**Theorem 1.** Assume conditions A1–A3 hold, and the initial smoothing matrix \( H_0 \) is positive definite and \( O(n^{-1}) \), as \( n \to \infty \). Then we have

\( i \) \( n^{1/2} (\tilde{\beta}_k - \beta_0) \xrightarrow{d} \mathcal{N}(0, \Gamma) \), for any \( k \geq 1 \);

\( ii \) \( \tilde{\Gamma}_k \xrightarrow{p} \Gamma \), for any \( k \geq 1 \).

**Remark.** Assumptions A1 and A2 impose some conditions on the conditional error distributions and the covariates, both of which will hold in most practical situations. Assumption A3 will be satisfied in many clinical studies with administrative censoring. Theorem 1(i) suggests that the smoothed coefficient estimator, \( \tilde{\beta}_k \), has the same asymptotic distribution as the unsmoothed estimator \( \hat{\beta} \). Theorem 1(ii) provides the theoretical justification for IS as a variance estimation procedure for the censored quantile regression model (1). In addition, Theorem 1 suggest that \( \beta \) and \( \hat{\Gamma} \), the coefficient and variance estimators at convergence, have the same asymptotic properties as the one-step estimators \( \tilde{\beta}_1 \) and \( \hat{\Gamma}_1 \), respectively. However, our numerical studies show that the IS variance estimator at convergence has smaller bias than the one-step estimator in finite samples. The proposed algorithm converges quite fast, usually within 4 iterations for the simulation study in Section 4 and within 4–8 iterations for the data analysis in Section 5.
4. Simulation

In this section, we conduct a simulation study to compare three variance estimation methods: the proposed IS method, the bootstrap method (BTBOOT) and the kernel smoothing method with prespecified bandwidths. For BTBOOT, the bootstrap coefficient estimates are obtained by minimizing (3) using bootstrap samples generated by resampling the triplets \( \{ Y_i, \Delta_i, Z_i \} \) with replacement. The bootstrap covariance matrix is calculated as the sample covariance of the bootstrap coefficient estimates. For kernel smoothing, the indicator functions \( I(Y_i - \beta'Z_i < 0) \) are approximated by \( \Phi \left\{ (Y_i - \beta'Z_i)/h \right\} \), where \( h \) is the bandwidth, and the covariance matrix is estimated by the sandwich formula using the M-estimation technique. We consider three bandwidths in the kernel smoothing \( h = n^{-0.1}, n^{-0.3} \) and \( n^{-0.7} \), and denote them as \( S_1, S_2 \) and \( S_3 \), respectively. Note that the induced smoothing approach can be viewed as a variation of the kernel smoothing method: in induced smoothing, rather than a fixed bandwidth \( h \), each subject receives a subject-specific bandwidth \( h_i = \sqrt{Z_i^T H Z_i} \) that is determined automatically through iteration.

We generate data from the following model:

\[
\log(T_i) = Z_i' \beta + \epsilon_i, \quad i = 1, \ldots, n,
\]

where \( Z_i = (1, Z_{i1}, Z_{i2})' \), \( Z_{i1} \sim \text{Unif}(0, 4) \), \( Z_{i2} \sim \text{Bernoulli}(0.5) \) and \( \beta = (\beta_1, \beta_2, \beta_3)' = (1, 1, 1)' \). We consider three different scenarios for generating the random error \( \epsilon_i \).

- Scenario 1: homoscedastic normal, \( \epsilon_i \sim \text{N}(0, 1) \);
- Scenario 2: heteroscedastic normal, \( \epsilon_i \sim \text{N}(0, (1 + 0.3Z_{i1})^2) \);
- Scenario 3: heteroscedastic logistic, \( \epsilon_i \sim \text{Logistic}(0, 1 + 0.3Z_{i1}) \).

The censoring times \( C_i \) are generated by \( \log(C_i) \sim \text{Unif}(0, k) \), where \( k \) is set as 23, 12 and 7, corresponding to 15%, 30% and 50% of censoring, respectively. For each scenario, 2000 simulated subjects of size \( n = 200 \) are generated. We focus on two quantiles, \( \tau = 0.25 \) and \( \tau = 0.5 \).

Tables 1–3 summarize the simulation results for 15%, 30% and 50% censoring, respectively. For each method, we report the mean bias and the mean squared error (MSE) of each estimate of \( \beta_i \), \( i = 1, 2, 3 \), the empirical coverage probability of 95% Wald-type confidence intervals, and SE/SD, where SD is the Monte Carlo standard deviation and SE is the average of estimated standard errors across 2000 simulations.

In cases with moderate censoring (15% and 30%), the coefficient estimates from IS and BTBOOT are essentially unbiased, while those from the kernel smoothing methods tend to be biased when the bandwidth is selected too large. Theorem 1 suggests that the smoothed coefficient estimator from IS, \( \hat{\beta}_i \), is asymptotically equivalent to the BT estimator. However, simulation results show that the ultimate point estimator from IS has smaller mean squared errors than the BT estimator and thus has some finite sample advantages. In terms of variance estimation, IS performs similarly as BTBOOT, and both provide reasonable variance estimates and confidence intervals with coverage probabilities close to the nominal level, 95%. In contrast, the kernel smoothing is very sensitive to the choice of bandwidth \( h \). Generally speaking, kernel smoothing tends to produce inflated variance estimation when more smoothing (larger \( h \)) is applied.

Estimation and inference are more challenging in cases with 50% censoring. The upper quantiles may not be identifiable. Table 3 shows that for data sets with 50% censoring, all methods yield point estimates with larger bias especially at median, which is possibly due to the identifiability issue. Consequently, the confidence intervals from IS and BTBOOT have coverage probabilities lower than 95% at \( \tau = 0.5 \). However, two methods still provide reasonable confidence intervals at the lower quantile \( \tau = 0.25 \).

In summary, the IS method is comparable to bootstrap in terms of variance estimation. However, IS is computationally more efficient than BTBOOT. For example, in Scenario 1 with \( n = 200 \) and 15% censoring, it took IS 4.27 s and BTBOOT 621 s to finish the computation for 100 data sets by using R 2.9.2 on a PC with an Intel i7 920 CPU clocked at 3.6 GHz. In this simulation study, we focused on data sets with 2 covariates. Additional simulation studies suggest that the proposed method also work well for data sets with more covariates at the cost of more computational time. For instance, the IS method required 134 s for analyzing 2000 simulated data sets with 6 covariates, compared to 94 s for analyzing 2000 data sets with 2 covariates.

5. Analysis of a multiple myeloma data

In this section, we apply the proposed method to analyze the survival times of multiple myeloma patients. Multiple myeloma is a plasma cell cancer that happens in 4 per 10 million in the human population. Almost all multiple myeloma patients have genetic lesions in their cancer cells. Recently, to identify specific genetic lesions associated with prognosis, a genetic analysis of malignant plasma cells from 192 multiple myeloma patients was done using high density SNP arrays (Avet-Loiseau et al., 2009). After discarding those observations with missing values in the covariates, we have 170 observations with 54% censoring. We fit the quantile regression model to study the relation between the log survival times and 4 covariates, including serum Beta-2 microglobulin (\( B_2M \)) level, genetic markers del(13) (deletion on chromosome 13), t(4; 14) (translocation between chromosome 4 and 14) and t(11; 14) (translocation between chromosome 11 and 14). We
focus on two quantile levels $\tau = 0.25$ and 0.5. In cancer studies, lower quantiles are of particular interest, as they correspond to high-risk cases with shorter survivals.

Table 4 summarizes the quantile coefficient estimates, the corresponding estimated standard errors, 95% confidence intervals and p-values obtained from three methods: IS, BT-boot and kernel smoothing with $h = n^{-0.3}$. In accordance with our simulation study, all three methods give similar coefficient estimates. All three methods suggest that $\delta(13)$ and $\tau(11; 14)$ have no significant effect, while $\tau(4; 14)$ has a significant adverse effect on the two quantiles of the survival times, which is consistent with the existing results (Keats et al., 2003). In addition, the quantile coefficient of $\tau(4; 14)$ is estimated as $-1.062$ at $\tau = 0.5$ and $-1.037$ at $\tau = 0.25$, indicating that the adverse effects of the genetic lesion $\tau(4; 14)$ on patients’ survival are about the same at both quantiles.

The main discrepancy of three methods lies in the assessment of the $B_2M$ effect. At $\tau = 0.5$, both IS and BT-boot suggest a marginal significance of $B_2M$, while the kernel smoothing method gives a much larger variance estimate, leading to an insignificance. Note that this data has heavy right censoring (54%). Simulation studies in Section 4 suggest that analysis at median could be unstable under such heavy censoring, and this may partly explain the different variance estimation of the kernel smoothing method. At the lower quantile $\tau = 0.25$, all three methods show that $B_2M$ has a significantly adverse effect. In the biomedical literature, $B_2M$ has been shown to be an important factor in multiple myeloma prognostis; see Avet-Loiseau et al. (2009), Greipp et al. (1993) and Facon et al. (2001). By studying different quantiles, our analysis reveals a richer picture. The results of IS show that the estimated quantile coefficient of $B_2M$ is $-0.083$ at $\tau = 0.5$ and $-0.134$ at $\tau = 0.25$, suggesting that high $B_2M$ level has an adverse effect on patients’ survival times, and this adverse effect tends to
be larger for those with shorter survivals, i.e. at the lower quantiles. Further, biological investigations are needed to provide a more definite explanation of the differential impacts of \( B_2 M \) on different locations of the survival distribution (Table 4).

6. Conclusion and remarks

We developed a new inference procedure for censored quantile regression based on induced smoothing, which is computationally fast and theoretically valid. In this paper, we focused on the Bang–Tsiatis estimator because of its computational convenience. The proposed idea can be extended to other censored quantile estimators. For example, for regression at the \( r \) th quantile, we can extend the median regression approach of Ying et al. (1995) and consider the following estimating equation

\[
\hat{\beta}_r = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{I(Y_i - Z_i' \hat{\beta} \geq 0)}{\hat{G}(Z_i' \hat{\beta})} - 1 + \tau \right\}.
\]

Note that in the above equation, \( \hat{\beta} \) is involved in both the numerator and the denominator, and thus the induced smoothing cannot be directly applied. However, the difficulty can be overcome if there exists a good initial estimator of \( \hat{\beta} \). For instance, we can use \( \hat{\beta}_0 \), the estimate obtained based on Eq. (2) as an initial estimate. Then an iterative estimation can proceed as follows: at each step, we plug the \( \hat{\beta} \) estimate from the previous step into \( \hat{G}(Z_i' \beta) \), apply induced smoothing to \( I(Y_i - Z_i' \hat{\beta} \geq 0) \) and update both the coefficient and variance estimates. Further investigation is needed in this direction.
In this paper, we focused on survival data by making the unconditional independence assumption of survival and censoring times as in Ying et al. (1995). The method can be directly extended to analyze the censored medical cost data with informative censoring, the original context studied in Bang and Tsiatis (2002). In survival analysis, it is more practical to assume the conditional independence of survival and censoring times, given covariates. To incorporate covariate-dependent censoring, we can adopt the local Kaplan–Meier estimates \( \hat{G}(Y_i|Z_i) \) in constructing the inverse-censoring-probability weights in (2); see Wang and Wang (2009) for related discussions of the local Kaplan–Meier estimates. The corresponding variance estimation can be carried out following the same line as in the current paper.

The R implementation of the proposed method is available at http://www4.stat.ncsu.edu/~wang/software.html.

Acknowledgements

Lu’s research was partly supported by NIH/NCI grant R01 CA-140632, and Wang’s research was partly supported by NSF grants DMS-0706963 and DMS-1007420. The authors would like to thank the editor, an associate editor, and two referees for their valuable suggestions.

Appendix

To prove Theorem 1, we need the following lemma.
Table 4
Results of the multiple myeloma data analysis. For each method, $\hat{\beta}$ denotes the quantile coefficient estimate, lb and ub denote the lower and the upper bounds of a 95% Wald-type confidence interval, and SE is the estimated standard error.

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>$\beta$</th>
<th>lb</th>
<th>ub</th>
<th>SE</th>
<th>$p$-value</th>
<th>lb</th>
<th>ub</th>
<th>SE</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$B_0$M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>def(13)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t(11; 14)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t(11; 14)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t(11; 14)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Lemma 1. Let $W = O(n^{-1})$ be any positive definite matrix, and define

$$\tilde{U}_n(\beta, W) = n^{-1} \sum_{i=1}^{n} Z_i \Delta_i \frac{G(Y_i)}{\sqrt{Z_i'WZ_i}} \Phi \left( \frac{\beta_i Z_i - Y_i}{\sqrt{Z_i'WZ_i}} \right)$$

as the smoothed estimating function. Under assumptions A1–A3, we have

$$\sup_{|\beta - \beta_0| \leq \epsilon_n} \left| n^{1/2} \{ \tilde{U}_n(\beta, W) - U_n(\beta) \} \right| \xrightarrow{p} 0, \quad \text{as } n \to \infty,$$

where $\{\epsilon_n\}$ is a positive sequence that converges to 0.

Proof of Lemma 1. Let $\sigma_i = (Z_i'WZ_i)^{1/2}$, $\epsilon_i^\beta = Y_i - Z_i'\beta$ and $d_i(\beta) = \text{sgn}(\epsilon_i^\beta) \Phi(-|\epsilon_i^\beta|/\sigma_i)$, where $\text{sgn}(\cdot)$ is the sign function. We have

$$n^{1/2} \{ \tilde{U}_n(\beta, W) - U_n(\beta) \} = n^{-1/2} \sum_{i=1}^{n} \frac{Z_i \Delta_i}{G(Y_i)} \left\{ \Phi \left( \frac{-\epsilon_i^\beta}{\sigma_i} \right) - I(\epsilon_i^\beta < 0) \right\} = n^{-1/2} \sum_{i=1}^{n} \frac{Z_i \Delta_i}{G(Y_i)} d_i(\beta).$$

Denote $D_n(\beta) = n^{-1/2} \sum_{i=1}^{n} \frac{Z_i \Delta_i}{G(Y_i)} d_i(\beta)$ and $D_n^C(\beta) = n^{-1/2} \sum_{i=1}^{n} \frac{Z_i \Delta_i}{G(Y_i)} d_i(\beta)$. It follows that

$$D_n(\beta) = D_n^C(\beta) - n^{-1/2} \sum_{i=1}^{n} \frac{Z_i \Delta_i \{ \hat{G}(Y_i) - G(Y_i) \}}{G^2(Y_i)} d_i(\beta) + o_p(1).$$

Taking the expectation of $D_n^C(\beta)$, we have

$$E[D_n^C(\beta)] = n^{-1/2} \sum_{i=1}^{n} Z_i E[d_i(\beta)].$$

where

$$E[d_i(\beta)] = \int_{-\infty}^{+\infty} \text{sgn}(\epsilon_i^\beta) \Phi \left( \frac{-\epsilon_i^\beta}{\sigma_i} \right) f(Y_i) \sigma_i d\epsilon_i^\beta,$$

$$= \sigma_i \int_{-\infty}^{+\infty} \Phi(-|t|)[2I(t > 0) - 1]f(\sigma_i t + Z_i' (\beta - \beta_0)) dt,$$

$$= \sigma_i \int_{-\infty}^{+\infty} \Phi(-|t|)[2I(t > 0) - 1][f(Z_i' (\beta - \beta_0)) + f'(\sigma_i^2(t)) \sigma_i t] dt,$$

where $f$ is the density function of $Y_i$. Please cite this article in press as: Pang, L., et al., Variance estimation in censored quantile regression via induced smoothing, Computational Statistics and Data Analysis (2010), doi:10.1016/j.csda.2010.10.018.
where \( f_i \) is the density of \( \epsilon_i = \epsilon_i^\beta \), and \( \omega_i^\beta(t) \) is between \( Z_i'(\beta - \beta_0) \) and \( Z_i'(\beta - \beta_0) + \sigma_i^\beta t \). Note that for \( \beta \) that satisfies \( \| \beta - \beta_0 \| \leq \varepsilon_n \), where \( \varepsilon_n \to 0 \), we have \( |Z_i'(\beta - \beta_0)| \to 0 \). It follows from assumption A1 that \( \sup f_i[Z_i'(\beta - \beta_0)] < \infty \), and since \( \int_{-\infty}^{+\infty} \Phi(-|t|)2I(t > 0) - 1 \) \( dt = 0 \), we have \( \int_{-\infty}^{+\infty} \Phi(-|t|)2I(t > 0) - 1 \left| f_i[Z_i'(\beta - \beta_0)] \right| \) \( dt = 0 \). In addition, by assumption A1, we can find \( M > 0 \) such that \( \sup \| f_i'[\omega_i^\beta(t)] \| < M \). Thus, it follows that
\[
|E[d_i(\beta)]| \leq \sigma_i^2 \int_{-\infty}^{+\infty} \left| t \Phi(-|t|) f_i'[\omega_i^\beta(t)] \right| \) \( dt \leq M \sigma_i^2 / 2, \]
where the last equality holds because \( \int_{-\infty}^{+\infty} \left| t \Phi(-|t|) \right| dt = 1 / 2). By assumption A2 and the fact that \( W = O(n^{-1}) \), \( \sum_{i=1}^{n} \sigma_i^2 = \text{tr}(WZ'Z) = \text{tr}(WZ'Z) \) is bounded, and \( \sum_{i=1}^{n} |E[d_i(\beta)]| \leq M \sum_{i=1}^{n} \sigma_i^2 / 2 \) is also bounded. Therefore,
\[
\|E[D_n^G(\beta)]\| \leq n^{-1/2} \sqrt{p} \sup_{i,j} |Z_{ij}^\beta| \sum_{i=1}^{n} |E[d_i(\beta)]| \to 0, \quad \text{as} \quad n \to \infty.
\]
In addition, by assumption A3,
\[
\text{Var} \{D_n^G(\beta)\} = \frac{1}{n} \sum_{i=1}^{n} Z_{ii}^\beta \text{Var} \left\{ \frac{\Delta_i}{G(Y_i)} d_i(\beta) \right\} \leq \frac{1}{n} \sum_{i=1}^{n} Z_{ii}^\beta \text{Var} \{d_i^2(\beta)\},
\]
where
\[
E[d_i^2(\beta)] = \int_{-\infty}^{+\infty} \Phi^2(-|s|) f_i[\sigma_i^\beta(s) + Z_i'(\beta - \beta_0)] ds = \int_{|s| > \Delta} \Phi^2(-|s|) f_i[\sigma_i^\beta(s) + Z_i'(\beta - \beta_0)] ds + \int_{|s| \leq \Delta} \Phi^2(-|s|) f_i[\sigma_i^\beta(s) + Z_i'(\beta - \beta_0)] ds \leq \Phi^2(-\Delta) + \sigma_i \Delta \Phi(\Delta) + 2 \sigma_i \Delta f_i(\omega_i^\beta) \]
\[
= \Phi^2(-\Delta) + 2 \sigma_i \Delta f_i(\omega_i^\beta).
\]
Note that \( \omega_i^\beta \in (Z_i'(\beta - \beta_0) - \sigma_i, Z_i'(\beta - \beta_0) + \sigma_i) \). Let \( \Delta = n^{1/4} \) and since \( \sigma_i = O(n^{-1/2}) \), both \( \sigma_i \Delta \) and \( \omega_i^\beta \) go to zero as \( n \) increases. As \( f_i(\cdot) \) is uniformly bounded around zero, both \( \Phi^2(-\Delta) \) and \( \sigma_i \Delta f_i(\omega_i^\beta) \) go to zero as \( n \to \infty \). Thus, it follows that \( \lim_{n \to \infty} E[d_i^2(\beta)] = 0 \), and \( \lim_{n \to \infty} \text{Var} \{D_n^G(\beta)\} = 0 \). By the Weak Law of Large Numbers, for \( \beta \) that satisfies \( \| \beta - \beta_0 \| \leq \varepsilon_n \), we have
\[
\|D_n^G(\beta)\| \to 0, \quad \text{as} \quad n \to \infty.
\]
The second term on the right side of (5) can be written as
\[
n^{-1/2} \sum_{i=1}^{n} \int_{0}^{L} \left\{ n^{-1} \sum_{i=1}^{n} \frac{\Delta_i Z_i d_i(\beta) Y_i(u)}{G(Y_i)} \right\} \frac{\text{d}M_j^c(u)}{y(u)} + o_p(1),
\]
where \( M_j^c(u) = N_j^c(u) - \int_{0}^{L} I(Y_i \geq u) d\Lambda^c(s), \) \( N_j^c(u) = (1 - \delta_i) I(Y_i \leq u) \), \( \Lambda^c(u) = -\text{log}[G(u)] \) is the censoring cumulative hazard, \( Y_i(u) = I(Y_i \geq u) \) is the i-th at-risk process, and \( y(u) = \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} Y_i(u) \) is bounded from below in \((0, L)\] by assumption A3.
Define \( l_i(u, \beta) = n^{-1} \sum_{i=1}^{n} \frac{\Delta_i Z_i d_i(\beta) Y_i(u)}{G(Y_i)} \), and \( l(u, \beta) = E[l_i(u, \beta)] \). We have
\[
l(u, \beta) = n^{-1} \sum_{i=1}^{n} Z_i E[d_i(\beta) Y_i(u)],
\]
where \( |E[d_i(\beta) Y_i(u)]| \leq E[d_i(\beta)] \), and
\[
E[d_i(\beta)] = \int_{-\infty}^{+\infty} \Phi \left( -\frac{\epsilon_i^\beta}{\sigma_i} \right) f_i[\epsilon_i^\beta + Z_i'(\beta - \beta_0)] d\epsilon_i^\beta
\]
\[
= \sigma_i \int_{-\infty}^{+\infty} \Phi(-|t|) f_i[\sigma_i t + Z_i'(\beta - \beta_0)] dt
\]
\[
= \sigma_i f_i[Z_i'(\beta - \beta_0)] \int_{-\infty}^{+\infty} \Phi(-|t|) dt + \sigma_i^2 \int_{-\infty}^{+\infty} t \Phi(-|t|) f_i'[\omega_i^\beta(t)] dt.
\]
By assumption A1, we have \( f_i(Z_i'(\beta - \beta_0)) \int^{-\infty}_0 \Phi(-|t|)dt < \infty \), and \( \int^{+\infty}_0 t \Phi(-|t|)f_i'(\alpha_i(t)) dt < \infty \). Thus, it follows that \( E[d_i(\beta)] = O(n^{-1/2}) \), and

\[
\|I(u, \beta)\| \leq \sqrt{p} \sup_{t_j} |Z_j| n^{-1} \sum_{i=1}^{n} E[d_i(\beta)] = O(n^{-1/2}) \to 0.
\]

Define \( \mathcal{F} = \{ \Delta_n d_i(Y_i) G_i(Y_i) \} \). \( \|\beta - \beta_0\| \leq \varepsilon_n \) and \( u \in (0, \infty) \). The function class \( \mathcal{F} \) is Glivenko–Cantelli (van der Vaart and Wellner, 1996) because the class of indicator functions is Glivenko–Cantelli, and \( Z_i, d_i(\beta) \), and \( 1/G(Y_i) \) are uniformly bounded. It follows that \( \sup_{\|\beta - \beta_0\| \leq \varepsilon_n, u \in (0, \infty)} \|I(u, \beta) - I(u, \beta)\| \to 0 \), and we have

\[
n^{-1/2} \sum_{j=1}^{n} \int_{0}^{L} I_n(u, \beta) \frac{dM_j(u)}{y(u)} = n^{-1/2} \sum_{j=1}^{n} \int_{0}^{L} I(u, \beta) \frac{dM_j(u)}{y(u)} + o_p(1).
\]

By the Martingale Central Limit Theorem (Fleming and Harrington, 1991), \( n^{-1/2} \sum_{j=1}^{n} \int_{0}^{L} I(u, \beta) \frac{dM_j(u)}{y(u)} \) is \( o_p(1) \) as \( n \) goes to infinity. It follows that, for \( \beta \) that satisfies \( \|\beta - \beta_0\| \leq \varepsilon_n \),

\[
\left\| n^{-1/2} \sum_{j=1}^{n} \int_{0}^{L} I_n(u, \beta) \frac{dM_j(u)}{y(u)} \right\| \to 0. \tag{7}
\]

Collating (6) and (7), we have

\[
\left\| n^{-1/2} \{ \tilde{U}_n(\beta, \mathbf{W}) - U_n(\beta) \} \right\| \to 0.
\]

for any \( \beta \) such that \( \|\beta - \beta_0\| \leq \varepsilon_n \). Lemma 1 is thus proven by the fact that both \( \tilde{U}_n(\beta, \mathbf{W}) \) and \( U_n(\beta) \) are monotone functions, thus the point-wise convergence could be strengthened to uniform convergence (Shorack and Wellner, 1986).

\textbf{Proof of Theorem 1(i).} Consider the one-step estimator of \( \beta_0 \),

\[
\tilde{\beta}_1 = \tilde{\beta}_0 + (\tilde{\mathbf{A}}_n(\tilde{\beta}_0, \mathbf{H}_0))^{-1} \tilde{U}_n(\tilde{\beta}_0, \mathbf{H}_0),
\]

where \( \tilde{\beta}_0 = \hat{\beta} \) is the unsmoothed estimator.

To show the asymptotic equivalency of \( \tilde{\beta}_1 \) and \( \hat{\beta}_0 \), we first prove that for any positive definite matrix \( \mathbf{W} = O(n^{-1}) \),

\[
\tilde{\mathbf{A}}_n(\beta_0, \mathbf{W}) \to \mathbf{A}.
\]

Define \( \sigma_i = \sqrt{Z_i} \mathbf{W} Z_i \), \( \epsilon_i = Y_i - Z_i \beta_0 \), and \( \eta_i = \phi(-\epsilon_i/\sigma_i)/\sigma_i \). Then

\[
\tilde{\mathbf{A}}_n(\beta_0, \mathbf{W}) = n^{-1} \sum_{i=1}^{n} Z_i Z_i' \frac{\Delta_i}{G(Y_i)} \eta_i.
\]

For any vectors \( a, b \in \mathbb{R}^p \),

\[
a'\tilde{\mathbf{A}}_n(\beta_0, \mathbf{W})b = a' \left\{ n^{-1} \sum_{i=1}^{n} Z_i Z_i' \frac{\Delta_i}{G(Y_i)} \eta_i \right\} b + o_p(1),
\]

where

\[
E \left[ a' \left\{ n^{-1} \sum_{i=1}^{n} Z_i Z_i' \frac{\Delta_i}{G(Y_i)} \eta_i \right\} b \right] = a' \left\{ n^{-1} \sum_{i=1}^{n} Z_i Z_i' E(\eta_i) \right\} b.
\]

and

\[
E(\eta_i) = \frac{1}{\sigma_i} \int \phi \left( -\frac{\epsilon_i}{\sigma_i} \right) f_i(\epsilon_i) d\epsilon_i = \int \phi(-t)f_i(\sigma_i) dt = \int \phi(-t)f_i(0) dt + \sigma_i \int t \phi(-t)f_i'(\omega_i) dt,
\]

where \( \int \phi(-t)f_i(0) dt = f_i(0) \), and \( \sigma_i \int t \phi(-t)f_i'(\omega_i) dt \leq M_\sigma \int |t| \phi(-t) dt \to 0 \).

It follows that \( E(\eta_i) \to f_i(0) \) and

\[
\lim_{n \to \infty} E \left[ a' \left\{ n^{-1} \sum_{i=1}^{n} Z_i Z_i' \frac{\Delta_i}{G(Y_i)} \right\} b \right] = a' \left\{ \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} Z_i Z_i' f_i(0) \right\} b = a' \mathbf{A} b. \tag{8}
\]
By assumption A1, 
\[
E(n\eta_i^2) = \frac{1}{\sigma_i} \int \phi^2 \left( -\frac{\epsilon_i}{\sigma_i} \right) f_i(\epsilon_i) d\epsilon_i = \frac{1}{\sigma_i} \int \phi^2 (-t) f_i(\sigma_i t) dt
\]
\[
= \frac{1}{\sigma_i} \int \phi^2 (-t) f_i(0) + f_i^\prime (\omega^* \sigma_i t) dt
\]
\[
= \frac{1}{\sigma_i} \int \phi^2 (-t) f_i(0) dt + \int t \phi^2 (-t) f_i^\prime (\omega^*) dt
\]
\[
= O(n^{1/2}).
\]

Consequently,
\[
\text{Var} \left[ \alpha' \left\{ -n^{-1} \sum_{i=1}^{n} Z_i Z_i^\prime \frac{\Delta_i}{G(Y_i)} \eta_i \right\} b \right] \leq \frac{1}{n^{2\beta}} \sum_{i=1}^{n} (\alpha' Z_i Z_i^\prime b)^2 E(\eta_i^2) \to 0. \tag{9}
\]

By (8) and (9), we have \( \tilde{A}_n(\beta_0, W) \to A \). Thus, \( \tilde{A}_n(\tilde{\beta}_0, \tilde{H}_0) \to A \) because \( \tilde{\beta}_0 \to \beta_0 \) (Bang and Tsiatis, 2002). By Lemma 1, \( n^{1/2} \tilde{U}_n(\tilde{\beta}_0, \tilde{H}_0) = n^{1/2} U_n(\tilde{\beta}_0) + o_p(1) = o_p(1) \).

Now consider 
\[
n^{1/2}(\tilde{\beta}_1 - \beta_0) = n^{1/2}(\tilde{\beta}_0 - \beta_0) + (\tilde{A}_n(\tilde{\beta}_0, \tilde{H}_0))^{-1} [n^{1/2} \tilde{U}_n(\tilde{\beta}_0, \tilde{H}_0)],
\]
where \( \tilde{A}_n(\tilde{\beta}_0, \tilde{H}_0) \to A \) and \( n^{1/2} U_n(\tilde{\beta}_0, \tilde{H}_0) \to 0 \). It follows that \( n^{1/2}(\tilde{\beta}_1 - \beta_0) \to n^{1/2}(\tilde{\beta}_0 - \beta_0) \), which together with the asymptotic normality of \( \beta_0 \) (Bang and Tsiatis, 2002) gives
\[
n^{1/2}(\tilde{\beta}_1 - \beta_0) \to \mathcal{N}(0, \Gamma).
\]

The proof for the asymptotic properties of \( \tilde{\beta}_k \) with \( k > 1 \) follows with similar arguments. \( \square \)

**Proof of Theorem 1(ii).** We define 
\[
\tilde{\Gamma}_k = \left\{ \tilde{A}_n(\tilde{\beta}_k, \tilde{H}_{k-1}) \right\}^{-1} \tilde{\Sigma}(\tilde{\beta}_k) \left\{ \tilde{A}_n(\tilde{\beta}_k, \tilde{H}_{k-1}) \right\}^{-1},
\]
where 
\[
\tilde{\Sigma}(\beta) = \frac{1}{4n} \sum_{i=1}^{n} \frac{\Delta_i}{G(Y_i)} Z_i Z_i^\prime \left[ n^{-1} \sum_{i=1}^{n} \frac{\Delta_i}{G(Y_i)} \left( I(Y_i - Z_i' \beta \leq 0) - \tau \right) Z_i \right]^\otimes 2
\]
\[
+ n^{-1} \int \frac{dN^c(u)}{G(Y)} \left[ K \left( I(Y_i - Z_i' \beta \leq 0) - \tau \right) Z_i^\otimes 2, u \right] - \tilde{K} \otimes 2 \left( I(Y_i - Z_i' \beta \leq 0) - \tau \right) Z_i, u \right]. \tag{10}
\]
Here \( K(W, u) = n^{-1} S^{-1}(u) \sum_{i=1}^{n} \frac{\Delta_i W_i I(Y_i \geq u)}{G(Y_i)} \), \( d^\otimes 2 = a d^2 \) for vector \( a \), and \( S(t) \) is the survival function for the failure time \( T \).

Since \( \tilde{\beta}_k \to \beta_0 \) for any \( k \), and \( \tilde{\Sigma}(\beta_0) \to \Sigma \) by Bang and Tsiatis (2002), it follows that \( \tilde{\Sigma}(\tilde{\beta}_k) \to \Sigma \). In addition, \( \tilde{H}_{k-1} = O(n^{-1}) \), thus we have \( \tilde{A}_n(\tilde{\beta}_k, \tilde{H}_{k-1}) \to A \), and consequently, \( \tilde{\Gamma}_k \to A^{-1} \Sigma A^{-1} = \Gamma \). \( \square \)

**References**


Please cite this article in press as: Fang, L., et al., Variance estimation in censored quantile regression via induced smoothing. Computational Statistics and Data Analysis (2010), doi:10.1016/j.csda.2010.10.018