Published in final edited form as:

Stat Methodol. 2007 October ; 4(4): 493–503. doi:10.1016/j.stamet.2007.05.002.

Censored Median Regression and Profile Empirical Likelihood

Sundarraman Subramanian¹

Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, New Jersey,
email: sundars@njit.edu

Abstract

We implement profile empirical likelihood based inference for censored median regression models. Inference for any specified sub-vector is carried out by profiling out the nuisance parameters from the “plug-in” empirical likelihood ratio function proposed by Qin and Tsao. To obtain the critical value of the profile empirical likelihood ratio statistic, we first investigate its asymptotic distribution. The limiting distribution is a sum of weighted chi square distributions. Unlike for the full empirical likelihood, however, the derived asymptotic distribution has intractable covariance structure. Therefore, we employ the bootstrap to obtain the critical value, and compare the resulting confidence intervals with the ones obtained through Basawa and Koul’s minimum dispersion statistic. Furthermore, we obtain confidence intervals for the age and treatment effects in a lung cancer data set.

Keywords

Inverse censoring weighted; Kaplan–Meier estimator of censoring; Lagrange multipliers; Local linearity; Normal approximation; Nuisance parameters

1. INTRODUCTION

Median regression (MR) models provide a robust alternative to the Cox and the accelerated failure time models. One focus recently has been inference for MR parameters from right censored data, see Powell (1986), Newey and Powell (1990), Ying, Jung, and Wei (1995), Lindgren (1997), Yang (1999), McKeague, Subramanian, and Sun (2001), Subramanian (2002), Bang and Tsiatis (2002), Portnoy (2003), and Qin and Tsao (2003), among others. For this problem, estimation typically follows a two-step “estimating equation” approach: One first sets up the least absolute deviation (LAD) estimating function, which is then minimized to specify the estimator. Due to the difficulty of estimating the covariance matrix of the estimator, however, alternatives to the usual Wald-type confidence regions have been explored. These are the normal approximation (NA) based and the empirical likelihood (EL) based methods, see Ying et al. (1995) and Qin and Tsao (2003). The latter authors, in particular, proposed a procedure that combines the inverse censoring weighted (ICW) estimating function of Ying et al. (1995) and the EL, and presented numerical evidence that the EL confidence region for the *full* MR parameter has better empirical coverage probabilities (ECPs) than the NA based

² Author’s address: Sundar Subramanian, Associate Professor, Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, New Jersey (email: sundars@njit.edu).

¹ This research was supported by U.S. National Cancer Institute grant R15 CA103845.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

confidence region. The inverse censoring weighted approach was first proposed by Koul, Susarla, and Van Ryzin (1981) for linear regression.

In applications, however, the focus may be only on a subset of the regression parameters. In particular, it may be often necessary to obtain confidence intervals for individual regression effects, see Ying et al. (1995) who computed confidence intervals for treatment and age effects in a lung cancer study using the minimum dispersion (MD) statistic of Basawa and Koul (1988). Their procedure however is NA based, which, as mentioned above, is reported to be inferior to the EL method in terms of finite sample performance. The EL of Qin and Tsao (2003), on the other hand, cannot be employed to compute confidence intervals for a subset. A desirable method would be the profile EL, which can be carried out by profiling out nuisance parameters from the “plug-in” EL ratio function. This approach, however, has not been implemented for censored median regression. Quite clearly, the method that provides confidence intervals with better coverage should be the preferred choice in practice. Toward that end, this paper investigates the efficacy of profile EL confidence intervals.

Let T be the response, Z a $p + 1$ dimensional covariate whose first component is 1, and suppose that $T = \beta^T Z + \varepsilon$, where β is the unknown regression parameter vector. The joint distribution of ε and Z is unspecified, but the conditional median of ε given Z is known to be 0. For the right censored version we observe $X = \min(T, C) = T \wedge C$, where C is the censoring variable independent of T and Z , see Ying et al. (1995), and Qin and Tsao (2003), and $\delta = I(X = T)$ is the censoring indicator. The data, thus, are n iid copies of (X, δ, Z) .

Our first approach of profile EL inference is to employ asymptotic methods to derive the limiting distribution of the profile EL ratio function and thereby attempt to compute its critical value. Since the adjusted LAD estimating functions are non-smooth, a profile EL theorem for censored median regression would require theoretical justification, unlike for smooth estimating functions (Owen, 2001). Indeed, we show that the limiting distribution of the profile EL ratio function is a weighted sum of independent chi-squared distributions each with one degree of freedom. Unlike for the full EL case, however, the weights are the eigenvalues of a matrix that depends on the conditional distribution of the error given the covariate, and hence difficult to estimate. The critical value for the profile EL ratio statistic, therefore, cannot be calculated easily using the derived distribution, necessitating the search for alternate methods. Our second approach is to employ the bootstrap to obtain the critical value (Namba, 2004). We compare the performances of the profile EL and NA based approaches through Monte Carlo simulation.

In Section 2, we review existing approaches, present a profile EL theorem, describe the bootstrap method, and produce results from a numerical study. In Section 3 we provide an illustration using a lung cancer data set. Proof of the profile EL theorem is in the Appendix.

2. INFERENCE FOR THE REGRESSION PARAMETER

This section begins with a review of the existing approaches. After that, we develop the profile EL approach and present a profile EL theorem. Finally, we introduce the bootstrap method and provide numerical justification in support of profile EL.

2.1 The ICW estimating function and NA based inference

Let $\|\cdot\|$ denote the Euclidean norm. Ying et al. (1995) proposed the estimating function

$$S_n(\beta) = \sum_{i=1}^n \left\{ \frac{I(X_i \geq \beta^T Z_i)}{\widehat{G}(\beta^T Z_i)} - \frac{1}{2} \right\} Z_i, \quad (1)$$

where $\hat{G}(t)$ is the Kaplan–Meier estimator of the censoring distribution. Let β_0 denote the true value of β and let $\Lambda_G(t)$ denote the censoring cumulative hazard function. Let $\Upsilon = -E[ZZ^T f_{\varepsilon}(0|Z)]$, where $f_{\varepsilon}(t|z) = f(t + \beta_0^T z | z)$ is the conditional density of ε given $Z = z$. Here $f(t|z)$ is the conditional density of T given Z . Furthermore, for any vector a , let $a^{\otimes 2} = aa^T$, and let $q(t)$ and $h(t)$ denote the limits of $n^{-1} \sum_{i=1}^n I(\beta_0^T Z_i \geq t) Z_i$ and $n^{-1} \sum_{i=1}^n I(X_i \geq t)$. Let $\Psi_0 \equiv \Psi(\beta_0)$ and $\Psi_0^* \equiv \Psi^*(\beta_0)$, where

$$\Psi_0 = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \left[\frac{I(X_i \geq \beta_0^T Z_i)}{G(\beta_0^T Z_i)} - \frac{1}{2} \right] Z_i^{\otimes 2}, \quad \Psi_0^* = \frac{1}{4} \int_{-\infty}^{\infty} \frac{q(t)q^T(t)}{h(t)} dt d\Lambda_G(t).$$

Under certain regularity conditions (cf. Section 2.2), the estimator $\hat{\beta}$, which is defined as a minimizer of $\|S_n(\beta)\|$, is consistent and $n^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{\mathcal{D}} N(0, \Upsilon^{-1} \Gamma \Upsilon^{-1})$, where $\Gamma = \Psi_0 - \Psi_0^*$. The conditional density of the error distribution given the covariate is difficult to estimate in practice, however. Therefore, asymptotic confidence regions for the MR parameters via Wald-type statistics are difficult to obtain. Alternatively, the test statistic

$V(\beta_0) = n^{-1} S_n^T(\beta_0) \hat{\Gamma}^{-1} S_n(\beta_0)$ is approximately chi-square distributed with $p + 1$ degrees of freedom, leading to a $100(1 - \alpha)\%$ confidence region $\{\beta: V(\beta) < \chi_{p+1}^2(1 - \alpha)\}$ for β_0 . Here $\hat{\Gamma}$ is a consistent estimator of Γ . On the other hand, when $\beta^T = (\beta^{(1)T}, \beta^{(2)T})^T$, where $\beta^{(1)}$ is the $q \times 1$ parameter sub-vector of interest, and $\beta_0^T = (\beta_0^{(1)T}, \beta_0^{(2)T})^T$, the MD statistic given by $D(\beta_0^{(1)}) = \min_{\beta_2} \left\{ n^{-1} S_n^T(\beta_0^{(1)T}, \beta_2^{(2)T}) \hat{\Gamma}^{-1} S_n(\beta_0^{(1)T}, \beta_2^{(2)T}) \right\}$ is approximately chi-square distributed with q degrees of freedom. This in turn results in a $100(1 - \alpha)\%$ confidence region for $\beta_0^{(1)}$ given by $\{\beta^{(1)}: D(\beta^{(1)}) < \chi_q^2(1 - \alpha)\}$.

2.2 EL inference for the full parameter

Let $E_i(\beta, G) = I(X_i \geq \beta^T Z_i) / G(\beta^T Z_i) - 1/2$. Since $E(E_i(\beta_0, G)) = 0$, the estimating function $\sum_{i=1}^n E_i(\beta, G)$ is unbiased when G is known. The EL ratio function is

$$L_n(\beta, G) = \max \left\{ \prod_{i=1}^n (n p_i) : \text{each } p_i \geq 0, \sum_{j=1}^n p_j = 1, \sum_{j=1}^n p_j E_j(\beta, G) = 0 \right\}. \quad (2)$$

From EL theory, $L_n(\beta_0, G) = \prod_{i=1}^n n p_i$, where the maximizing weights p_i are given by $p_i = n^{-1} (1 + \lambda_0^T E_i(\beta_0, G))^{-1}$ and $\lambda_0 = \lambda(\beta_0)$ is the vector of Lagrange multipliers satisfying the set of equations $n^{-1} \sum_{i=1}^n E_i(\beta_0, G) (1 + \lambda_0^T E_i(\beta_0, G))^{-1} = 0$. This leads to a confidence region $\{\beta: -2 \log L_n(\beta, G) < \chi_{p+1}^2(1 - \alpha)\}$ for β_0 , see Owen (1990). The nuisance parameter G , however, is unknown so, employing a plug-in $\hat{G}$, the EL ratio function is $L_n(\beta, \hat{G})$. Let $l(\beta) = -2 \log L_n(\beta, \hat{G})$. Assume that the true value β_0 is in the interior of a known bounded convex region $\mathcal{D}$. Suppose also that the following conditions hold:

1. The covariate vector Z is bounded, say $\|Z\| \leq L$.
2. For $\beta \in \mathcal{D}$, there exists a t_0 such that $P(X \geq t_0 | Z) > 0$ and $\beta^T Z \leq t_0$ almost surely.
3. Let $\mathcal{F} = (-\infty, t_0] \times [-L, L]$. The derivatives of the conditional survival function $F(t|z)$ of T given Z and $G(t|z)$ with respect to t are uniformly bounded in $(t, z) \in \mathcal{F}$.

4. The matrix $E[ZZ^T f_{\varepsilon}(0|Z)]$ is positive definite.

Qin and Tsao (2003) proved that $l(\beta_0)$ converges in distribution to L , where L is a weighted sum of p independent chi-squared distributions each with one degree of freedom. The weights are the eigenvalues of the matrix $\Psi_0^{-1}\Gamma$, which can be estimated from the data. Using their result, a confidence region $R_{\alpha}(\beta)$ for the full regression parameter vector can be obtained by inverting $l(\beta)$ as $R_{\alpha}(\beta) = \{\beta : l(\beta) \leq c_{\alpha}\}$, where c_{α} is the upper α quantile of the distribution of L .

2.3 A profile EL theorem

Our focus is EL confidence region for the $q \times 1$ subvector $\beta^{(1)}$ of $\beta = (\beta^{(1)T}, \beta^{(2)T})^T$. Define $\beta_0 = (\beta_0^{(1)T}, \beta_0^{(2)T})^T$ and $\mathcal{N} = \{\beta^{(2)} : \|\beta^{(2)} - \widehat{\beta}^{(2)}\| = O(n^{-1/3})\}$. Denote the profile EL ratio function by $l_s(\beta^{(1)}) = \min_{\beta^{(2)} \in \mathcal{N}} l((\beta^{(1)T}, \beta^{(2)T})^T)$. The following theorem gives the asymptotic distribution of $l_s(\beta_0^{(1)})$.

Theorem 1—Assume conditions 1–4. Then the profile EL ratio function $l_s(\beta_0^{(1)})$ converges in distribution to $k_1 U_1 + \dots + k_q U_q$, where the weights $k_1, \dots, k_q$ are the eigenvalues of $([I_q, 0] \Upsilon^{-1} \Psi_0 \Upsilon^{-1} [I_q, 0]^T)^{-1} ([I_q, 0] \Upsilon^{-1} \Gamma \Upsilon^{-1} [I_q, 0]^T)$ and $U_1, \dots, U_q$ are independent chi-squared random variables each with 1 degree of freedom.

The weights are difficult to estimate because the matrix Υ depends on $f_{\varepsilon}(0|Z)$ which, for arbitrary Z , is difficult to estimate nonparametrically. Therefore, unlike for the full EL case, the profile EL theorem does not readily lead to a profile EL confidence region for $\beta_0^{(1)}$. In the next section we describe the bootstrap method for computing the critical value of the profile EL ratio statistic. We remark that the profile EL theorem also extends to the doubly censored version of the MR model considered by Subramanian (2007).

2.4 Bootstrapping profile EL

Following Namba (2004), we propose the following bootstrap approach to test the null hypothesis $H_0: \beta^{(1)} = \beta_0^{(1)}$ against a general alternative. Resample $(X_i^*, \delta_i^*, Z_i^*)_{1 \leq i \leq n}$ from the original sample $(X_i, \delta_i, Z_i)_{1 \leq i \leq n}$ with replacement. For each bootstrap sample, compute

$\left\{ E_i^* ((\beta_0^{(1)T}, \beta^{(2)T})^T, \widehat{G}^*) \right\}_{1 \leq i \leq n}$, where $\widehat{G}^*$ is the Kaplan–Meier estimator of G based on the bootstrap sample, and $\beta^{(2)}$ varies over $\mathcal{N}$. Recenter (cf. Hall and Horowitz, 1996) each

$\left\{ E_i^* ((\beta_0^{(1)T}, \beta^{(2)T})^T, \widehat{G}^*) \right\}_{1 \leq i \leq n}$ by subtracting their average over n values. Then compute the

sequence of values $l^* ((\beta_0^{(1)T}, \beta^{(2)T})^T) = -2 \log L_n^* ((\beta_0^{(1)T}, \beta^{(2)T})^T, \widehat{G}^*)$ for each bootstrap sample.

Next obtain the bootstrap profile EL ratio $l_s^*(\beta^{(1)})$ by finding the minimum of the above mentioned sequence. Repeat the process b times, arrange the b values in increasing order and choose the $b(1 - \alpha)$ th value. This is the bootstrap critical value.

2.5 A numerical study

We considered a simple linear regression model with $Z^T = (1, Z_2)$, where Z_2 was uniform on $(0, 1)$, and $\beta_0 = (0, 1)^T$. We investigated with two models. The error distributions for Model 1 and Model 2 were, respectively, $N(0, Z_2)$ and logistic with mean 0 and scale parameter Z_2 . The censoring C was uniform on $(0, c)$, where c was chosen to get 10%, 20% and 30% censoring rates (CRs). A grid search over $[-1, 1] \times [-1, 3]$ was employed to obtain the point estimates

of β . The minimization for computing the profile EL ratio statistic was performed over a grid of values of the intercept parameter centered at its point estimate. The grid width was 0.2 or 0.3. The bootstrap critical value was the 95th percentile of the bootstrap profile EL ratio statistic values evaluated at the true parameter value of 1. The ECPs of the 95% asymptotic confidence regions of the slope parameter were based on 1,000 Monte Carlo replications at sample size $n = 100$. The number of bootstrap replications was $b = 10,000$. The ECPs of the profile EL are close to the nominal 95% level, see Table 1.

3. ILLUSTRATION USING A LUNG CANCER DATA SET

Ying et al. (1995) used a lung cancer study to illustrate the MR model for right censored data. Patients with small cell lung cancer were assigned randomly to two treatments, namely Arm A, which is Cisplatin (P) followed by Etoposide (E), and Arm B, which is E followed by P. The response is the base 10 log failure time and the covariates are age and treatment indicator. Using a grid search, our point estimates of the intercept, age, and treatment effects are respectively 2.90, -0.002 , -0.161 . Ying et al. (1995), however, obtained their estimates as 3.028, -0.004 , -0.163 . Treating the intercept $\beta^{(1)}$ and the treatment effect $\beta^{(2)}$ as nuisance parameters, a 95% confidence interval for the age effect $\beta_0^{(3)}$ using the MD statistic is $(-0.0064, 0.0025)$. The bootstrap critical value requires knowledge of $\beta_0^{(3)}$, however. We computed the critical value for various values of $\beta_0^{(3)}$ in the interval $(-0.0063, 0.0023)$. The value of $\beta_0^{(3)} = -0.0064$ provided the most conservative interval, namely $(-0.0063, 0.0023)$. Other values resulted in slightly narrow confidence intervals. Treating the intercept $\beta^{(1)}$ and the age effect $\beta^{(3)}$ as nuisance parameters, a 95% confidence interval for the treatment effect $\beta_0^{(2)}$ using the MD statistic is $(-0.3893, -0.0285)$. The bootstrap critical value with $\beta_0^{(2)} = -0.39$ provided the most conservative interval of $(-0.3893, -0.0486)$ for $\beta_0^{(2)}$.

4. CONCLUSIONS

For inference involving any subset of the MR parameter vector, the profile EL, implemented in this article, is a competing alternative to the currently available methodology based on the MD statistic. Ying et al. (1995) have given theoretical justification for the applicability of the latter methodology. Unlike for many other models, however, such justification for the profile EL in the case of censored MR requires considerable care because it involves minimization of non-smooth functions over an appropriate nuisance parameter range. We have shown that the limiting distribution of the profile EL ratio function is a weighted sum of independent chi-squared distributions each with one degree of freedom, but the weights are difficult to compute in practice for arbitrary Z . The bootstrap provides a viable means of obtaining the critical value of the profile EL ratio statistic. The consistency of the bootstrap critical value remains to be proved, however. This is a worthwhile direction for future research. On the basis of our simulation study, we report that the bootstrap profile EL confidence intervals provide better coverage than their counterpart obtained using the MD statistic. We have also provided competing confidence intervals for individual effects in the case of the lung cancer data set, which is not possible using Qin and Tsao's (2003) full EL method. The specification of the "bandwidth" for computing the bootstrap profile EL ratio would require further investigation, however, since an improper choice can deleteriously affect coverage.

One reviewer was interested in understanding why the empirical likelihood produces improved finite sample performance over the NA based methods. Indeed this phenomenon has been observed in several other cases also, see Qin and Jing (2001), Li and Wang (2003), and Wang and Veraverbeke (2006) among others. Qin and Jing (2001) point to the instability of

covariance estimates as contributing to inferior coverage in the case of the NA intervals. Furthermore, EL confidence intervals are asymmetric, which perhaps better describes the true underlying (possibly skewed) distribution of the parameter estimate, see Li and Wang (2003). This is also a worthwhile direction for further investigations.

Acknowledgements

The author is grateful to Professor Jogesh Babu for providing a timely review of the manuscript, and for giving the opportunity for submitting a revision. The author also thanks the Associate Editor and a referee for their critical comments. Finally, the author wishes to express his thanks to Professor Gang Li for useful discussions and for providing many references involving the bootstrap.

References

1. Bang H, Tsiatis AA. Median regression with censored cost data. *Biometrics* 2002;58:643–649. [PubMed: 12229999]
2. Basawa IV, Koul HL. Large-sample statistics based on quadratic dispersion. *Int Statist Rev* 1980;56:199–219.
3. Csörgő S, Horváth L. The rate of strong uniform consistency for the product-limit estimator. *Z Wahrsch Verw Gebiete* 1983;62:411–426.
4. Hall P, Horowitz JL. Bootstrap critical values for tests based on generalized-method-of-moments estimators. *Econometrica* 1996;64:891–916.
5. Koul H, Susarla V, Van Ryzin J. Regression analysis with randomly right censored data. *Ann Statist* 1981;9:1276–1288.
6. Li G, Wang QH. Empirical likelihood regression analysis for right censored data. *Statist Sinica* 2003;13:51–68.
7. Lindgren A. Quantile regression with censored data using generalized L_1 minimization. *Comput Statist, Data Anal* 1997;23:509–524.
8. McKeague IW, Subramanian S, Sun Y. Median regression and the missing information principle. *J Nonpara Statist* 2001;13:709–727.
9. Namba A. Simulation studies on bootstrap empirical likelihood tests. *Comm Statist Simula Comput* 2004;33:99–108.
10. Newey WK, Powell JL. Efficient estimation of linear and type I censored regression models under conditional quantile restrictions. *Econometric Theory* 1990;6:295–317.
11. Owen A. Empirical likelihood ratio confidence regions. *Ann Statist* 1990;18:90–120.
12. Owen, A. *Empirical Likelihood*. Chapman & Hall/CRC; Boca Raton: 2001.
13. Powell JL. Censored regression quantiles. *J Econometrics* 1986;32:143–155.
14. Portnoy S. Censored Regression Quantiles. *J Ameri Statist Assoc* 2003;98:1001–1012.
15. Qin G, Jing BY. Empirical likelihood for censored linear regression. *Scand J Statist* 2001;28:661–673.
16. Qin G, Tsao M. Empirical likelihood inference for median regression models for censored survival data. *J Mult Anal* 2003;85:416–430.
17. Subramanian S. Median regression using nonparametric kernel estimation. *J Nonpara Statist* 2002;14:583–605.
18. Subramanian S. Median regression analysis from data with left and right censored observations. *Statist Methodology* 2007;4:121–131.
19. Wang L, Veraverbeke N. Empirical likelihood in a semi-parametric model for missing response data. *Comm Statist Th Meth* 2006;35:625–639.
20. Wei LJ, Ying Z, Lin DY. Linear regression analysis of censored survival data based on rank tests. *Biometrika* 1990;77:845–851.
21. Yang S. Censored median regression using weighted empirical survival and hazard functions. *J Ameri Statist Assoc* 1999;94:137–145.

22. Ying Z, Jung S, Wei LJ. Survival analysis with median regression models. J Amer Statist Assoc 1995;90:178–184.

APPENDIX 1

Preliminaries

We will need a result of Ying et al. (1995): For any β in the $n^{-1/3}$ neighborhood of β_0 ,

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

Let $\mathcal{B}_n^*$ denote the $n^{-1/3}$ neighborhood of β inside $\mathcal{D}$. Recall that $E_i(\beta, G) \equiv E_i(\beta)$ is the quantity $\{I(X_i \geq \beta^T Z_i)/G(\beta^T Z_i) - 1/2\} Z_i$. Also let $E_i(\beta, \hat{G}) \equiv \hat{E}_i(\beta)$. Define $\widehat{\Psi}(\beta) = n^{-1} \sum_{i=1}^n \widehat{E}_i(\beta) \widehat{E}_i^T(\beta)$, and $\Psi_n(\beta) = n^{-1} \sum_{i=1}^n E_i(\beta) E_i^T(\beta)$. We have the following results:

Lemma 1

For every $\eta > 0$, $\widehat{\Psi}(\beta) = \Psi_n(\beta) + o(n^{-1/2+\eta})$ almost surely, uniformly for $\beta \in \mathcal{B}_n^*$.

Proof

For any $a \in \mathbb{R}^{p+1}$, we employ the decomposition (Qin and Tsao 2003)

$$a^T (\widehat{\Psi} - \Psi_n) a = \frac{1}{n} \sum_{i=1}^n (a^T (\widehat{E}_i(\beta) - E_i(\beta)))^2 + \frac{2}{n} \sum_{i=1}^n (a^T E_i(\beta)) (a^T (\widehat{E}_i(\beta) - E_i(\beta))).$$

We only consider the second term above. Note that $\sup_{t \leq t_0} |\hat{G}(t) - G(t)| = o(n^{-1/2+\eta})$, almost surely, see Csörgő and Horváth (1983). Also, since $\beta \in \mathcal{B}_n^* \subset \mathcal{D}$, by Condition 2, $\inf_{\beta \in \mathcal{B}_n^*} G(\beta^T Z) \geq G(t_0)$ almost surely. Therefore, the absolute value of the second term is bounded above by

$$\left| \frac{1}{G(t_0)} - \frac{1}{\widehat{G}(t_0)} \right| \frac{1}{G(t_0)} \sup_{t \leq t_0} |\widehat{G}(t) - G(t)| \frac{1}{n} \sum_{i=1}^n (a^T Z_i)^2 I(X_i \geq \beta^T Z_i),$$

which is $o(n^{-1/2+\eta})$ almost surely, for all $\eta > 0$. Likewise, the first term. $\square$

Lemma 2

We have $\widehat{\Psi}(\beta) = \widehat{\Psi}(\beta_0) + O_p(\|\beta - \beta_0\|)$ and $\Psi_n(\beta) = \Psi_n(\beta_0) + O_p(\|\beta - \beta_0\|)$.

Proof

We use the decomposition specified in the proof of Lemma 1 and Eq. (A.7) of Ying et al. (1995). The steps are similar to the proof of Lemma 1, so we omit the details.

Recall that $\beta_0 = (\beta_0^{(1)T}, \beta_0^{(2)T})^T$ and $\mathcal{N} = \{\beta^{(2)}: \|\beta^{(2)} - \widehat{\beta}^{(2)}\| = O(n^{-1/3})\}$. Let $\mathcal{B}_n = \left\{ \beta: \beta = \begin{pmatrix} \beta_0^{(1)T} \\ \beta_0^{(2)T} \end{pmatrix}^T, \beta^{(2)} \in \mathcal{N} \right\} \subset \mathcal{B}_n^*$. Note that $S_n(\beta) = \sum_{i=1}^n \widehat{E}_i(\beta)$, and $l(\beta) = 2 \sum_{i=1}^n \log(1 + \lambda^T(\beta) \widehat{E}_i(\beta))$, where $\lambda(\beta)$ is the solution of

$$g(\lambda(\beta)) \equiv \sum_{i=1}^n \frac{\widehat{E}_i(\beta)}{1 + \lambda^T(\beta) \widehat{E}_i(\beta)} = 0. \quad (\text{A.2})$$

Using a standard argument we have that $\|\lambda(\beta_0)\| = O_p(n^{-1/2})$, see page 220 of Owen (2001). The following lemma involving $\lambda(\beta)$ is required to prove Theorem 1:

Lemma 3

For any $\beta \in \mathcal{B}_n$, the quantity $\lambda(\beta)$ satisfies $\|\lambda(\beta)\| = O_p(\|\beta - \beta_0\|)$.

Proof

Note that $\theta^T(\beta)g(\lambda(\beta)) = 0$. First, following p. 220 of Owen (2001), we obtain

$$\theta^T(\beta)n^{-1}S_n(\beta) = \|\lambda(\beta)\| \theta^T(\beta)\widehat{\Omega}(\beta)\theta(\beta),$$

where $\widehat{\Omega}(\beta) = n^{-1} \sum_{i=1}^n (\widehat{E}_i(\beta)\widehat{E}_i^T(\beta))/(1 + \lambda^T(\beta)\widehat{E}_i(\beta))$, and $\theta(\beta)$ is the unit vector defined by $\lambda(\beta) = \|\lambda(\beta)\|\theta(\beta)$. Again, following p. 220 of Owen (2001), we next obtain

$$\|\lambda(\beta)\| \theta^T(\beta)\widehat{\Psi}(\beta)\theta(\beta) \leq \|\lambda(\beta)\| \theta^T(\beta)\widehat{\Omega}(\beta)\theta(\beta) \left(1 + \max_{1 \leq i \leq n} \lambda^T(\beta)\widehat{E}_i(\beta)\right). \quad (\text{A.3})$$

Since $\max_{1 \leq i \leq n} \lambda^T(\beta)\widehat{E}_i(\beta) \leq \|\lambda(\beta)\| \max_{1 \leq i \leq n} \|\widehat{E}_i(\beta)\|$ we then obtain using Eq. (A.3) that

$$\|\lambda(\beta)\| \theta^T(\beta)\widehat{\Psi}(\beta)\theta(\beta) \leq \theta^T(\beta)n^{-1}S_n(\beta) \left(1 + \|\lambda(\beta)\| \max_{1 \leq i \leq n} \|\widehat{E}_i(\beta)\|\right). \quad (\text{A.4})$$

It then follows that

$$\|\lambda(\beta)\| \left\{ \theta^T(\beta)\widehat{\Psi}(\beta)\theta(\beta) - \max_{1 \leq i \leq n} \|\widehat{E}_i(\beta)\| \theta^T(\beta)n^{-1}S_n(\beta) \right\} \leq \theta^T(\beta)n^{-1}S_n(\beta).$$

Let $R_n = o_p(\max(n^{-1/2}, \|\beta - \beta_0\|))$. Using Eq. (A.1) we then obtain

$$\begin{aligned} & \|\lambda(\beta)\| \left\{ \theta^T(\beta)\widehat{\Psi}(\beta)\theta(\beta) - \max_{1 \leq i \leq n} \|\widehat{E}_i(\beta)\| \theta^T(\beta)n^{-1}S_n(\beta_0) + \theta^T(\beta)\Upsilon(\beta - \beta_0) + R_n \right\} \\ & \leq \theta^T(\beta)n^{-1}S_n(\beta_0) + \theta^T(\beta)\Upsilon(\beta - \beta_0) + R_n = O_p(\|\beta - \beta_0\|), \end{aligned}$$

the last step following because $n^{-1}S_n(\beta_0) = O_p(n^{-1/2}) = o_p(\|\beta - \beta_0\|)$, since $\beta \in \mathcal{B}_n$. Since $\max_{1 \leq i \leq n} \|\widehat{E}_i(\beta)\| = O_p(1)$, it follows that the second term inside the brace above is $o_p(1)$. Using Lemma 2 and some routine calculations it can be shown that the first term inside the brace above is $O_p(1)$. The statement of Lemma 3 follows. The proof is complete.

The next significant link for proving Theorem 1 is provided by the following lemma:

Lemma 4

We have that $\min_{\beta \in \mathcal{B}_n} \|\lambda(\beta)\| = O_p(n^{-1/2})$.

Proof

Using Eq. (A.2) we obtain the following standard equation (Qin and Tsao, 2003):

$$\lambda(\beta) = \left(\sum_{j=1}^n \widehat{E}_j(\beta) \widehat{E}_j^T(\beta) \right)^{-1} S_n(\beta) + \left(\sum_{j=1}^n \widehat{E}_j(\beta) \widehat{E}_j^T(\beta) \right)^{-1} \sum_{i=1}^n \frac{\widehat{E}_i(\beta) (\lambda^T(\beta) \widehat{E}_i(\beta))^2}{1 + \lambda^T(\beta) \widehat{E}_i(\beta)}.$$

Since $\max_{1 \leq i \leq n} \|\widehat{E}_i(\beta)\| = O_p(1)$, $\sum_{j=1}^n \widehat{E}_j(\beta) \widehat{E}_j^T(\beta) = O_p(n)$, and $\beta \in \mathcal{B}_n$, by Lemma 3 the second term is $O_p(\|\lambda(\beta)\|^2 (\max_{1 \leq i \leq n} \|\widehat{E}_i(\beta)\|)^2) = o_p(n^{-1/2})$. Therefore, using Eq. (A.1) and Lemma 2 we have that, uniformly for $\beta \in \mathcal{B}_n$,

$$\begin{aligned} \lambda(\beta) &= \left(n \widehat{\Psi}(\beta) \right)^{-1} S_n(\beta) + o_p(n^{-1/2}) \\ &= \left(n \widehat{\Psi}(\beta) \right)^{-1} S_n(\beta_0) + \left(\widehat{\Psi}(\beta) \right)^{-1} \Upsilon(\beta - \beta_0) + o_p(\max(n^{-1/2}, \|\beta - \beta_0\|)) \\ &= \left(n \widehat{\Psi}(\beta_0) \right)^{-1} S_n(\beta_0) + \left[\left(n \widehat{\Psi}(\beta) \right)^{-1} - \left(n \widehat{\Psi}(\beta_0) \right)^{-1} \right] S_n(\beta_0) \\ &\quad + \left(\widehat{\Psi}(\beta) \right)^{-1} \Upsilon(\beta - \beta_0) + o_p(\max(n^{-1/2}, \|\beta - \beta_0\|)) \\ &= \left(n \widehat{\Psi}(\beta_0) \right)^{-1} S_n(\beta_0) - n^{-1} \left(\widehat{\Psi}(\beta) \right)^{-1} \left(\widehat{\Psi}(\beta) - \widehat{\Psi}(\beta_0) \right) \left(\widehat{\Psi}(\beta_0) \right)^{-1} S_n(\beta_0) \\ &\quad + \left(\widehat{\Psi}(\beta) \right)^{-1} \Upsilon(\beta - \beta_0) + o_p(\max(n^{-1/2}, \|\beta - \beta_0\|)) \\ &= \left(n \widehat{\Psi}(\beta_0) \right)^{-1} S_n(\beta_0) + \left(\widehat{\Psi}(\beta) \right)^{-1} \Upsilon(\beta - \beta_0) + o_p(\max(n^{-1/2}, \|\beta - \beta_0\|)) \\ &= \lambda(\beta_0) + \left(\widehat{\Psi}(\beta) \right)^{-1} \Upsilon(\beta - \beta_0) + o_p(\max(n^{-1/2}, \|\beta - \beta_0\|)) \\ &= \lambda(\beta_0) + \left(\widehat{\Psi}(\beta) \right)^{-1} \Upsilon[0, I_{p-q}]^T \{(\beta^{(2)} - \widehat{\beta}^{(2)}) + (\widehat{\beta}^{(2)} - \beta_0^{(2)})\} \\ &\quad + o_p(\max(n^{-1/2}, \|\beta - \beta_0\|)), \end{aligned} \tag{A.5}$$

Note that $\widehat{\beta}^{(2)} - \beta_0^{(2)} = O_p(n^{-1/2})$. Then, $\min_{\beta \in \mathcal{B}_n} \|\lambda(\beta)\| = \|\lambda(\beta_0)\| + o_p(n^{-1/2})$.

APPENDIX 2

Proof of Theorem 1

Using (A.2), we can show that

$$\sum_{i=1}^n \left(\lambda^T(\beta) \widehat{E}_i(\beta) \right) = \sum_{i=1}^n \left(\lambda^T(\beta) \widehat{E}_i(\beta) \right)^2 - \sum_{i=1}^n \frac{\left(\lambda^T(\beta) \widehat{E}_i(\beta) \right)^3}{1 + \lambda^T(\beta) \widehat{E}_i(\beta)} \tag{A.6}$$

Let $\widetilde{l}(\beta) = \sum_{i=1}^n \left(\lambda^T(\beta) \widehat{E}_i(\beta) \right)^2$. Using a Taylor expansion and Eq. (A.6) we have that

$$\begin{aligned} l(\beta) &= \sum_{i=1}^n \left(\lambda^T(\beta) \widehat{E}_i(\beta) \right) - \frac{1}{2} \left(\lambda^T(\beta) \widehat{E}_i(\beta) \right)^2 + \frac{1}{3} \left(\lambda^T(\beta) \widehat{E}_i(\beta) \right)^3 + o_p(1) \\ &= \widetilde{l}(\beta) + r_n(\beta) + o_p(1), \end{aligned}$$

uniformly for $\beta \in \mathcal{B}_n$, where $r_n(\beta) = 2 \sum_{i=1}^n \left(\lambda^T(\beta) \widehat{E}_i(\beta) \right)^3 \left(1/3 - 1/(1 + \lambda^T(\beta) \widehat{E}_i(\beta)) \right)$. For any $\beta \in \mathcal{B}_n$, we have that $|r_n(\beta)| \leq kn \|\lambda(\beta)\|^3 (\max_{1 \leq i \leq n} \|\widehat{E}_i(\beta)\|)^3$, for some constant k . It follows by Lemma 4 that $\min_{\beta \in \mathcal{B}_n} |r_n(\beta)| = o_p(1)$. Then the quantities $\min_{\beta \in \mathcal{B}_n} l(\beta)$ and $\min_{\beta \in \mathcal{B}_n} \widetilde{l}(\beta)$ have the same asymptotic distribution and, therefore, it suffices to consider $\min_{\beta \in \mathcal{B}_n} l(\beta)$. Using Eq. (A.5) it follows that, uniformly for $\beta \in \mathcal{B}_n$,

$$\begin{aligned} \widetilde{l}(\beta) &= \lambda^T(\beta) n \widehat{\Psi}(\beta) \lambda(\beta) \\ &= n^{-1} S_n^T(\beta) \Psi_0^{-1} S_n(\beta) + (n^{-1/2} S_n^T(\beta)) \left(\widehat{\Psi}^{-1}(\beta) - \Psi_0^{-1} \right) (n^{-1/2} S_n(\beta)) + o_p(1). \end{aligned}$$

Using Eq. (A.1), Lemma 1, and Lemma 2 it follows that the minimum over $\mathcal{N}$ of the second term above is $o_p(1)$. Note that for $\beta \in \mathcal{B}_n$ we may write $\beta = (\beta_0^{(1)T}, \beta^{(2)T})^T$. We want to minimize over $\beta^{(2)} \in \mathcal{N}$ the quantity $n^{-1} S_n^T(\beta) \Psi_0^{-1} S_n(\beta)$. Using the local linearity property (A.1) this translates to minimizing over $\beta^{(2)} \in \mathcal{N}$ the function

$$Q(\beta^{(2)}) = n B^T \Upsilon \Psi_0^{-1} \Upsilon B = n B^T \Delta B, \quad (\text{A.7})$$

where $B^T = (\widehat{\beta}^{(1)} - \beta_0^{(1)})^T, (\widehat{\beta}^{(2)} - \beta^{(2)})^T$; see Wei, Ying, and Lin (1990). Partition the symmetric matrix $\Delta = \Upsilon \Psi_0^{-1} \Upsilon$ into the block submatrices Δ_{11} , Δ_{12} , and Δ_{22} . Here, for example, $\Delta_{11} = [I_q, 0] \Delta [I_q, 0]^T$, which would be the $q \times q$ top left block matrix. Setting $\partial Q(\beta^{(2)}) / \partial \beta^{(2)} = 0$, we obtain

$$\widehat{\beta}^{(2)} - \beta^{(2)} = -\Delta_{22}^{-1} \Delta_{12} (\widehat{\beta}^{(1)} - \beta_0^{(1)}). \quad (\text{A.8})$$

Substituting $\widehat{\beta}^{(2)} - \beta^{(2)}$ obtained from Eq. (A.8) above into Eq. (A.7), it follows that

$$\begin{aligned} \min_{\beta^{(2)} \in \mathcal{N}} Q(\beta^{(2)}) &= n (\widehat{\beta}^{(1)} - \beta_0^{(1)})^T \{ \Delta_{11} - \Delta_{12} \Delta_{22}^{-1} \Delta_{12} \} (\widehat{\beta}^{(1)} - \beta_0^{(1)}) \\ &= n (\widehat{\beta}^{(1)} - \beta_0^{(1)})^T \{ [I_q, 0] \Upsilon^{-1} \Psi_0 \Upsilon^{-1} [I_q, 0]^T \}^{-1} (\widehat{\beta}^{(1)} - \beta_0^{(1)}) \\ &= Y_n^T \{ [I_q, 0] \Upsilon^{-1} \Gamma \Upsilon^{-1} [I_q, 0]^T \}^{1/2} \{ [I_q, 0] \Upsilon^{-1} \Psi_0 \Upsilon^{-1} [I_q, 0]^T \}^{-1} \\ &\quad \times \{ [I_q, 0] \Upsilon^{-1} \Gamma \Upsilon^{-1} [I_q, 0]^T \}^{1/2} Y_n, \end{aligned}$$

where

$$Y_n = n^{1/2} \{ [I_q, 0] \Upsilon^{-1} \Gamma \Upsilon^{-1} [I_q, 0]^T \}^{-1/2} (\widehat{\beta}^{(1)} - \beta_0^{(1)}).$$

Note that $Y_n \xrightarrow{\mathcal{D}} N(0, I_q)$. Furthermore, note that the $q \times q$ matrices

$$([I_q, 0] \Upsilon^{-1} \Gamma \Upsilon^{-1} [I_q, 0]^T)^{1/2} ([I_q, 0] \Upsilon^{-1} \Psi_0 \Upsilon^{-1} [I_q, 0]^T)^{-1} ([I_q, 0] \Upsilon^{-1} \Gamma \Upsilon^{-1} [I_q, 0]^T)^{1/2}$$

and

$$([I_q, 0] \Upsilon^{-1} \Psi_0 \Upsilon^{-1} [I_q, 0]^T)^{-1} ([I_q, 0] \Upsilon^{-1} \Gamma \Upsilon^{-1} [I_q, 0]^T)$$

have the same eigenvalues. It follows by lemma A.3 of Qin and Tsao (2003) that $l_s(\beta_0^{(1)})$ converges in distribution to $k_1 U_1 + \dots + k_q U_q$ where $k_1, \dots, k_q$ are the eigenvalues of

$$([I_q, 0] \Upsilon^{-1} \Psi_0 \Upsilon^{-1} [I_q, 0]^T)^{-1} ([I_q, 0] \Upsilon^{-1} \Gamma \Upsilon^{-1} [I_q, 0]^T).$$

The proof is complete.

Table 1
Empirical coverage probability (ECP) of the 95% confidence regions for the slope of a linear regression model using the minimum dispersion (MD) statistic and the bootstrap profile empirical likelihood (BPEL). For Model 1 the error distribution is $N(0, Z_2)$, and for Model 2 logistic with mean 0 and scale parameter Z_2 , where Z_2 is standard uniform. CR is the censoring rate and BW is the “bandwidth” for the bootstrap profile EL.

CR	Model 1		Model 2	
	MD	BPEL	MD	BPEL
		BW=0.2	BW=0.2	BW=0.3
10%	97.60%	98.00%	95.10%	93.20%
20%	97.60%	98.00%	95.00%	93.00%
30%	97.90%	98.30%	96.20%	93.90%