应用概率统计 第二十二卷
第二期 2006 年 5 月

Chinese Journal of Applied Probability
and Statistics Vol. 22 No. 2 May 2006

Censored Regression in Semiparametric Transformation Models for Survival Data

ZHAO XIAOBING

(*School of Science, Southern Yangtze University, Wuxi, 214063*)

Abstract

We consider a class of semi-parametric transformation models, under which an unknown transformation of the survival time is linear related to the covariates with various error distributions, which are parametrically specified with unknown parameters. Estimators of the coefficients of covariates are obtained from linear least squares procedures and the Class-K method with censored observations. We show that the estimators are consistent and asymptotically normal. This transformation model, coupled with proposed inference procedures, provides many alternatives to the Cox models and survival analysis.

Keywords: Proportional hazards model, semi-parametric transformation, censored regression, least squares, Class-K method.

AMS Subject Classification: 62G05, 62N01.

§ 1. Introduction

Consider the following semiparametric transformation model introduced by Cox (1972)

$$h(T) = -z^\top \beta + \varepsilon, \tag{1.1}$$

where $h(\cdot)$ is assumed to be a smooth, invertible and strictly monotonically increasing function on R^1, z is a $q \times 1$ covariate vector, β is a $q \times 1$ coefficient vector, and ε has distribution W with density $w > 0$ on R^1. The response T is continuous, z and β are bounded. Generally, we assume that the distribution of ε does not vary with z. The data are generated by n i.i.d. samples of (z_i, ε_i) with $T_i = h^{-1}(-z_i^\top \beta + \varepsilon_i)$, for $i = 1, 2, \cdots, n$. Our focus is the estimation of β by using the resulting n copies of (T, z).

Many methods proposed to estimate β vary with the different treatment of $h(\cdot)$ and a distinct distribution of W. With h specified up to a finite-dimensional parameter vector, model (1.1) has been discussed extensively by Box and Cox (1964). For a completely unspecified h, rank method for analysis failure time data in model (1.1) has been proposed, for example, by Cuzick (1988), an alternative method based on estimating equations for β in model (1.1) was proposed by Cheng et al. (1995, 1997), Cheung and Fine (2001), among others.

On the other hand, if W, a distribution function for ε, is the extreme value distribution $W(s) = 1 - \exp\{-\exp(s)\}$, model (1.1) is the proportional hazards model. Rank likelihood estimation has

Received 2004. 12. 27.

been proposed to estimate regression coefficients by Doksum (1987). If W is the standard logistic distribution, model (1.1) is the proportional odds model. Approaches, such as, modified likelihood approach by Pettitt (1984), likelihood sampler method by Dabrowska and Doksum (1988), profile likelihood by Murphy et al. (1997), and recently, sieve likelihood proposed by Shen (1998), are used to investigate proportional odds model. If W is the standard normal distribution, model (1.1) is the probit model (Fine and Bosch, 2000).

Although semiparameteric transformation models have been extensively applied to analyze survival data, the inferences are essentially based on the likelihood function, and the inference for the monotone transformation $h(\cdot)$ (hence the baseline hazards rate) is after the estimation for β. In this paper, we propose a different approach: an estimator of the monotone transformation $h(\cdot)$ is given firstly, then based on a transformation of the observed data, such as Class-K method, an estimator of coefficients for covariates is derived from the ordinary linear least squares procedure and the large sample properties are also obtained. This new "two-step" estimation procedure provides a simple and effective methodology to analyze survival data with covariates.

We next describe briefly the structure of the article. In section 2 we give a specification of semiparametric transformation models for analysis survival data. And in section 3, an estimation of $h(\cdot)$ will be given. And the linear least squares estimators and their large sample properties are obtained in sections 4-5. A small simulation preformed in section 6 is given to illustrate our inference.

§2. Model Specification

Following the usual formulation in survival analysis, we postulate a "true" survival time T_i for each individual i, which is only observed if it does not exceed individual i's censoring time c_i; otherwise, we observe c_i. We also know whether an individual i is censored or not. This is recorded in a censoring indicator δ_i, with $\delta_i = 1$ if T_i is an actual failure time (uncensored) and $\delta_i = 0$ if T_i is censored. The observable survival time Y_i, possibly censored, is then given by $Y_i = T_i \wedge c_i = \min(T_i, c_i)$, $i = 1, \cdots, n$. The censoring indicators can be written as

$$\delta_i = \begin{cases} 1, & \text{if } T_i \le c_i. \\ 0, & \text{otherwise,} \end{cases} \qquad i = 1, \cdots, n.$$

In general, both T_i and c_i are random variables, so are Y_i and δ_i. We further assume that T_i is independent of c_i for each i, and (T_i, c_i), $i = 1, \cdots, n$, are mutual independent pairs. In addition, c_i are assumed to have the same cumulative distribution function G, which is referred to as an independent and identically distribution (i.i.d.) censoring model. The distribution of T_i, on the other hand, are not necessarily identical, and may depend on such covariates as age, gender, treatment method, etc.

Let T denote a nonnegative random variable representing the survival time of an individual, and $F(t)$ be the cumulative distribution function (cdf) of T. Let z be a $q \times 1$ vector of covariates associated with the survival time T of an individual under study. The Cox Proportional Hazards (PH) model specifies the survival function of T with covariate vector z by

$$S(t, z) = (1 - F_0(t))^{\exp(z^\top \beta)}, \tag{2.1}$$

where $F_0(t)$ is a baseline survival function independent of covariates, $\beta = (\beta_1, \cdots, \beta_q)^\top$ is an unknown vector of regression parameters (coefficients of covariates) to be estimated.

Suppose that T_i is a survival time with distribution $F_i(t) = 1 - (1 - F_0(t))^{\exp(z_i^\top \beta)}$, $i = 1, 2, \cdots, n$, then

$$\log\{-\log[1 - F_i(t)]\} = z_i^\top \beta + \log\{-\log[1 - F_0(t)]\}. \tag{2.2}$$

Let $h(t) = \log\{-\log[1 - F_0(t)]\}$. Then for a random variable T_i with distribution function $F_i(t)$, by (2.2) we can write

$$h(T_i) = -z_i^\top \beta + \varepsilon_i, \tag{2.3}$$

where $\{\varepsilon_i\}$ are i.i.d. with the distribution function $W(x) = 1 - \exp\{-\exp(x)\}$, $x \in (-\infty, \infty)$.

From (2.3), we say, the independent random variables $T_1, T_2, \cdots, T_n$ follow a linear transformation model. In this paper, we investigate a semiparametric transformation models with a unknown $h(t)$ and the distribution functions of ε_i is known. From the definition of $h(t)$ in model (2.3) we know that $h(t)$ is an increasing monotone function.

Following Doksum (1987), we can and will reparameterize to have $\sum_{i=1}^{n} z_{ij} = 0$, $j = 1, 2, \cdots, q$. Here, to unify model (2.3), is a summary of assumptions and notation:

$$\begin{aligned} &h(T_i) = -z_i^\top \beta + \varepsilon_i, \qquad i = 1, 2, \cdots, n, \\ &\sum_{i=1}^{n} z_{ij} = 0, \qquad z = (z_{ij})_{n \times q} \text{ has rank } q, \end{aligned} \tag{2.4}$$

where $h(t)$ is increasing on the real line and $W(x)$, a distribution function of ε, is absolutely continuous and with a density function $w(x)$ that satisfies $w(x) > 0$, $x \in R$.

§ 3. Estimation of the Transformation

In order to develop linear least squares regression analysis, we shall first give a consistent estimator of $h(t)$. Following Doksum (1987) we write

$$h(t_i) = -\mu_i + \varepsilon_i, \qquad \mu_i = \sum_{j=1}^{q} z_{ij} \beta_j,$$

and let F_i denote the distribution of T_i and W the distribution of ε_i. Note that $F_i(t) = \mathsf{P}(T_i \le t) = \mathsf{P}(h(T_i) \le h(t)) = W(h(t) + \mu_i)$. Then, we can write $h(t) = W^{-1}\{F_i(t)\} - \mu_i$. In our parameterization $\sum_{i=1}^{n}\mu_i=0$; thus

$$h(t)=\frac{1}{n}\sum_{i=1}^{n}W^{-1}(F_i(t)). \tag{3.1}$$

We assume that the μ's are not all zero.

Doksum (1987) gives two consistent estimators of $h(t)$ with some uncensored data, this paper, we extend the results of Doksum (1987) from uncensored survival data to possibly censored survival data.

In order to avoid a lengthy technical discussion about the tail behavior, we make the following assumptions (see also, Murphy et al., 1997): For τ_0, a finite time point, assume that $\mathsf{P}\{c\ge\tau_0\}=\mathsf{P}\{c=\tau_0\}>0$. That is, the study ends at time τ_0, and any remaining live individuals are considered censored at time τ_0. Also assume that on average, some individuals are at risk at time τ_0, that is, $\mathsf{P}\{T>\tau_0\}>0$. Finally, for any possible covariate pattern, the chance of observing a survival time should be positive, that is, $\mathsf{P}\{T\le c|z\}>0$ almost surely. For convenience, denote those assumptions by

$$\mathsf{P}\{c\ge\tau_0\}=\mathsf{P}\{c=\tau_0\}>0,\qquad \mathsf{P}\{T>\tau_0\}>0,\qquad \mathsf{P}\{T\le c|z\}>0. \tag{3.2}$$

3.1 Fixed parameters

We consider the nonlocal case (see Doksum, 1987) with β_j and μ_i fixed as sample size increases. From (3.1) we see that if we can estimate the F_i, then we can estimate $h(t)$. This can be done in analysis of variance models with several observations per cell. These models can be written as

$$h(t_{jk})=\theta_j+\varepsilon_{jk},\qquad k=1,2,\cdots,n_j,\ j=1,2,\cdots,q,$$

where θ_j and n_j are the mean and sample size in cell j, respectively. Now define $\lambda_{jn}=n_j/n$, let F_j denote the distribution of t_{jk} and let $\widehat{F}_j$ be the Kaplan-Meier estimator of F_j in cell j. Assume that the following arguments hold

$$\lim_{n\to\infty}\lambda_{jn}=\lambda_j,\qquad 0<\lambda_j<1,\ j=1,2,\cdots,q.$$

Now we can write $h(t)=\sum_{j=1}^{q}\lambda_{jn}W^{-1}(F_j(t))$ and our estimator of $h(t)$ is defined by

$$\dot{h}(t)=\sum_{j=1}^{q}\lambda_{jn}W^{-1}(\widehat{F}_j(t)).$$

Throughout the rest of this paper, we will assume that $\widetilde{F}(t)=\lim_{n\to\infty}n^{-1}\sum_{i=1}^{n}F_i(t)$ exists for all $t\in[0,\tau_{F_0}]$, where $\tau_{F_0}=\sup\{t:F_0(t)<1\}$ is the right extreme point of F_0. Let $\widetilde{Y}=\max_{1\le i\le n}Y_i$ and denote $h(\cdot\wedge\widetilde{Y})$ by $h_{\widetilde{Y}}(\cdot)$. Define $S_j(t)=1-F_j(t)$,

$$C_j(t)=\int_0^t\frac{\mathrm{d}F_j(u)}{[1-F_j(u)]^2[1-G(u)]}\qquad\text{and}\qquad U(t)=\int_0^t\frac{\mathrm{d}G(u)}{[1-G(u)]^2[1-\widetilde{F}(u)]},$$

where G is the censoring distribution function. We establish weak convergence of $\dot{h}_{\widetilde{Y}}(t)$ in what follows.

Theorem 1 Suppose that W has a continuous derivative w bounded away from 0 and ∞ on R^1. Assume that $\tau_{F_0} \le \tau_G$ and $\sup_i \int_0^{\tau_{F_0}} [1 - G(u-)]^{-1} \mathrm{d}F_i(u) < \infty$. Then the process $\sqrt{n}[\dot{h}_{\widetilde{Y}}(t) - h_{\widetilde{Y}}(t)]$ converges weakly on $(0, \tau_0]$ to the Gaussian process

$$\sum_{j=1}^{q} \frac{\sqrt{\lambda_j} S_j(t) B_j(C_j(t))}{w(W^{-1}(F_j(t)))},$$

where $B_1(\cdot), \cdots, B_n(\cdot)$ are independent standard Brownian Motion processes.

Proof Write $u_{jn} = \widehat{F}_j(t)$, $u = F_j(t)$ and

$$D_{jn} = \sqrt{n}[W^{-1}(\widehat{F}_j(t)) - W^{-1}(F_j(t))] = \frac{W^{-1}(u_{jn}) - W^{-1}(u)}{u_{jn} - u} \sqrt{n}[u_{jn} - u].$$

Note that $D_{jn}(t)$ converges weakly to $B_j(F_j(t))/\sqrt{\lambda_j} w(W^{-1}(F_j(t)))$. Then for any $t \in (0, \tau_0]$, this together with Theorem 3.1 of Gill (1983) implies that $D_{jn}(t \wedge \widetilde{Y})$ converges weakly to $S_j(t) \cdot B_j(C_j(t))/\sqrt{\lambda_j} w(W^{-1}(F_j(t)))$. As a result,

$$\sqrt{n}[\dot{h}_{\widetilde{Y}}(t) - h_{\widetilde{Y}}(t)] = \sum_{j=1}^{q} D_{jn}(t \wedge \widetilde{Y}) \xrightarrow{d} \sum_{j=1}^{q} \frac{\sqrt{\lambda_j} S_j(t) B_j(C_j(t))}{w(W^{-1}(F_j(t)))}. \qquad \#$$

Note that $W^{-1}(F_j(t)) = h(t) - \mu_j$, hence by Theorem 1, $\dot{h}(t)$ is approximately normally distributed with mean $h(t)$ and variance $\sum_{j=1}^{q} \lambda_j S_j(t)^2 C_j(t)/w^2(h(t) - \mu_j)$.

3.2 Local functions space

The method in section 3.1 requires multiple observations per cell. If that is not available for a given data set, we may adopt an alternative approach described below. Following the notations of Doksum (1987), if $\mathsf{E}(\varepsilon_i)$ exists, without loss of generality we can view $\mu_i = \sum_{j=1}^{q} z_{ij}\beta$ as the mean of $h(y)$. Moreover, by (2.4), $\overline{\mu} = n^{-1} \sum_{i=1}^{n} \mu_i = 0$. We will assume that β belongs to the set

$$\Omega_n = \Big\{\beta : \sum_{i=1}^{n} \mu_i^2 \le C^2, \max_{1\le i\le n} |\mu_i| \to 0\Big\}, \tag{3.3}$$

where C^2 is a constant independent of n, while β and μ_i may depend on n although this is suppressed in the notations.

Define $\overline{L}(\cdot) = 1 - L(\cdot)$ for any distribution function $L(\cdot)$. Define $F^{(n)}(y) = n^{-1}\#[Y_i > y]$ to estimate each $\overline{H}_i(y)$, and let $\widehat{\overline{G}}(y)$ be the Kaplan-Meier estimator for the censoring survival function $\overline{G}$. Because for each y,

$$\overline{W}(h(y) - \mu_i) = \overline{F}_i(y) = \overline{H}_i(y)/\overline{G}(y),$$

where $\overline{H}_i(y) = \mathsf{P}(Y_i > y)$, $h(y)$ can be estimated naturally by

$$\dot{h}(y) = \overline{W}^{-1}(F^{(n)}(y)/\widehat{\overline{G}}(y)).$$

Theorem 2 Suppose that W has a continuous derivative w bounded away from 0 and ∞ on R^1, and that G is continuous on $[0,\tau_G]$. Assume $\tau_{F_0}\le\tau_G$ and $\displaystyle\int_0^{\tau_G}[1-\widetilde{F}(u-)]^{-1}\mathrm{d}G(u)<\infty$. Then the process $\sqrt{n}[\dot{h}_{\widetilde{Y}}(y)-h_{\widetilde{Y}}(y)]$ converges weakly on $(0,\tau_0)$ to a Gaussian process.

Proof Note that, for any $t\in(0,\tau_0)$, $\sqrt{n}(\widehat{\overline{G}}_{\widetilde{Y}}-\overline{G}_{\widetilde{Y}})$ converges weakly to a Gaussian process $[1-G(t)]B(U(t))$ (Shorach and Wellner, 1986, p. 329), where B is a standard Brownian process. Let $v_n=F^{(n)}(y)/\widehat{\overline{G}}(y)$, $v=\overline{W}(h(y))$, and

$$D_n(y)=\sqrt{n}[\dot{h}(y)-h(y)]=\frac{\overline{W}^{-1}(v_n)-\overline{W}^{-1}(v)}{v_n-v}\sqrt{n}[v_n-v].$$

Then

$$\begin{aligned}v_n-v &= \frac{\widehat{W}_n(h(y),\mu)}{\widehat{\overline{G}}(y)}-\overline{W}(h(y))\\ &= n^{-1}\sum_{i=1}^{n}\Big[\frac{I[\varepsilon_i>(h(y)-\mu_i)]}{\widehat{\overline{G}}(y)}-\overline{W}(h(y)-\mu_i)\Big]\\ &\quad+n^{-1}\sum_{i=1}^{n}[\overline{W}(h(y)-\mu_i)-\overline{W}(h(y))].\end{aligned}$$

By the results in Appendix B of Cheng et al. (1997, p. 234), the first term converges uniformly to a zero mean Gaussian process. Let $a_i=|\overline{W}(h(y)-\mu_i)-\overline{W}(h(y)|$. By the Taylor expansion, $a_i=w(h(y_0))\mu_i$ with $|h(y)-h(y_0)|\le|\mu_i|$ (see Doksum, 1987, p. 341). This implies that the second term converges uniformly to zero. Finally, note that $[W^{-1}(v_n)-W^{-1}(v)]/(v_n-v)$ converges to $1/wW^{-1}(\widetilde{F}(y))$ as in the proof of Theorem 1. The proof is thus complete. #

Remark 1 In any sample, $\dot{h}(y)=-\infty$ for $y\le Y_{(1)}$ and $=\infty$ for $y\ge Y_{(n)}$, where $Y_{(1)},\cdots,Y_{(n)}$ are the order statistics of $Y_1,\cdots,Y_n$. To ensure that the estimator is finite outside the range of the data with small n, Cheung and Fin (2001) proposed a modified estimator of $h(\cdot)$, say $\widehat{h}(\cdot)$, i.e.,

$$\widehat{h}(y)=\{\dot{h}(Y_{(1)}^{+}),y\le Y_{(1)};\dot{h}(y),Y_{(1)}<y<Y_{(n)};\dot{h}(Y_{(n)}^{-}),y\ge Y_{(n)}\}.\tag{3.4}$$

Since with probability 1 $(0,\tau_0)$ contains $(Y_{(1)},Y_{(n)})$ as $n\to\infty$. Hence Theorems 1-2, but now with $\dot{h}(\cdot)$ replaced by $\widehat{h}(\cdot)$, also hold for $y\in(0,\tau_0)$.

§4. Estimators in Semiparametric Regression

There are many literatures to discuss regression in non-censored case, but for censored data, this techniques are not directly applicable because some of the response and/or covariates may not be observed — they may be censored. A pioneering effort in regression analysis of censored data was made by Miller (1976) and Buckley and James (1979), however these two procedures may fail to converge. Koul, Susarla and Van Ryzin (1981) proposed another approach and gave their consistent and asymptotically normal estimators, they supposed that censoring variables c_i are i.i.d. and suggested replacing T_i by $T_i^* = \delta_i Y_i/(1-G(Y_i))$. Zheng (1987) extend the results of Koul et al. (1981) and proposed a class transformation of data, which depends only on the censoring distribution (assumed to be independent of the covariates). Once such a transformation is carried through, one can apply a variety of statistical techniques to analyze the transformed data as if they were uncensored. This paper, we follow Zheng (1987) and use his transformation of data to investigate our semiparametric transformation model.

Let us recall semiparametric regression model (2.4) in section 2. First we consider the model (2.4) with known h and known censoring distribution G, this is a ordinary regression model with censored data. Note that $h(T_i)$ is censored if and only if T_i is censored, a naive idea is that if $h(T_i)$ is censored we add something to it to make up for the censored part and if $h(T_i)$ is uncensored we also modify it appropriately to ensure unbiasedness in the sense that the modification $h^*(T_i)$ has the same expectation as $h(T_i)$. In view of this consideration, we suggest using $h^*(T_i)$ of the form (for known h and G)

$$h^*(T_i) = \delta_i \varphi_1(Y_i) + (1-\delta_i)\varphi_2(Y_i), \tag{4.1}$$

where φ_1, φ_2 are continuous such that

$$[1-G(t)]\varphi_1(t) + \int_0^t \varphi_2(s)\mathrm{d}G(s) = h(t), \tag{4.2}$$

and φ_1, φ_2 are independent of F_i (but may depend on G). The class of all pairs (φ_1, φ_2) of such functions will be denoted by K (for details, see Zheng (1987)). Various choice $(\varphi_1, \varphi_2) \in K$ can be found in Zheng (1987). In this paper, to circumvent the difficulty in estimating the derivative of $h(\cdot)$, we follow Koul et al. (1981) to obtain the transformed data (this means $\varphi_2(\cdot) = 0$)

$$h^*(T_i) = \frac{\delta_i h(Y_i)}{1-G(Y_i)}. \tag{4.3}$$

In the sequel, for simplicity we consider the case of 1-dimensional regression parameter. Now we return to the linear regression model (2.4). We assume that G and h are known and consider the least squares estimators

$$\widehat{\beta}(G, h^*) = \frac{\sum\limits_{i=1}^{n}(z_i - \overline{z})h^*(T_i)}{\sum\limits_{i=1}^{n}(z_i - \overline{z})^2}, \tag{4.4}$$

based on the transformed data $(z_i, h^*(T_i))$, $(i = 1, 2, \cdots, n)$, where $h^*(T_i)$ is given by (4.3).

If monotone function h is unknown ($G(\cdot)$ is known), it is natural to substitute it by a consistent estimator of h, say $\widehat{h}$ (cf. Remark 2 below), then an estimator of h^* is given by substituting $\widehat{h}$ for h into (4.3), we have

$$\widehat{h}^*(T_i) = \frac{\delta_i \widehat{h}(Y_i)}{1-G(Y_i)}. \tag{4.5}$$

Clearly $\widehat{h}^*(\cdot)$ is a consistent estimator of $h^*(\cdot)$ from Theorems 1-2 in section 2, but now with $\widehat{\overline{G}}$ replaced by $\overline{G}$. Then the least squares estimators based on the consistent estimator $\widehat{h}^*(\cdot)$ and (4.4) is given by

$$\widehat{\beta}(G,\widehat{h}^*)=\frac{\sum\limits_{i=1}^{n}(z_i-\overline{z})\widehat{h}^*(T_i)}{\sum\limits_{i=1}^{n}(z_i-\overline{z})^2}. \tag{4.6}$$

Also when censoring distribution G is unknown ($h(\cdot)$ is known), an estimator of G is used by Koul et al. (1981), Zheng (1987), among others, i.e., an estimator of G, asymptotically like the product-limit estimator, is given in Koul et al. (1981), and a modified Kaplan-Meier estimator is given in Zheng (1987).

Substituting $\widehat{G}$ for G into (4.3), we have

$$\widetilde{h}^*(T_i)=\frac{\delta_i h(Y_i)}{1-\widehat{G}(Y_i)}, \tag{4.7}$$

then the estimator of β is given by

$$\widehat{\beta}(\widehat{G},\widetilde{h}^*)=\frac{\sum\limits_{i=1}^{n}(z_i-\overline{z})\widetilde{h}^*(T_i)}{\sum\limits_{i=1}^{n}(z_i-\overline{z})^2}, \tag{4.8}$$

If censoring distribution G and monotone transformation h are all unknown, the modified Kaplan-Meier estimator of G and the estimator of h defined in Theorems 1-2 are used to replace G and h in (4.3), respectively, then the estimator of h^* and the least squares estimator of β are given by

$$\overline{h}^*(T_i)=\frac{\delta_i \widehat{h}(Y_i)}{1-\widehat{G}(Y_i)}, \tag{4.9}$$

$$\widehat{\beta}(\widehat{G},\overline{h}^*)=\frac{\sum\limits_{i=1}^{n}(z_i-\overline{z})\overline{h}^*(T_i)}{\sum\limits_{i=1}^{n}(z_i-\overline{z})^2}, \tag{4.10}$$

§ 5. Asymptotic Properties of Estimators

Now we return to discuss the asymptotic properties of estimators. We first investigate the case where $h(t)$ and $G(t)$ are all known. For estimator (4.4), note that $\mathsf{E}\{h^*(T_i)\}=\mathsf{E}\{h(T_i)\}$ and $\overline{z}=0$, then

$$\widehat{\beta}(G,h^*)-\beta=\frac{\sum\limits_{i=1}^{n}(z_i-\overline{z})}{\sum\limits_{i=1}^{n}(z_i-\overline{z})^2}[h^*(T_i)-\mathsf{E}h^*(T_i)], \tag{5.1}$$

since $[h^*(T_i)-\mathsf{E}h^*(T_i)]$ ($i=1,2,\cdots,n$) are independent random variables with zero mean, the strong consistent and asymptotic normality of $\widehat{\beta}(G,h^*)$ follow from known results on least squares estimators in regression model with independent errors. The following Theorem 3 is direct derived from Zheng (1987) for known h and known G.

Theorem 3 Suppose that

$$\sup_i \mathsf{Var}\{h^*(T_i)\} < \infty, \qquad S_n^2 = \sum_{i=1}^n (z_i - \overline{z})^2 \to \infty. \tag{A}$$

Then $\widehat{\beta}(G, h^*) - \beta \xrightarrow{\text{a.s.}} 0$. Furthermore if

$$0 < \inf_i \mathsf{Var}\{h^*(T_i)\} \leq \sup_i \mathsf{Var}\{h^*(T_i)\} < \infty, \tag{B}$$

$$\sup_i |z_i| < \infty, \tag{C}$$

and for $\eta > 0$,

$$\sum_{i=1}^n (z_i - \overline{z})^2 \mathsf{E}[h^*(T_i) - \mathsf{E}\{h^*(T_i)\}]^2 I_{(|h^*(T_i) - \mathsf{E}\{h^*(T_i)\}| \cdot |z_i - \overline{z}| \geq \eta V_n)} = o(V_n^2). \tag{D}$$

Then

$$\frac{\sum\limits_{i=1}(z_i - \overline{z})^2}{\left[\sum\limits_{i=1}(z_i - \overline{z})^2 \mathsf{Var}\{h^*(T_i)\}\right]^{1/2}} \{\widehat{\beta}(G, h^*) - \beta\} \xrightarrow{d} N(0, 1),$$

where $V_n = \left[\sum\limits_{i=1}(z_i - \overline{z})^2 \mathsf{Var}\{h^*(T_i)\}\right]^{1/2}$.

If one or all of h and G are unknown, then the least squares estimators of β are given by (4.6), (4.8), (4.10), respectively. To establish the asymptotic properties of $\widehat{\beta}(\cdot, \cdot)$, we need the following lemmas.

Lemma 1 Under the conditions listed in Theorem 1 and (3.2), then $\widehat{h}(Y) - h(Y) \xrightarrow{p} 0$, where $\widehat{h}(t)$ is defined in (3.4) through Theorem 1 and Y is a radom variable of $Y_1, \cdots, Y_n$.

Proof Recall an estimator $\dot{h}(t)$ defined in Theorem 1, i.e., $\dot{h}(t) = \sum\limits_{j=1}^q \lambda_{jn} W^{-1}(\widehat{F}_j(t))$, where $\widehat{F}_j(t)$ is the Kaplan-Meier estimator of $F_j(t)$ from a i.i.d. sample $\{Y_{j1}, \cdots, Y_{jn_j}\}$ in cell j (see Theorem 1). Note that under the i.i.d. censoring model, (5.2) is due to Peterson (1977) and Wang (1987)

$$\sup_{t \in [0, \tau_0]} |\widehat{F}_j(t) - F_j(t)| \xrightarrow{\text{a.s.}} 0, \qquad \text{as } n \to \infty. \tag{5.2}$$

Thus

$$|\widehat{F}_j(Y) - F_j(Y)| \xrightarrow{p} 0, \qquad \text{as } n \to \infty. \tag{5.3}$$

Note that

$$\left(\left|\sum_{i=1}^q \widehat{F}_j(Y) - \sum_{i=1}^q F_j(Y)\right| > \varepsilon\right) \subseteq \bigcup_{i=1}^q (|\widehat{F}_j(Y) - F_j(Y)| > \varepsilon / q)$$

and $W^{-1}(\cdot)$ is continuous, thus $\widehat{h}(Y) - h(Y) \xrightarrow{p} 0$. This follows from Theorem 1, the assumption (3.2), the modified estimator $\widehat{h}(\cdot)$ in (3.4) and (5.3). #

Lemma 2 Under the conditions listed in Theorem 2 and (3.2), then $\widehat{h}(Y)-h(Y)\xrightarrow{p}0$, where $\widehat{h}(t)$ is defined in (3.4) through Theorem 2 and Y is a random variable of $Y_1,\cdots,Y_n$.

Proof Under the assumptions (3.2) and the modified estimator $\widehat{h}(y)$ in (3.4), Following the notation from Theorem 2, we need to show $v_n(Y)-v(Y)\xrightarrow{p}0$. Note from Theorem 2 that

$$\begin{aligned} v_n(Y)-v(Y) &= \frac{1}{n}\sum_{i=1}^{n}\Big\{\frac{I[\varepsilon_i>(h(Y)-\mu_i)]}{\widehat{\overline{G}}(Y)}-\overline{W}(h(Y))\Big\} \\ &= \frac{1}{n}\sum_{i=1}^{n}\Big\{\Big[\frac{I[\varepsilon_i>(h(Y)-\mu_i)]}{\overline{G}(Y)}-\overline{W}(h(Y)-\mu_i)\Big]\Big\} \\ &\quad+\frac{1}{n}\sum_{i=1}^{n}\Big\{\frac{I[\varepsilon_i>(h(Y)-\mu_i)]}{\overline{G}(Y)}\Big[\frac{\overline{G}(Y)-\widehat{\overline{G}}(Y)}{\widehat{\overline{G}}(Y)}\Big]\Big\} \\ &\quad+\frac{1}{n}\sum_{i=1}^{n}\{\overline{W}(h(Y)-\mu_i)-\overline{W}(h(Y))\} \\ &= I_1+I_2+I_3. \end{aligned} \tag{5.4}$$

Mimicking the proof of Proposition 6.2 in Doksum (1987, p. 341), we can show $I_3\xrightarrow{p}0$, then we need only to show $I_1\xrightarrow{p}0$ and $I_2\xrightarrow{p}0$.

For I_1, let $\xi_i=I[\varepsilon_i>(h(Y)-\mu_i)]/\overline{G}(Y)-\overline{W}(h(Y)-\mu_i)$, $i=1,\cdots,n$, we should mention that $\xi_1,\cdots,\xi_n$ are dependent, but conditionally independent with $\mathsf{E}\{\xi_i\}=0$, thus

$$\mathsf{Var}\{I_1\}=n^{-2}\mathsf{E}\Big\{\sum_{i=1}^{n}\mathsf{Var}\{\xi_i|Y\}\Big\}=n^{-2}\sum_{i=1}^{n}\mathsf{E}\{\mathsf{Var}\{\xi_i|Y\}\}. \tag{5.5}$$

On the other hand, since random variables $\varepsilon_1,\cdots,\varepsilon_n$ are mutually conditional independence, the conditionally standard deviation of ξ_i can be described by

$$\begin{aligned} \mathsf{Var}\{\xi_i|Y\} &= \mathsf{Var}\Big\{\frac{I[\varepsilon_i>(h(Y)-\mu_i)]}{\overline{G}(Y)}\Big|Y\Big\} \\ &= \frac{1}{\overline{G}^2(Y)}\mathsf{Var}\{I[\varepsilon_i>(h(Y)-\mu_i)]|Y\} \\ &= \frac{1}{\overline{G}^2(Y)}W(h(Y)-\mu_i)[1-W(h(Y)-\mu_i)] \\ &\le \frac{1}{2\overline{G}^2(Y)}<\frac{1}{2\overline{G}^2(\tau_0)}, \end{aligned} \tag{5.6}$$

from (5.5)-(5.6), we have

$$\mathsf{Var}\{I_1\}\le\frac{1}{2n\overline{G}^2(\tau_0)}. \tag{5.7}$$

Thus $I_1\xrightarrow{p}0$ holds from (5.7).

For I_2, since $\widetilde{F}(t)=n^{-1}\sum_{i=1}^{n}F_i(t)$ exists for $t>0$, $1-H_i(t)=[1-F_i(t)][1-G(t)]$. It is well known that (see also Zheng (1988), p. 311)

$$\sup_{t\in[0,t_0]}|\widehat{\overline{G}}(t)-\overline{G}(t)|\xrightarrow{\text{a.s.}}0,\qquad t_0<\sup_i\tau_{H_i}, \tag{5.8}$$

then by the assumptions (3.2)

$$|\overline{G}(Y) - \widehat{\overline{G}}(Y)| \xrightarrow{p} 0, \qquad \text{as } n \to \infty. \tag{5.9}$$

Next by the assumptions (3.2), (5.8) and $\overline{G}(Y) \geq \overline{G}(\tau_0)$, then we have

$$n^{-1} \sum_{i=1}^{n} \left\{ \frac{I[\varepsilon_i > (h(Y) - \mu_i)]}{\overline{G}(Y)\widehat{\overline{G}}(Y)} \right\} \leq \frac{1}{\overline{G}^2(\tau_0)}. \tag{5.10}$$

From (5.9)-(5.10), we obtain that $I_2 \xrightarrow{p} 0$ as $n \to \infty$. This implies that $v_n(Y) - v(Y) \xrightarrow{p} 0$.

Note that $\overline{W}^{-1}(\cdot)$ is continuous function, Then we obtain that $\widehat{h}(Y) - h(Y)$ converges to zero in probability. #

Remark 2 If G is known and h is unknown, an estimator of h can be defined $\dot{h}(y) = \overline{W}^{-1}\{n^{-1}\#[Y_i > y]\}/\overline{G}(y)$, it is similar to Theorem 2 together with lemma 2 we can find the asymptotic properties of $\dot{h}(y)$, hence $\widehat{h}(Y)$.

Now we turn to show Theorem 4.

Theorem 4 Under conditions (A) and (C) listed in Theorem 3, and

$$\liminf_{n\to\infty} n^{-1} \sum_{i=1}^{n} (z_i - \overline{z})^2 > 0. \tag{E}$$

(a) If h is unknown but G is known, we estimate h by $\widehat{h}$ defined in Remark 2 above. Then $\widehat{\beta}(G, \widehat{h}^*) \xrightarrow{p} \beta$.

(b) If G is unknown but h is known, we replace G with its Kaplan-Meier estimator. Then $\widehat{\beta}(\widehat{G}, \widetilde{h}^*) \xrightarrow{p} \beta$.

(c) If G and h are all unknown, we replace G with its Kaplan-Meier estimator and replace h with its estimator defined in Theorems 1-2 and (3.4). Then $\widehat{\beta}(\widehat{G}, \overline{h}^*) \xrightarrow{p} \beta$.

proof Note that the least squares estimator, such as, $\widehat{\beta}(\widehat{G}, \overline{h}^*)$ can be decomposed into

$$\widehat{\beta}(\widehat{G}, \overline{h}^*) - \beta = [\widehat{\beta}(G, h^*) - \beta] + [\widehat{\beta}(\widehat{G}, \overline{h}^*) - \widehat{\beta}(G, h^*)]. \tag{5.11}$$

Since the convergence properties of $(\widehat{\beta}(G, h^*) - \beta)$ have been established (cf. Theorem 3), it remains to show the difference $(\widehat{\beta}(\widehat{G}, \overline{h}^*) - \widehat{\beta}(G, h^*))$. Letting $a_n = \sum_{i=1}^{n} (z_i - \overline{z})^2$, we write

$$\widehat{\beta}(\widehat{G}, \overline{h}^*) - \widehat{\beta}(G, h^*) = a_n^{-1} \Big\{ \sum_{i=1}^{n} (z_i - \overline{z}) \Big[\frac{\delta_i \widehat{h}(Y_i)}{1 - \widehat{G}(Y_i)} - \frac{\delta_i h(Y_i)}{1 - G(Y_i)} \Big] \Big\}.$$

Note that

$$\begin{aligned} & \Big| \frac{\widehat{h}(Y_i)}{1 - \widehat{G}(Y_i)} - \frac{h(Y_i)}{1 - G(Y_i)} \Big| \\ = & \Big| \frac{\widehat{h}(Y_i)}{1 - \widehat{G}(Y_i)} - \frac{\widehat{h}(Y_i)}{1 - G(Y_i)} \Big| + \Big| \frac{\widehat{h}(Y_i)}{1 - G(Y_i)} - \frac{h(Y_i)}{1 - G(Y_i)} \Big| \\ = & I_4 + I_5. \end{aligned} \tag{5.12}$$

For I_4, note that $1-G(Y_i)\geq 1-G(\tau_0)$ and $\widehat{h}(Y_i)\xrightarrow{p} h(Y_i)$ from Lemmas 1-2, thus together with (5.10), we have

$$I_4 \leq \frac{|\widehat{h}(Y_i)|}{(1-G(\tau_0))^2}|\widehat{G}(Y_i)-G(Y_i)| \xrightarrow{p} 0. \tag{5.13}$$

For I_5, from lemmas 1-2,

$$I_5 \leq \frac{1}{1-G(\tau_0)}|\widehat{h}(Y_i)-h(Y_i)| \xrightarrow{p} 0. \tag{5.14}$$

Also note that $(|\delta_i\widehat{h}(Y_i)-\delta_i h(Y_i)|>\varepsilon)\subseteq(|\widehat{h}(Y_i)-h(Y_i)|>\varepsilon)$ for any $\varepsilon>0$. This completes the proof of (5.11) from (5.12)-(5.14).

It's similar to show (a) and (b). This completes the proof. #

Remark 3 Based on the assumptions in (3.2) and the modified estimator $\widehat{h}(\cdot)$ in (3.4), the tail of $\widehat{\widehat{G}}$ can be deal with easily. Then condition $\tau_{F_0}<\tau_G$ in Zheng (1987) is redundant.

The limiting normal distribution of $\sqrt{n}(\widehat{\beta}(G,\widehat{h}_G^*)-\beta)$ is similar to Theorem 4 in Zheng (1987).

Theorem 5 Under conditions (B)-(D) in Theorem 3 and condition (E) in Theorem 4, and conditions of Theorems 1-2 hold. Moreover assume

$$\begin{aligned} R(t_1,t_2) &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n F_i(t_1)F_i(t_2) \qquad \text{exists for all } t_1,t_2\geq 0,\\ R_1(t) &= \lim_{n\to\infty}\frac{1}{\sum\limits_{i=1}^n(z_i-\overline{z})^2}\sum_{i=1}^n(z_i-\overline{z})\mathrm{P}(\delta_i=1,Y_i\leq t),\\ R_2(t) &= \lim_{n\to\infty}\frac{1}{\sum\limits_{i=1}^n(z_i-\overline{z})^2}\sum_{i=1}^n(z_i-\overline{z})\mathrm{P}(\delta_i=0,Y_i\leq t), \end{aligned}$$

exist for all $t\geq 0$ and $\widetilde{F}_n(\tau_{F_0})=1-0(n^{-1/2})$,

$$\int_0^{\tau_{F_0}}\frac{\mathrm{d}G(s)}{1-\overline{F}(s-)}<\infty.$$

Then $\sqrt{n}(\widehat{\beta}(G,\widehat{h}^*)-\beta)$ in case (a) of Theorem 4, $\sqrt{n}(\widehat{\beta}(\widehat{G},\widetilde{h}^*)-\beta)$ in case (b) of Theorem 4, or $\sqrt{n}(\widehat{\beta}(\widehat{G},\overline{h}^*)-\beta)$ in case (c) of Theorem 4, has a limiting normal distribution with zero mean and a finite variance matrix.

The proof of the limiting normal distribution in Theorem 5 are tedious. We omit the details here, because the method is standard, such as: develop the convergence of $\sqrt{n}(\widehat{h}-h)$ to a Gaussian process $\widetilde{W}$ (cf. Theorems 1-2), express, such as, $\sqrt{n}(\widehat{\beta}(G,\widehat{h}^*)-\widehat{\beta}(G,h^*))$ as a function of $\widetilde{W}$ and a remainder, and show the remainder tends to zero, thus completing the proof.

§6. Simulation Results

In the small simulation study, we compare the performance of the estimators of the parameters with the true values. The general calculations can be seen more clearly in special cases. We consider the two sample problem with exponential distribution lifetimes. The two samples are of sizes, say, n_1 and n_2, respectively, $n_1 + n_2 = n$, with sample membership being indicated by the dummy variable

$$z_i = \begin{cases} -1, & \text{if individual } i \text{ is in sample 1} \\ 1, & \text{if individual } i \text{ is in sample 2,} \end{cases} \qquad i = 1, 2, \cdots, n.$$

Data are generated from the survival functions $S_1(t) = (1 - F_0(t))^{\exp(-\beta)}$ (with respect to sample 1) and $S_2(t) = (1 - F_0(t))^{\exp(\beta)}$ (with respect to sample 2). Let $F_0(t)$ be an exponential distribution with parameter $\psi = 0.058$. The coefficient of covariates is $\beta = 0.3581$. Censoring times c are generated from a uniform distribution between 0 to 100. For this simulation, $\dot{h}$ is given by Theorem 2 and a sample of size $n_1 = n_2 = 100$ and $n_1 = n_2 = 400$ were replicated 10000 times.

Let $\beta_1^{(1)}, \beta_2^{(1)}, \beta_3^{(1)}, \beta_4^{(1)}$ denote the linear least squares estimators based on (4.4), (4.6), (4.8), (4.10) with the sample size $n_1 = n_2 = 100$, respectively. And $\beta_1^{(2)}, \beta_2^{(2)}, \beta_3^{(2)}, \beta_4^{(2)}$ denote the linear least squares estimators based on (4.4), (4.6), (4.8), (4.10) with the sample size $n_1 = n_2 = 400$, respectively. The means and the standard deviations of the simulated estimates with the sample size $n = 100$ and $n = 400$ are displayed in Tables 1 below.

From Table 1, we see that the estimates are reasonably close to the true values of the parameters and the accuracy improves as the sample sizes increase from $n_1 = n_2 = 100$ to $n_1 = n_2 = 400$.

Table 1 Summary of the simulation studies on the estimators of β

	$\beta_1^{(1)}$	$\beta_2^{(1)}$	$\beta_3^{(1)}$	$\beta_4^{(1)}$	$\beta_1^{(2)}$	$\beta_2^{(2)}$	$\beta_3^{(2)}$	$\beta_4^{(2)}$
Mean	0.3265	0.3106	0.3238	0.3085	0.3553	0.3570	0.3527	0.3547
STD	0.0845	0.0759	0.0843	0.0759	0.0468	0.0443	0.0457	0.0452

References

[1] Box, G.E.P. and Cox, D.R., An analysis of transformations (with discussion), *J. Roy. Stat. Soc. B*, **26**(1964), 211–252.

[2] Buckley, J. and James, I., Linear regression with right-censored data, *Biometrika*, **66(3)**(1979), 429–436.

[3] Cheng, S.C., Wei, L.J. and Ying, Z.L., Analysis transformation models with censored data, *Biometrika*, **82(4)**1995, 835–845.

[4] Cheng, S.C., Wei, L.J. and Ying, Z.L., Predictive survival probabilities with semiparametric transformation model, *J. Am. Stat. Assoc.*, **92(1)**1997, 227–235.

[5] Cheung, Y.K. and Fine, J.P., Likelihood estimation after nonparametric transformation, *Stat. Probabil. Lett.*, **55(1)**(2001), 1–7.

[6] Cox, D.R., Regression models and life-tables (with discussion), *J. Roy. Stat. Soc. B*, **34**(1972), 187–220.

[7] Cuzick, J., Rank regression, *Ann. Stat.*, **16(4)**(1988), 1369–1389.

[8] Doksum, K.A., An extension of partial likelihood methods for proportional hazards models to general transformation models, *Ann. Stat.*, **15(1)**(1987), 325–345.

[9] Dabrowska, D.M. and Doksum, K.A., Partial likelihood in transformation models with censored data, *Scand. J. Stat.*, **15(1)**(1988), 1–23.

[10] Fine, J.P. and Bosch, R.J., Risk assessment via a robust probit model with application to toxicology, *J. Am. Stat. Assoc.*, **95(540)**(2000), 375–382.

[11] Gill, R., Large sample behavior of the product-limit estimator on the whole line, *Ann. Stat.*, **11(1)**(1983), 49–83.

[12] Koul, H., Susarla, V. and Van Ryzin, J., Regression analysis with randomly right-censored data, *Ann. Stat.*, **9(6)**(1981), 1276–1288.

[13] Mill, R.G., Least squares regression with censored data, *Biometrika*, **63(3)**(1976), 449–464.

[14] Murphy, S.A., Rossini, A.J. and Van der Vaart, A.W., Maximum likelihood estimation in the proportional odds model, *J. Am. Stat. Assoc.*, **92(439)**(1997), 968–976.

[15] Peterson, Arthur V., Jr., Expressing the Kaplan-Meier estimator as a function of empirical subsurvival functions, *J. Am. Stat. Assoc.*, **72(360)**(1977), 854–858.

[16] Pettitt, A.N., Proportional odds models for survival data and estimates using rank, *J. Appl. Stat.*, **33(2)**(1984), 169–175.

[17] Shen, X.T., Proportional odds regression and sieve maximum likelihood estimation, *Biometrika*, **85(1)**(1998), 165–177.

[18] Shorack, G.R. and Wellner, J.A., *Empirical Processes with Applications to Statistics*, John Wiley and Sons, New York, 1986.

[19] Wang, J.G., A note on the uniform consistency of the Kaplan-Meier estimator, *Ann. Stat.*, **15(3)**(1987), 1313–1316.

[20] Zheng, Z.K., Strong consistency of nonparametric regression estimates with censored data, *J. Math. Res. Exposition*, **8(2)**(1988), 307–313.

[21] Zheng, Z.K., A class of estimators of the parameters in linear regression with censored data, *Acta Math. Sinica*, **3(3)**(1987), 231–241.

基于生存数据的半参数变换的删失回归估计

赵 晓 兵

(江南大学理学院，无锡， 214063)

生存数据经过未知的单调变换后等于协变量的线性函数加上随机误差，随机误差的分布函数已知或是带未知参数的已知函数．本文先给出未知单调变换的一个相合估计，再对删失数据做变换，在此基础上给出了协变量系数的最小二乘估计，并讨论它的大样本性质．

关 键 词: 比例危险率模型，半参数变换，删失回归，最小二乘， Class-K 方法.

学 科 分 类 号: O212.