Far East J. Theo. Stat. **15**(1) (2005), 53-73

SEMIPARAMETRIC REGRESSION WITH CURRENT STATUS DATA

SHUANGGE MA

Collaborative Health Studies Coordinating Center
Department of Biostatistics, University of Washington
Seattle, WA 98115, U. S. A.
e-mail: shuangge@u.washington.edu

Abstract

Current status data arises in survival analysis and reliability analysis, when a continuous response is reduced to an indicator of whether the response is greater or less than an observed random threshold value. In this article, we study the accelerated failure time (AFT) model with current status data. With empirical processes techniques, we prove that the semiparametric least squares (SLS) estimators are consistent, with convergence rates $n^{1/3}$ and $n^{1/2}$ for estimating the unknown error distribution function and the unknown regression parameter, respectively. Asymptotic normality and inference for the regression parameter based on the weighted bootstrap are also investigated. A simulation study is applied to evaluate the finite sample efficacy and asymptotic properties. We analyze the California Partner Study data for demonstration.

1. Introduction

Current status data (also known as case I interval censored data) often arises in survival analysis and reliability analysis, where the target measurement is the time of occurrence of some event, but observations are limited to indicators of whether or not the event has occurred at the

2000 Mathematics Subject Classification: 62F12, 62N02, 62G09.

Key words and phrases: AFT model, current status data, semiparametric inference.

Received September 21, 2004

time the sample is collected: only the *current status* of each individual with respect to event occurrence is observed. Such data are commonly encountered in biomedicine, economics, biography and sociology. Consider the tumorigenicitys study described in Hoel and Walburg [3]. 144 RFM mice were assigned to either a germ-free or a conventional environment. The purpose of this study was to investigate the environmental effect on the incidence of lung tumors. Since lung tumors cannot be observed before death in RFM mice, it is appropriate to treat this data as current status data. Another example of current status data is the partner study of HIV transmission, which will be discussed in Section 4. For detailed discussions of the history and methodologies for current status data, see Jewell and van der Laan [7].

Let Y and T denote the failure time of interest and the censoring time, respectively. Then a data observation consists of $X = (T, \Delta, Z) \in \mathbb{R}^+ \times \{0, 1\} \times \mathbb{R}^d$, where $\Delta = I_{(Y \le T)}$ indicates whether the event time of interest Y has occurred or not at censoring time T and Z is a d-dimensional real valued covariate. Previous study of semiparametric models with current status data includes the nonlinear regression model in Honda [4], the Cox model in Huang [5], the additive models in Lin, Oakes and Ying [9] and the partly linear transformation model in Ma [10]. In this article, we investigate the accelerated failure time (AFT) model.

The AFT model has been of extensive interest (Shen [12], Abrevaya [1], Aragon and Quiroz [2]) because of its simple format and successes with data analysis. More precisely, the AFT model assumes

$$Y = \beta' Z + \varepsilon, \tag{1}$$

where β is the unknown regression parameter and ε has an unknown distribution function F with support on $\mathbb{R}$. Another commonly used version of the AFT model assumes $\log(Y) = \beta' Z + \varepsilon$, where the *log* function can be replaced with other known monotone transformation functions. See Klein and Moeschberger [8] for reference.

Linear regression with current status data has been studied by Shen

[12] using a sieve approach. It is shown that (β, F) can be consistently estimated. However, Shen [12] also points out that his approach suffers from great computational complexities, especially for high dimensional covariates or large sample size cases. As an alternative, rank based estimator is investigated in Aragon and Quiroz [2] and Abrevaya [1]. With the rank estimator, only β can be estimated. Considering the computational complexity of Shen's approach and the interest in F in a lot of data analysis (Jewell and Shiboski [6], Ma [10]), we study the semiparametric least square (SLS) estimator in this article, which can fill the gap between the rank estimator and the sieve estimator.

The goal is to develop (1) an estimation algorithm, with which (β, F) can be consistently estimated, (2) a computationally algorithm, which is more efficient compared with Shen [12], and (3) an inference procedure for the estimator of β. The motivations of this article come from the following concerns. The best possible convergence rates for estimating β and F are $n^{1/2}$ and $n^{1/3}$, respectively. This has been observed in general for current status data in van der Vaart [15]. It is of theoretical interest to investigate the convergence rates and asymptotic behaviors of the regression parameters for linear regression models. Secondly, since the objective function is not the likelihood, estimation and inference results in Huang [5], van der Vaart [15] and Murphy and van der Vaart [11] do not apply. Thirdly, a computationally efficient algorithm is needed to estimate F under the constraint that F is a distribution function.

The rest of the article is organized as follows. The SLS estimator is proposed in Section 2, followed by the pseudo-information calculation, consistency and convergence rates results. We also discuss asymptotic distribution and inference for the estimator of β in the same section. In Section 3, computational strategy is discussed and a small simulation study is employed to evaluate the efficacy of the proposed estimator. We apply the proposed approach to the analysis of the California Partner Study (CPS) data in Section 4. Concluding remarks are in Section 4. Proofs are provided in the Appendix.

2. Semiparametric Least Squares Estimator

2.1. Estimation

Assume that there are n i.i.d. observations $X_1 = (T_1, \delta_1, Z_1)$, ..., $X_n = (T_n, \delta_n, Z_n)$ generated from model (1). Denote the true value of (β, F) as (β_0, F_0). Since $\Delta = I_{(Y \le T)}$, we have the identity $E(\Delta) = \Pr(Y \le T) = \Pr(\varepsilon \le T - \beta_0' Z) = E(F_0(T - \beta_0' Z))$. Inspired by this, we consider the following semiparametric least squares estimator $(\hat{\beta}, \widetilde{F})$ of (β_0, F_0) defined by

$$(\hat{\beta}, \widetilde{F}) = \arg\min_{\beta, F} \left\{ \sum_{i=1}^{n} (\delta_i - F(T_i - \beta' Z_i))^2 \right\}, \tag{2}$$

under the constraint that F is a distribution function. Note that $\widetilde{F}$ is not well defined out of $[\min\{T_i - \hat{\beta}' Z_i\}, \max\{T_i - \hat{\beta}' Z_i\}]$. So in the downstream analysis, we consider the modified SLS estimator $(\hat{\beta}, \hat{F})$ defined by

$$\hat{F}(u) \equiv \begin{cases} \widetilde{F}(\max\{T_i - \hat{\beta}' Z_i\}), & u > \max\{T_i - \hat{\beta}' Z_i\} \\ \widetilde{F}(\min\{T_i - \hat{\beta}' Z_i\}), & u < \min\{T_i - \hat{\beta}' Z_i\} \\ \widetilde{F}(u), & \text{otherwise.} \end{cases}$$

Other tail definitions can also be adopted. See Klein and Moeschberger [8] for relevant discussions. The following data and model assumptions are needed:

(A1) $\beta_0 \in \mathbb{B}_1$ and $Z \in \mathbb{B}_2$, where $\mathbb{B}_1$, $\mathbb{B}_2$ are known compact subsets of $\mathbb{R}^d$. All components of β_0 are not equal to zero and β_0 is an interior point of $\mathbb{B}_1$.

(A2) The censoring time T and event time Y are conditionally independent given Z.

(A3) $T \in [\tau_l, \tau_u]$, where $-\infty < \tau_l < \tau_u < \infty$.

(A4) $E(\varepsilon) = 0$ and $0 < \mathrm{var}(\varepsilon) < \infty$. ε is independent of (Z, T).

(A5) (i) There exists a fixed $M > 0$, such that $1/M < F_0(T - \beta' Z) < M$. (ii) F has first order derivative f bounded away from 0.

Conditions A1-A4 are standard model assumptions for regression models with current status data, as discussed in Huang [5] and Jewell and van der Laan [7]. Condition A5 (i) is a crucial assumption for properly controlling the size of the nonparametric estimator set. It also guarantees that the partial derivatives are bounded away from 0. This can be seen in the proof of Lemma 2 (in the Appendix). For most distribution functions F with support on $\mathbb{R}$, this can be achieved under the boundedness assumptions A1 and A3. For theoretical proof and computational purpose, the exact value of M is not needed. We only need to assume F_0 is bounded away from 0 and 1. Condition A5 (ii) is a technical assumption needed for the asymptotic consistency of our estimators, as will be discussed in Section 2.3. Intuitively, this condition assumes that the distribution of $T - \beta_0' Z$ is not too sparse for any finite interval in the range of $\{T - \beta' Z\}$. We allow negative T to incorporate certain situations arising in liability analysis and transformation of survival times.

It can be seen from the equation (2) that for any fixed β, the value $\hat{F}$ is determined by the relative rank of the set $\{T_i - \beta' Z_i\}$, instead of the actual values. So as for the rank estimators (Abrevaya [1]), for any finite sample cases, the least squares estimator $\hat{\beta}$ will usually be a finite interval of unions of finite intervals instead of a single value. However, it is expected that if the length of the intervals shrink at least at the rate of $O_p(n^{-1/2-a})$, for a constant $a > 0$, then asymptotic properties of $\hat{\beta}$ will not be affected. This can be achieved under the assumption A5 (ii). The estimators $\hat{F}$ are step functions with jumps at the observed $T_i - \hat{\beta}' Z_i$ only. This is also observed in Huang [5].

2.2. Pseudo-information calculation

It is well known that for semiparametric models, it is not trivial to estimate the regression parameters at the $\sqrt{n}$ rate. For semiparametric maximum likelihood estimators, van der Vaart [15] and van der Vaart and Wellner [16] show that a necessary condition is to have non-singular

information, which corresponds to inverse of the asymptotic variance matrix. For semiparametric maximum likelihood estimators, the information matrix is expressed in an efficient score function. The efficient score functions are equal to the score functions of the regression parameters minus the score functions of the nonparametric parameters, evaluated at the least favorable directions. See van der Vaart [15] for more discussions.

However, the results in van der Vaart [15] are not applicable here, since the target function is not a likelihood. For the SLS estimator, proof of Lemma 3 (in the Appendix) shows that we need to assume non-singular "pseudo-information", which still corresponds to the inverse of the asymptotic variance matrix. However, the information explanation no longer holds. Generally speaking, the pseudo-information cannot be constructed with the projection approach. The pseudo-information can be obtained as follows.

Consider a functional set $\mathbb{A}$ composed of real valued differentiable functions defined on $\mathbb{R}$, where for any $a \in \mathbb{A}$, $\int a = 0$ and for $t \in \mathbb{R}$ small enough $F_0 + ta$ is a distribution function. $\mathbb{A}$ is the functional set composed of "proper perturbations". The properties of the sets of proper to perturbations for maximum likelihood estimators have been discussed in van der Vaart and Wellner [16]. Similar results hold for the perturbation direction for the SLS estimator.

Denote $m = (\Delta - F(T - \beta'Z))^2$. Then the first order partial derivatives of m are $m_1 = 2(\Delta - F(T - \beta'Z))f(T - \beta'Z)Z$ and $m_2(a_1) = -2(\Delta - F(T - \beta'Z))a_1(T - \beta'Z)$, evaluated at the direction $a_1 \in \mathbb{A}$. The second order partial derivatives are

$$m_{11} = 2(\Delta - F(T - \beta'Z))Z^2(f^2(T - \beta'Z) - f^{(1)}(T - \beta'Z)),$$

$$m_{12}(a_1) = 2(\Delta - F(T - \beta'Z))Z(a_1^{(1)}(T - \beta'Z) - a_1(T - \beta'Z)f(T - \beta'Z)),$$

and

$$m_{22}(a_1, a_2) = 2(\Delta - F(T - \beta'Z))a_1^{(1)}(T - \beta'Z)a_2^{(1)}(T - \beta'Z),$$

where in the above equations $a_1, a_2 \in \mathbb{A}$ and the superscript "(1)" denotes the first order derivatives of smooth functions.

Denote $\mathbb{P}_n$ and P as the empirical measure and the expectation, respectively. Following the scheme for semiparametric maximum likelihood estimators, we assume the special perturbation direction a^* for our SLS estimator satisfies $\mathrm{P}[m_{12}(a) - m_{22}(a^*, a)] = 0$ for any $a \in \mathbb{A}$, which is equivalent to $\mathrm{P}((\Delta - F(T - \beta' Z))(Z(a^{(1)} - af) - aa^*)) = 0$. a^* corresponds to the least favorable direction for the maximum likelihood estimators. The pseudo-information is defined as $I^* = \{\mathrm{P}(m_{11} - m_{12}(a^*))\}^{-1} \cdot \mathrm{P}[m_1 - m_2(a^*)]^2 \{\mathrm{P}(m_{11} - m_{12}(a^*))\}^{-1}$. As for the maximum likelihood estimators, we need to assume

(I1) There exists $a^* \in \mathbb{A}$, such that for any $a \in \mathbb{A}$, $\mathrm{P}((\Delta - F(T - \beta' Z))(Z(a^{(1)} - af) - aa^*)) = 0$.

(I2) (Finite variance) $0 < \det(I^*) < \infty$, where det denotes the determinant of a matrix.

In conditions I1 and I2, all the functions are evaluated at (β_0, F_0). Similar information assumptions are made in Huang [5] and van der Vaart and Wellner [16] for maximum likelihood estimators. Although generally speaking checking the non-singularity of I^* is not trivial, it is expected that for most encountered statistical models, the non-singularity can be achieved under mild assumptions (see van der Vaart and Wellner, [15] for reference).

2.3. Asymptotic results

Identifiability of the SLS estimators $(\hat{\beta}, \hat{F})$ can be proved in a similar manner as in Shen [12]. The proof is omitted here. Regularity conditions as those in Shen [12] are also needed for our estimators. We demonstrate the uniqueness of our estimator schematically in Section 3 with a simulation study. The asymptotic properties of the estimator in (2) can be summarized into the following two lemmas. Denote U_L and U_U as the maximum and minimum of the set $\{T - \beta_0' Z\}$, respectively.

Lemma 1 (Consistency and convergence rate)**.** *Under assumptions* A1-A5, *it can be shown that* $\|(\hat{\beta}, \hat{F}) - (\beta_0, F_0)\|_2 = (\|\hat{\beta} - \beta_0\|^2 + \|\hat{F} - F_0\|_2^2)^{1/2} = O_p(n^{-1/3})$, *where* $\|\cdot\|$ *is the usual* L_2 *norm in* $\mathbb{R}^d$ *and* $\|\hat{F} - F_0\|_2^2$ *is defined as* $\int_{U_L}^{U_U} (\hat{F}(\varepsilon) - F_0(\varepsilon))^2 d\varepsilon$.

Note the L_2 norm on the subspace of F is only defined on the interval where F is observable.

Lemma 1 shows that with the SLS estimator, we are able to estimate (β_0, F_0) consistently. As discussed in Huang [5], the best possible convergence rate for estimating F_0 is $n^{1/3}$, which can be achieved with the SLS estimator. Ma [10] shows that uniform consistency of $\hat{F}$ cannot be achieved, unless stronger assumptions are made. The proof is postponed to the Appendix. The key of the proof is that the asymptotic behaviors of the M-estimators depend on the size of the parameter sets, which can be measured with the entropy integrals. See van der Vaart and Wellner [16] for more discussions on convergence rate results of M-estimators. We now investigate the asymptotic distribution of $\hat{\beta}$.

Lemma 2 (Asymptotic normality)**.** *Under assumptions* A1-A5 *and* I1-I2, *it can be shown that* $\sqrt{n}(\hat{\beta} - \beta_0) \to_d N(0, I^{*-1})$.

Although $\hat{F}$ only converges at the $n^{1/3}$ rate, Lemma 2 says $\hat{\beta}$ is still $\sqrt{n}$ consistent and asymptotically normally distributed. The proof is shown in the Appendix. The key is the Donsker properties of the partial derivatives of the object function, in a neighborhood of the true values.

Inference for $\hat{\beta}$ can be built based on the asymptotic normality result given in Lemma 2. However, it can be seen that the asymptotic variance I^{*-1} takes a very complicated form. Plug-in estimators as in Huang [5] will involve estimations of complicated nonparametric functionals, which is computationally and theoretically difficult. As an alternative, we consider the following weighted bootstrap.

2.4. Inference

Denote $w_1, ..., w_n$ as n i.i.d. random variables, satisfying $E(W)=1, 0 < \text{var}(W)=v_0 < \infty$ and $0 < w < L < \infty$ for known positive constants v_0 and L. Denote $(\hat{\beta}^*, \tilde{F}^*)$ as the minimizer of the weighted sum of squared errors, i.e.,

$$(\hat{\beta}^*, \tilde{F}^*) = \arg\min\left\{\sum_{i=1}^{n} w_i(\delta_i - F(T_i - \beta' Z_i))^2\right\}.$$

$\hat{F}^*$ can be defined accordingly, following the definition of $\hat{F}$.

Lemma 3 (Validity of the weighted bootstrap)**.** *Under assumptions* A1-A5 *and* I1-I2, *for the weighted least squares estimator* $(\hat{\beta}^*, \hat{F}^*)$, *we have*:

$$\sqrt{n/v_0}(\hat{\beta}^* - \hat{\beta} \mid X_1, ..., X_n) \to_d \sqrt{n}(\hat{\beta} - \beta_0).$$

This justifies the validity of the following weighted bootstrap inference scheme. n i.i.d positive random weights are first generated. Proper random weights distributions include, but are not limited to, the truncated exponential distribution, the uniform distribution, and the truncated Normal distribution. Then for each set of weights, the weighted least squares estimator $\hat{\beta}^*$ can be computed. This procedure is repeated a number of times, until the variance estimators become stable. After proper scale adjust, the sample variance of $\hat{\beta}^*$ is a consistent estimator of the variance of $\hat{\beta}$.

The advantage of the weighted bootstrap is that estimation of complicated functionals are not needed. Software used to maximize (2) can be used to maximize the weighted sum, with only minor modifications. One drawback of the weighted bootstrap is that multiple computations are involved. The computational cost may be considerable for large sample size cases.

3. Numerical Study

3.1. Computational strategy

For simplicity, we consider the one dimensional case, where $\beta \in [l, u]$ with $-\infty < l < u < \infty$ for known l and u. For any fixed $\beta_1 \in [l, u]$, we assume the order of $\{T_i - \beta_1 Z_i\}$ is

$$T_{(1)} - \beta_1 Z_{(1)} < T_{(2)} - \beta_1 Z_{(2)} < \cdots < T_{(n)} - \beta_1 Z_{(n)}, \tag{3}$$

where $(T_{(i)}, Z_{(i)})$ for $i = 1, ..., n$ are paired observations from $\{(T_i, Z_i), i = 1, ..., n\}$. As we increase (or decrease) the value of β_1, there is a point β_{1*}, such that the order (3) no longer holds. We call points as β_{1*} the switching points, where the *order switches*. Asymptotically under assumption A5 (ii), the number of switching points is of order n^{α}, where $1 \le \alpha \le 2$. The target function $\sum_{i=1}^{n} (\delta_i - F(T_i - \beta' Z_i))^2$ is a constant between any two adjacent switching points.

Computationally, we can search equally spaced over the interval $[l, u]$. As long as the search scale is $o_p(n^{-1/2})$, the convergence rate of β will not be affected. For each fixed β, the estimator for F can be computed with the PAV (pooled-adjacent-violator) algorithm (Stout [13]). For d-dimensional cases, we search over small rectangles of $\mathbb{R}^d$. Since the convergence rate for β cannot exceed $n^{1/2}$ in our case, as long as the volume of the rectangles are of order $o_p(n^{-1/2})$, the asymptotic properties of $\hat{\beta}$ will not be affected. As can be seen from the proof of Lemma 2, we can release the maximization condition to the nearly maximization condition. This also justifies the searching procedure.

When the dimension of covariates increases, the computational complexity increases at a rate much slower than that of Shen [12]. The computational complexity of searching for β is of the same order as the rank estimator. For simulated data with one dimensional covariates and sample size equal to 500, it takes less than 30 seconds to compute the SLS estimator.

3.2. Simulation study

A small simulation study is employed to evaluate the finite sample performance of the proposed estimator. We compare our estimator with the rank based estimator in Abrevaya [1] numerically. Two censoring scenarios as suggested in Abrevaya [1] are considered: continuously distributed censoring and discretely distributed censoring. Again we allow negative event and censoring "times" to incorporate liability data.

Continuous censoring. Consider the following Monte Carlo study with continuously distributed censoring times. Assume ε is $N(0, 1)$ distributed, Z is $N(0, 1)$ the censoring is uniform $[-3.5, 3.5]$ and $\beta_0 = 1$. The continuous outcome of interests is generated from model (1) and independent of the censoring. We first show a plot of the empirical minimums of the SLS target function for each fixed β (minimizer over F) versus β under this scenario, based on sample size 500 and 200 realizations. The convexity of the empirical SLS target function, as shown in Figure 1, justifies the uniqueness of the least square estimator and our "search" approach.

Discrete censoring. Another case we investigate is the discrete censoring. We consider an extreme case, where it is assumed $T = -1$ or 1, each with probability 1/2. Other conditions are the same as for the continuous censoring case.

We show comparison of the least squares estimators with the rank based estimators in Table 1, under both censoring scenarios. We can see that the proposed estimator yields satisfactory results for data sets with sample size as small as 125. The sample standard deviations computed from 200 realizations shrink at a reasonable rate, which supports the $\sqrt{n}$ convergence rate result in Lemma 2. Since we estimate an infinite dimensional parameter, computationally our estimator has larger sample variances in general. In terms of bias and variance, our estimator is comparable to the rank estimator, although an extra infinite dimensional covariate is estimated simultaneously. The asymptotic normality of $\hat{\beta}$ under two censoring scenarios is shown in Figure 1 with two histograms. We also show the estimators of the unknown distribution function F in

Figure 2, with comparisons to the true underlying distribution function. The means of our estimators computed based on 200 realizations match the unknown distribution function very well. Pointwise 95% confidence intervals are also given.

4. CPS Data Analysis

One example of current status data arises naturally in the study of infectious diseases, particularly when infection is an unobservable event, that is one with often no or few clinical indications. The prototypical example is infection with the Human Immunodeficiency Virus (HIV). Consider the California Partner Study (CPS) of HIV infection (Jewell and Shiboski [6]). The most straightforward partner study occurs when HIV infection data is collected on both partners in a long-term sexual relationship. These partnerships are assumed to include a primary infected individual (called the index case) who has been infected via some external source, and a susceptible partner who has no other means of infection other than sexual contact with the index case. Suppose Y denotes the time from infection of the infected case to the infection of the susceptible partner, and that the partnership is evaluated at a single time T after infection of the infected case. Then the infection status of the susceptible partner provides current status data on Y at time T. A schematic representation is available in *Encyclopedia of Biostatistics*.

The partial HIV partner dataset we analyze consists of 302 observations of partners with the male partners as the index cases. The following data analysis is carried out with 295 complete records only. The followup time for the 295 partners ranges from 0.08 year to 14.9 years. 55 partners developed HIV when monitored. The covariate effect of interest is the average sexual contact rate. Previous study in Jewell and Shiboski [6] suggests *log* transformation of the covariate.

We assume $Y = \beta \log(\textit{sexual contact rate}) + \varepsilon$, and use the proposed SLS estimator to estimate β and F, the distribution function of ε. The weighted bootstrap with truncated exponential weights is used for inference. We obtain $\hat{\beta} = -5.6$, which is statistically significant at the 0.05 level. This result supports the common belief that sexual contacts

increase the hazard of HIV transmission significantly. The estimated $\hat{F}$ and its *lowess* smoother is shown in Figure 3.

5. Concluding Remarks

Current status data is studied in this article assuming the AFT model. Empirical processes techniques are applied to prove the convergence rates and asymptotic distribution results. An inference procedure based on the weighted bootstrap is proposed. Simulation studies and data analysis show the satisfactory numerical properties of the proposed approach. We propose a new estimator for studying current status data and show a way of adopting techniques applicable only to maximum likelihood estimators to least squares estimators.

A natural extension of the present model is the nonlinear regression model considered in Honda [4], where it is assumed that $Y = g(Z) + \varepsilon$, for an unknown function g. More generally, we can consider the partly linear additive model, assuming $Y = \beta_1' Z_1 + g(Z_2) + \varepsilon$. To avoid overfitting, we usually make certain smoothness assumptions on g, for example $g \in \Im = \{g : \int (g^{(m)})^2 < \infty\}$, i.e., the Sobolev space indexed by the order of derivative m. Following the approach discussed above, we may consider

$$(\hat{\beta}, \hat{g}, \hat{F}) = \arg\min\left\{\sum_{i=1}^{n} (\delta_i - F(T_i - \beta' Z_{1i} - g(Z_{2i})))^2\right\}. \tag{4}$$

The key in investigating the estimator (4) is that the size of the Sobolev space, which is measured with the entropy integral, is available. Entropy calculations for the subspace (g, F) can be found in van de Geer [14]. It is expected that results similar to those in the current article can be obtained.

The estimators of the distribution function F are step functions, with jumps at $\{T_i - \hat{\beta}' Z_i\}$. Practically speaking, it is possible to make stronger assumptions on F, for example, F can be assumed to belong to a known Sobolev space, or a well defined parametric subsets. With stronger assumptions, it is possible to improve the convergence rates of $\hat{F}$. Similar discussions can be found in van de Geer [14].

Appendix

We provide sketches of the proofs of Lemmas 1-3. Certain irrelevant regularity conditions are omitted. Denote the set of distribution functions satisfying condition A5 as $\mathbb{F}$. Define the norm $\|\cdot\|_2$ on the set $\mathbb{B}_1 \times \mathbb{F}$ as $\|(\beta_1, F_1) - (\beta_2, F_2)\|_2^2 = (\|\beta_1 - \beta_2\|^2 + \int_{U_L}^{U_U}(F_1(\varepsilon) - F_2(\varepsilon))^2 d\varepsilon)$. This is just a natural extension of the L_2 norm.

Definition (Bracketing number)**.** Given two functions l and u, the bracket $[l, u]$ is the set of all functions f with $l \le f \le u$. An ε bracket is a bracket $[l, u]$ with $\|l - u\| < \varepsilon$. The bracketing number $N_{[]}(\varepsilon, \mathbb{F}, \|\cdot\|)$ is the minimum number of ε brackets needed to cover $\mathbb{F}$. The entropy with bracketing is the logarithm of the bracketing number. For a given norm $\|\cdot\|$, define a bracketing integral of a class of functions $\mathbb{F}$ as

$$J_{[]}(\delta, \mathbb{F}, \|\cdot\|) = \int_0^{\delta} \sqrt{1 + \log N_{[]}(\varepsilon \|F\|, \mathbb{F}, \|\cdot\|)}\, d\varepsilon, \quad \text{or} \tag{5}$$

Technical tool. Under conditions A1-A6, there exists a constant C such that, for every $\eta > 0$, and for $\mathbb{B}_1 \times \mathbb{F}$, we have $\log N_{[]}(\eta, \mathbb{B}_1 \times \mathbb{F}, \|\cdot\|_2) \le C\left(\frac{1}{\eta}\right)$. For a proof, see Lemma 25.84 of van der Vaart [15].

Proof of Lemma 1.

Consistency. The parameter set for β is compact by assumption A1, and the parameter set for F is compact relative to the weak topology. Under certain mild regularity conditions as those in Shen [12], the maximum of the target function is "well separated" in the sense defined in van der Vaart [15]. Consistency for the semiparametric least squares estimator can be obtained by the Argmax theorem in van der Vaart and Wellner [16].

Convergence rate. We apply Theorem 3.2.1 of van der Vaart and Wellner [16] here. $P[m(\beta, F)]$ is maximized at $\beta = \beta_0$ and $F = F_0$. So its first order partial derivatives at (β_0, F_0) are equal to 0. Considering the

identifiability results shown in Shen [12] and the special format of m_1 and m_2, we can combine the boundedness conditions A1 and A5 and get

$$\mathrm{P}[m(\beta_0, F_0) - m(\beta, F)] \le -k_1 \| (\beta_0, F_0) - (\beta, F) \|_2^2,$$

for a positive constant k_1 in a small neighborhood of (β_0, F_0). Thus

$$\sup_{\eta/2 \le \| (\beta_0, F_0) - (\beta, F) \|_2 \le \eta} \mathrm{P}[m(\beta_0, F_0) - m(\beta, F)] \le -\frac{k_1}{4} \eta^2.$$

With the results shown in the Technical tool and Lemma 3.2.2 of van der Vaart and Wellner [16],

$$\mathrm{P}^* \sup_{\| (\beta_0, F_0) - (\beta, F) \|_2 \le \eta} |\sqrt{n}(\mathbb{P}_n - P)(m(\beta_0, F_0) - m(\beta, F))|$$

$$= O_p(1)\eta^{1/2}\left(1 + \frac{\eta^{1/2}}{\eta^2\sqrt{n}} k_2\right),$$

for a positive constant k_2 and here P^* denotes the outer expectation. Denote $\phi_n(\eta) = \eta^{1/2}\left(1 + \frac{\eta^{1/2}}{\eta^2\sqrt{n}} k_2\right)$. Then $\phi_n(\eta)/\eta$ is a decreasing function, with $n^{2/3}\phi_n(n^{-1/3}) = O(\sqrt{n})$, as $n \to \infty$. Hence the conditions in Theorem 3.2.1 of van der Vaart and Wellner [16] are satisfied. This implies that $\|(\beta_0, F_0) - (\hat{\beta}, \hat{F})\|_2^2 = O_p(n^{-2/3})$. So Lemma 1 holds.

Proof of Lemma 2. The asymptotic behaviors of $\hat{\beta}$ are built on the convergence rates results and the pseudo-information calculation. The key is to investigate the Donsker properties of the empirical target function in a small neighborhood of (β_0, F_0). The maximization condition in (2) can be released to the following "nearly maximization" condition.

$$\mathbb{P}_n[m_1(\hat{\beta}, \hat{F})] = o_p(n^{1/2}), \quad \text{and} \quad \mathbb{P}_n[m_2(a^*)(\hat{\beta}, \hat{F})] = o_p(n^{1/2}).$$

For the estimators exactly maximize the target function, the $o_p(n^{-1/2})$ in the above equations can be replaced by 0. We also notice that $\mathrm{P}m_1(\beta_0, F_0) = 0$ and $\mathrm{P}m_2(a^*)(\beta_0, F_0) = 0$.

Lemma 1 shows $\|\hat{\beta} - \beta_0\| = O_p(n^{-1/3})$ and $\|\hat{F} - F_0\|_2 = O_p(n^{-1/3})$. Combining the entropy results in the Technical tool with Lemma 3.2.2 of van der Vaart and Wellner [16], we can conclude for any $\eta_n \to 0$ and $C_1 > 0$,

$$\sup_{\|\beta-\beta_0\|\le\eta_n, \|F-F_0\|_2\le C_1 n^{-1/3}} |\sqrt{n}(\mathbb{P}_n - \mathrm{P})(m_1(\beta, F) - m_1(\beta_0, F_0))| = o_p(1)$$

and

$$\sup_{\|\beta-\beta_0\|\le\eta_n, \|F-F_0\|_2\le C_1 n^{-1/3}} |\sqrt{n}(\mathbb{P}_n - \mathrm{P})(m_2(a^*)(\beta, F) - m_2(a^*)(\beta_0, F_0))| = o_p(1).$$

Combining the above equations, we can get the following identity

$$\sqrt{n}\mathbb{P}_n[m_1(\hat{\beta}, \hat{F}) - m_1(\beta_0, F_0)] = o_p(1)$$

and

$$\sqrt{n}\mathbb{P}_n[m_2(a^*)(\hat{\beta}, \hat{F}) - m_2(a^*)(\beta_0, F_0)] = o_p(1). \tag{6}$$

From Taylor expansion, we have for $\|\beta - \beta_0\| \le \eta_n$ and $\|F - F_0\| \le C_1 n^{-1/3}$,

$$\left|\mathrm{P}\left\{m_1(\beta, F) - m_1(\beta_0, F_0) - m_{11}(\beta_0, F_0)\times(\beta-\beta_0) - m_{12}\left(\frac{F-F_0}{\|F-F_0\|_2}\right)(\beta_0, F_0)\times\|F-F_0\|\right\}\right| = o_p(\|\beta-\beta_0\|) + O_p(\|F-F_0\|_2^2),$$

and

$$\left|\mathrm{P}\left\{m_2(a^*)(\beta, F) - m_2(a^*)(\beta_0, F_0) - m_{21}(a^*)(\beta_0, F_0)\times(\beta-\beta_0) - m_{22}\left(a^*, \frac{F-F_0}{\|F-F_0\|_2}\right)(\beta_0, F_0)\times\|F-F_0\|\right\}\right| = o_p(\|\beta-\beta_0\|) + O_p(\|F-F_0\|_2^2). \tag{7}$$

The equations (6) and (7) give

$$\text{(a) } \mathrm{P}m_{11}(\beta_0, F_0) \times (\hat{\beta} - \beta_0) + \mathrm{P}m_{12}\left(\frac{F - F_0}{\| F - F_0 \|_2}\right)(\beta_0, F_0) \times \| F - F_0 \|$$

$$- \mathbb{P}_n m_1(\beta_0, F_0) + o_p(\| \beta - \beta_0 \|) + O_p(\| F - F_0 \|_2^2) = o_p(n^{-1/2}),$$

and

$$\text{(b) } \mathrm{P}m_{21}(a^*)(\beta_0, F_0) \times (\hat{\beta} - \beta_0) + \mathrm{P}m_{22}\left(a^*, \frac{F - F_0}{\| F - F_0 \|_2}\right)(\beta_0, F_0) \times \| F - F_0 \|$$

$$- \mathbb{P}_n m_2(a^*)(\beta_0, F_0) + o_p(\| \beta - \beta_0 \|) + O_p(\| F - F_0 \|_2^2) = o_p(n^{-1/2}). \quad (8)$$

Note that from the definition of the pseudo-information, we have $\mathrm{P}[m_{12}(a) - m_{22}(a, a^*)] = 0$ for any $a \in \mathbb{A}$. So (8)(a) minus (8)(b) gives

$$[\mathrm{P}m_{11}(\beta_0, F_0) - \mathrm{P}m_{21}(a^*)(\beta_0, F_0)] \times (\hat{\beta} - \beta_0) + o_p(\| \beta - \beta_0 \|)$$

$$= -\mathbb{P}_n[m_1(\beta_0, F_0) - m_2(a^*)(\beta_0, F_0)] + o_p(n^{-1/2}).$$

So it follows that

$$\sqrt{n}(\hat{\beta} - \beta_0) = -[\mathrm{P}m_{11}(\beta_0, F_0) - \mathrm{P}m_{21}(a^*)(\beta_0, F_0)]^{-1} \mathbb{P}_n[m_1(\beta_0, F_0)$$

$$- m_2(a^*)(\beta_0, F_0)] + o_p(1).$$

Combining this result with the pseudo-information assumptions I1-I2, we can conclude that Lemma 2 holds.

Proof of Lemma 3. First it can be seen that the entropy results shown in the Technical tool also hold for the subspace $w\mathbb{F}$, where w denotes the positive random weights discussed in Section 2.4. Since the weights are independent of the data and (β, F), we can conclude $\mathrm{P}(wm) = \mathrm{P}(m)$. This result also holds for other functions of (X_i, β, F).

Careful investigation of the proof of Lemma 2 shows that, if we replace $\mathrm{P}(\cdot)$ with $\mathrm{P}(w \times \cdot)$, and $\mathbb{P}_n(\cdot)$ with $\mathbb{P}_n(w \times \cdot)$, then all the arguments hold. This leads to the conclusion that

$$\sqrt{n}(\hat{\beta}^* - \beta_0) = -[\mathrm{P}m_{11}(\beta_0, F_0) - \mathrm{P}m_{21}(a^*)(\beta_0, F_0)]^{-1}\mathbb{P}_n[w(m_1(\beta_0, F_0) - m_2(a^*)(\beta_0, F_0))] + o_p(1).$$

Considering the independence of w and $\mathrm{var}(w) = v_0$, the above results lead to the validity of the weighted bootstrap.

Acknowledgment

The author was supported in part by the MESA project funded by the National Heart Lung and Blood Institute of the National Institutes of Health. The author would like to thank Dr. Jason P. Fine (University of Wisconsin) for several insightful discussions.

References

[1] J. Abrevaya, Rank regression for current status data: asymptotic normality, Statistics and Probability Letters 43 (1999), 275-287.

[2] J. Aragon and A. J. Quiroz, Rank regression for current status data, Statistics and Probability Letters 24 (1995), 251-256.

[3] D. G. Hoel and H. E. Walburg, Statistical analysis of survival experiments, J. National Cancer Institute 49 (1972), 361-372.

[4] T. Honda, Nonparametric regression with current status data, Unpublished Manuscript, 2002.

[5] J. Huang, Efficient estimation for the proportional hazard model with interval censoring, The Annals of Statistics 24 (1996), 540-568.

[6] N. P. Jewell and S. C. Shiboski, Statistical analysis of HIV infectivity based on partner studies, Biometrics 46 (1990), 1133-1150.

[7] N. P. Jewell and M. J. van der Laan, Current Status Data: Review, Recent Developments and Open Problems, Technical Report, Department of Biostatistics, University of California at Berkeley, 2002.

[8] J. P. Klein and M. L. Moeschberger, Survival Analysis: Techniques for Censored and Truncated Data, Springer, 1997.

[9] D. Y. Lin, D. Oakes and Z. Ying, Additive hazards regression with current status data, Biometrika 85 (1998), 289-298.

[10] S. Ma, Penalized *M*-estimations with Current Status Data, Ph.D. thesis, University of Wisconsin, 2004.

[11] S. A. Murphy and A. W. van der Vaart, Semiparametric likelihood ratio inference, Annals of Statistics 25 (1997), 1471-1509.

[12] X. Shen, Linear regression with current status data, JASA 95 (2000), 842-852.

[13] Q. F. Stout, Optimal algorithms for unimodal regression, Computing Science and Statistics 32 (2000), 348-355.

[14] S. van de Geer, Empirical Processes in *M*-estimation, Cambridge University Press, Cambridge, 2000.

[15] A. W. van der Vaart, Asymptotic Statistics, Cambridge University Press, 1998.

[16] A. W. van der Vaart and J. A. Wellner, Weak Convergence and Empirical Processes: with Applications to Statistics, Springer, New York, 1996.

Table 1. Comparison of sample mean and standard deviation (s.d.) of the SLS (semiparametric least squares) estimator with the rank based estimator. Based on 200 realizations

		SLS	Rank
	Sample Size	Mean (s.d.)	Mean (s.d.)
Continuous Censoring	125	1.011 (0.278)	0.985 (0.219)
	250	1.013 (0.189)	1.029 (0.153)
	500	0.995 (0.126)	1.014 (0.099)
Discrete Censoring	125	1.033 (0.231)	1.018 (0.216)
	250	1.015 (0.175)	1.002 (0.116)
	500	1.002 (0.116)	0.998 (0.096)

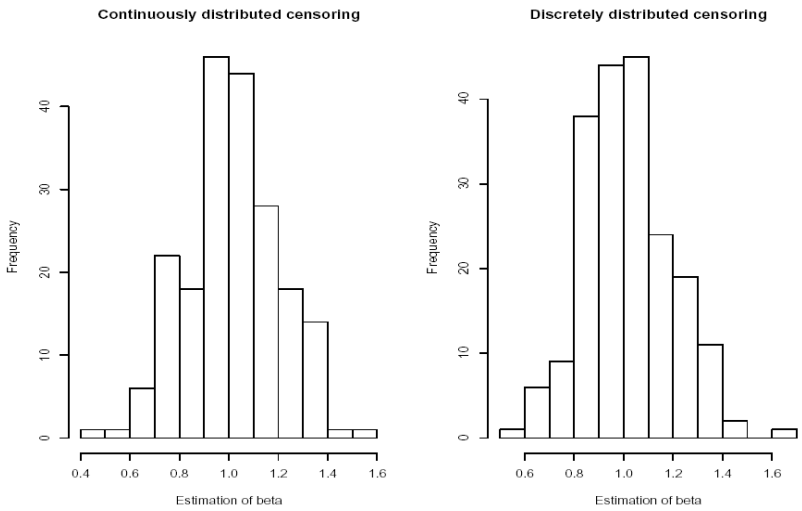

Figure 1. The histograms of $\hat{\beta}_{250}$ for the semiparametric least squares estimators under continuous and discrete censoring. Sample size is equal to 250, with 200 realizations

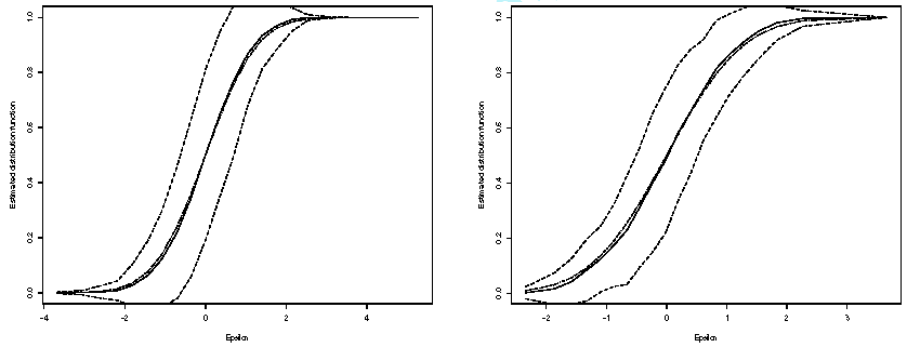

Figure 2. The estimated distribution function of ε: the solid line is the true (unknown) distribution function. The dashed lines are mean of estimated distribution functions and corresponding pointwise 95% confidence intervals. Sample size is equal to 250, based on 200 realizations. Left: continuous censoring. Right: discrete censoring

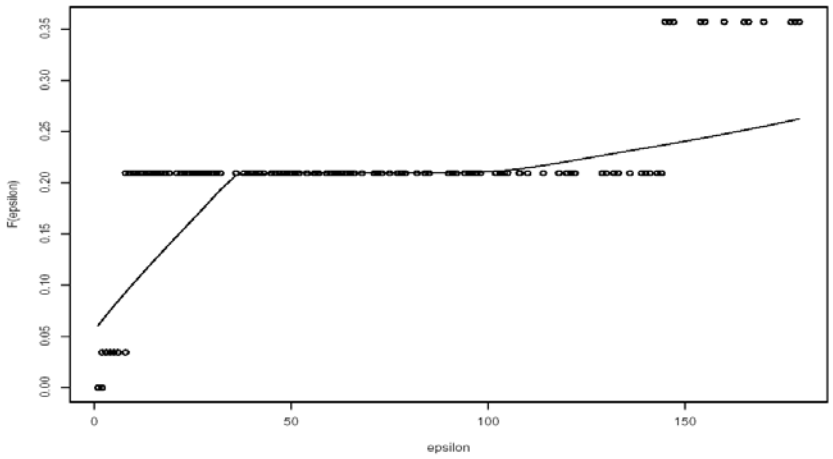

Figure 3. Estimation of the error distribution function for the CPS data, with its lowess smoother

■