A Hybrid Newton-Type Method for the Linear Regression in Case-cohort Studies

Menggang Yu* Bin Nan†

*University of Indiana, meyu@iupui.edu

†University of Michigan, bnan@umich.edu

A Hybrid Newton-Type Method for the Linear Regression in Case-cohort Studies

Menggang Yu and Bin Nan

Abstract

Case-cohort designs are increasingly commonly used in large epidemiological cohort studies. Nan, Yu, and Kalbeisch (2004) provided the asymptotic results for censored linear regression models in case-cohort studies. In this article, we consider computational aspects of their proposed rank based estimating methods. We show that the rank based discontinuous estimating functions for case-cohort studies are monotone, a property established for cohort data in the literature, when generalized Gehan type of weights are used. Though the estimating problem can be formulated to a linear programming problem as that for cohort data, due to its easily uncontrollable large scale even for a moderate sample size, we instead propose a Newton-type iterated method to search for an approximate root for the discontinuous monotone estimating function. Simulation results provide a good demonstration of the proposed method.

A Hybrid Newton-Type Method for the Linear Regression in Case-cohort Studies

Menggang Yu and Bin Nan

Department of Medicine/Biostatistics, Indiana University
1050 Wishard Boulevard, RG4101
Indianapolis, IN 46202.
(E-mail: meyu@iupui.edu)
and
Department of Biostatistics,
University of Michigan, Ann Arbor, MI 48109
(E-mail: bnan@umich.edu)

December 16, 2004

Abstract

Case-cohort designs are increasingly commonly used in large epidemiological cohort studies. Nan, Yu, and Kalbfleisch (2004) provided the asymptotic results for censored linear regression models in case-cohort studies. In this article, we consider computational aspects of their proposed rank based estimating methods. We show that the rank based discontinuous estimating functions for case-cohort studies are monotone, a property established for cohort data in the literature, when generalized Gehan type of weights are used. Though the estimating problem can be formulated to a linear programming problem as that for cohort data, due to its easily uncontrollable large scale even for a moderate sample size, we instead propose a Newton-type iterated method to search for an approximate root for the discontinuous monotone estimating function. Simulation results provide a good demonstration of the proposed method.

Key Words: Case-cohort design; Censored linear regression; Gehan-type weights; Linear Programming; Monotone estimating function; Newton-type method.

1 Introduction

Case-cohort designs have attracted much attention in statistics since they were proposed by Prentice (1986) and their applications are increasingly common in the medical literature. This type of designs are especially desirable for large epidemiological cohort studies with few observed failures in which enormous resources may be required to ascertain some important covariates. In case-cohort studies, data collection for censored individuals is reduced dramatically by only measuring expensive covariates for a small sub-sample of the censored individuals in addition to failures.

Methods of analyzing case-cohort data mainly are modifications of the standard Cox regression approach (see e.g. Prentice, 1986; Self and Prentice, 1988; Barlow, et al., 1999). Recently, Nan et al. (2004) proposed a semiparametric censored linear regression approach to analyze case-cohort data and provided asymptotic results. The estimating procedure is an extension of the rank-based analysis for full cohort data (see e.g. Tsiatis, 1990). The linear model is desirable because its interpretation is much more straightforward than hazard regression models. Due to the non-smoothness of the estimating equations, however, solving those estimating equations tends to be much more challenging than searching for a root of a smooth function (or a system of smooth functions). For full cohort data, Lin and Geyer (1992) proposed a simulated annealing method to estimate the regression parameter. Fygenson and Ritov (1994) established the monotonicity of the weighted estimating function when Gehan weights are used. Lin et al. (1998) then noticed that solving the Gehan weighted estimating equation is equivalent to an optimization problem formulated by linear programming, and thus solvable by the simplex algorithm. Jin et al. (2003) considered using linear programming in solving estimating functions with Gehan and other type of weights.

In this article, we investigate the global behavior of the estimating function proposed by Nan et al. (2004) for case-cohort studies. In particular, we prove componentwise monotonicity of the weighted estimating function with Gehan-type weights. As a result, we are able to provide a linear programming formulation for the estimation of regression parameters. We find, however, that the solvability of the linear programming problem for the estimation of regression parameters depends largely on the capacity of a computer since the numbers of unknown variables and of linear constraints are in the order of square of the sample size. Usually the regression parameter has much lower dimension comparing to the sample size.

We thus propose a Newton-type iterated method by approximating the slope in each iteration and adjusting iteration rules based on the knowledge of monotonicity and discontinuity. We call it a hybrid Newton-type method. The rest of the paper is organized as follows. In Section 2, we introduce the rank based estimating function for the linear regression parameter in case-cohort studies. In Section 3, we introduce generalized Gehan-type weights and show monotonicity of the corresponding weighted estimating function. We discuss the linear programming formulation in Section 4, and propose the hybrid Newton-type method in Section 5. We illustrate numerical implementations in Section 6 followed by a brief discussion section where we discuss similar extensions to arbitrary weights as in Jin et al. (2003).

2 Estimating Functions for Linear Models

Let T be the monotonically transformed failure time with a known transformation, C be the transformed censoring time with the same transformation. The log transformation is often used in practice and the corresponding model is called the accelerated failure time model, see e.g. Kalbfleisch and Prentice (2002). For subject i in the cohort, we only observe the minimum of T_i and C_i, denoted as $X_i \equiv T_i \wedge C_i$, and the failure indicator $\Delta_i \equiv I\{T_i \leq C_i\}$. Let $\mathbf{Z}_i$ be the d-dimensional covariate. The model of interest is:

$$T_i = \boldsymbol{\beta}_0' \mathbf{Z}_i + e_i, \quad i = 1, \dots, n, \tag{2.1}$$

where n is the total number of individuals in the cohort. We assume that e_i's are independent and identically distributed with an unknown distribution, and e_i is independent of $(\mathbf{Z}_i, C_i)$ for all i.

When $(\mathbf{Z}_i, X_i, \Delta_i)$ are observable for the entire cohort, Tsiatis (1990) introduced the following estimating function for $\boldsymbol{\beta}_0$:

$$S_n(\boldsymbol{\beta}, W_n) = \sum_{i=1}^{n} \int W_n(u, \boldsymbol{\beta}) \{\mathbf{Z}_i - \bar{\mathbf{Z}}(u, \boldsymbol{\beta})\} \, dN_i(u + \boldsymbol{\beta}' \mathbf{Z}_i) \tag{2.2}$$

where $N_i(u + \boldsymbol{\beta}'\mathbf{Z}_i) = I(X_i - \boldsymbol{\beta}'\mathbf{Z}_i \leq u, \Delta_i = 1)$ is the failure counting process for subject i and

$$\bar{\mathbf{Z}}(u, \boldsymbol{\beta}) \equiv \frac{\sum_{j=1}^{n} \mathbf{Z}_j Y_j(u + \boldsymbol{\beta}'\mathbf{Z}_j)}{\sum_{j=1}^{n} Y_j(u + \boldsymbol{\beta}'\mathbf{Z}_j)}$$

with $Y_j(u + \boldsymbol{\beta}'\mathbf{Z}_j) = I(X_j - \boldsymbol{\beta}'\mathbf{Z}_j \geq u)$. The stochastic function $W_n(u, \boldsymbol{\beta})$ is a weight process. Apparently estimating function (2.2) is based on ranks of residuals $X_i - \boldsymbol{\beta}'\mathbf{Z}_i$, $i = 1, \cdots, n$, it is thus a step function of $\boldsymbol{\beta}$ in contrast to a nicely continuous estimating function in the Cox model. When $d = 1$, the solution $\hat{\boldsymbol{\beta}}$ to equation $S_n(\boldsymbol{\beta}, W_n) = 0$ is usually defined as a zero crossing where $S_n(\boldsymbol{\beta}, W_n)$ changes sign.

Tsiatis (1990) showed that $S_n(\boldsymbol{\beta}, W_n)$ is asymptotically linear in a $n^{-1/2}$-neighborhood of the true value $\boldsymbol{\beta}_0$. Then the proof of asymptotic normality of $\hat{\boldsymbol{\beta}}$, the (approximate) root of $S_n(\boldsymbol{\beta}, W_n)$, became straightforward. Ritov (1990) and Ying (1993) also studied the theory for the same problem from different angles. The fact that $S_n(\boldsymbol{\beta}, W_n)$ is neither continuous nor componentwise monotone in $\boldsymbol{\beta}$ in general, however, makes it difficult to solve $S_n(\boldsymbol{\beta}, W_n) = 0$ numerically, especially when $d > 1$. The variance estimation is also difficult as its asymptotic form involves the derivative of the hazard function of the error e in (2.1).

Fygenson and Ritov (1994) proved that if $W_n(u, \boldsymbol{\beta})$ is taken to be the Gehan weight, i.e., $W_n(u, \boldsymbol{\beta}) = \sum_{i=1}^n Y_i(u + \boldsymbol{\beta}'\mathbf{Z}_i)/n$, then

$$S_n(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \Delta_i (\mathbf{Z}_i - \mathbf{Z}_j) I(X_i - \boldsymbol{\beta}'\mathbf{Z}_i \leq X_j - \boldsymbol{\beta}'\mathbf{Z}_j)$$

is monotone in each component of $\boldsymbol{\beta}$. Notice that we dropped the argument W_n of S_n in the above equation for the Gehan weight. It turns out that the above $S_n(\boldsymbol{\beta})$ enables a linear programming formulation (See e.g. Lin et al., 1998; Jin et al., 2003), thus makes the root searching problem a standard optimization problem.

In a case-cohort study, we only observe complete data $(\mathbf{Z}_i, X_i, \Delta_i)$ when subject i is either a failure or a member of the subcohort that is a sub-sample of the entire cohort. Let $\mathcal{D}$ denote the set of failures observed during the study period, and $\mathcal{C}$ denote the subcohort. Note that the intersection of these two sets may not be empty. The subcohort $\mathcal{C}$ can be either a simple random sub-sample or a stratified sub-sample of the entire cohort, corresponding to a classical case-cohort design or a stratified case-cohort design discussed in Nan et al. (2004). For either case, the proposed estimating function for $\boldsymbol{\beta}_0$ has the following form that has a similar structure as the estimating functions proposed by Self and Prentice (1988) and by Borgan et al. (2000) for hazard regression models:

$$\widetilde{S}_n(\boldsymbol{\beta}, W_n) = \sum_{i=1}^n \int W_n(u, \boldsymbol{\beta}) \{\mathbf{Z}_i - \widetilde{\mathbf{Z}}(u, \boldsymbol{\beta})\} \mathrm{d}N_i(u + \boldsymbol{\beta}'\mathbf{Z}_i), \tag{2.3}$$

where

$$\widetilde{\mathbf{Z}}(u, \boldsymbol{\beta}) = \frac{\sum_{j=1}^{n} \rho_j \mathbf{Z}_j Y_j(u + \boldsymbol{\beta}' \mathbf{Z}_j)}{\sum_{j=1}^{n} \rho_j Y_j(u + \boldsymbol{\beta}' \mathbf{Z}_j)}$$

and ρ_j is a weight function for subject j. Let η_j be the subcohort membership indicator variable for subject j, i.e. $\eta_j \equiv I(j \in \mathcal{C})$. In the case that $\mathcal{C}$ is a simple random sub-sample, $\rho_j = n/n_1$ where n_1 is the subcohort size. In the case that $\mathcal{C}$ is a stratified sub-sample $\rho_j = \eta_j/\pi(\mathbf{Z}_j^*)$, where $P(\eta_j = 1) = \pi(\mathbf{Z}_j^*)$ is the probability of being selected into the subcohort for subject j and $\mathbf{Z}^*$ is a set of auxiliary variables available for all subjects in the cohort. We assume that $\pi(\mathbf{Z}_j^*) \geq \alpha > 0$ for some small quantity α. Other types of weights may also be considered, see e.g. Borgan et al. (2000). But we will focus on the above two types of weights in this article. Apparently $\rho_j \equiv 1$ for all j yields an identical estimating function in both (2.2) and (2.3).

Nan et al. (2004) proved that under certain regularity conditions, $\widetilde{S}_n(\boldsymbol{\beta}, W_n)$ is asymptotically linear in a $n^{-1/2}$-neighborhood of the true value $\boldsymbol{\beta}_0$, and the estimator $\hat{\boldsymbol{\beta}}$ is asymptotically normal. The variance of $\hat{\boldsymbol{\beta}}$ can be estimated using the idea of Huang (2002). Specifically, decompose the variance matrix of $n^{-1/2}\widetilde{S}_n(\boldsymbol{\beta}, W_n)$ at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ (denoted by $\Sigma(\hat{\boldsymbol{\beta}})$) as $\Sigma(\hat{\boldsymbol{\beta}}) = \mathbf{C}\mathbf{C}^T$, where $\mathbf{C} = (\mathbf{c}_1, \ldots, \mathbf{c}_d)$; Then solve equations $n^{-1/2}\widetilde{S}_n(\widetilde{\boldsymbol{\beta}}_j, W_n) = \mathbf{c}_j$ for $\widetilde{\boldsymbol{\beta}}_j$, $j = 1, \ldots, d$. Let $\mathbf{D} = (\widetilde{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}, \ldots, \widetilde{\boldsymbol{\beta}}_d - \hat{\boldsymbol{\beta}})$. Then $n\mathbf{D}\mathbf{D}^T$ is a consistent variance estimator of $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$. Hence to obtain both $\hat{\boldsymbol{\beta}}$ and its variance estimator, we need to solve the equation $\widetilde{S}_n(\boldsymbol{\beta}, W_n) = \mathbf{b}$ many times for different constant vector $\mathbf{b}$. Developing an efficient and reliable numerical algorithm becomes crucial. To achieve the goal, we first show in the following section that $\widetilde{S}_n(\boldsymbol{\beta}, W_n)$ is componentwise monotone in $\boldsymbol{\beta}$ if generalized Gehan-type weights are used.

3 Generalized Gehan-type Weights and Monotonicity

Without loss of generality, we work on the case of one-dimensional covariate in this section. Considering the two types of weights ρ_j introduced in the previous section, we can rewrite $\widetilde{S}_n(\beta, W_n)$ as

$$\sum_{i=1}^{n} \Delta_i W_{n,i}^{\beta} \left\{ Z_i - \frac{\sum_{j \in \mathcal{C}} \rho_j Z_j I(X_j - \beta Z_j \geq X_i - \beta Z_i)}{\sum_{j \in \mathcal{C}} \rho_j I(X_j - \beta Z_j \geq X_i - \beta Z_i)} \right\},$$

where we explicitly write $W_{n,i}^{\beta}$ to emphasize that $W_{n,i}$ for subject i depends on β. Note that the above statistic as a function of β depends on ranks of residuals $X_i - \beta Z_i$ only for $i \in \mathcal{C} \cup \mathcal{D}$, the set of observations with complete data. Hence when β varies to β^+, the statistic changes whenever there is a change of ranks of residuals $X_i - \beta^+ Z_i$ comparing to that of $X_i - \beta Z_i$, $i \in \mathcal{C} \cup \mathcal{D}$.

Let us denote the id index for the ith order statistic of residuals $X_i - \beta Z_i$, $i \in \mathcal{C} \cup \mathcal{D}$, as $(i)^{\beta}$, and the corresponding observed time, failure indicator, covariate, and subcohort membership indicator as $X_{(i)}^{\beta}$, $\Delta_{(i)}^{\beta}$, $Z_{(i)}^{\beta}$, and $\eta_{(i)}^{\beta}$. Let $\widetilde{\mathcal{R}}_{(i)}^{\beta}$ be the risk set corresponding to $(i)^{\beta}$ in the subcohort $\mathcal{C}$. Then $\widetilde{S}_n(\beta, W_n)$ can be written as

$$\widetilde{S}_n(\beta, W_n) = \sum_{i \in \mathcal{C} \cup \mathcal{D}} \Delta_{(i)}^{\beta} W_{n,(i)}^{\beta} \left\{ Z_{(i)}^{\beta} - \widetilde{Z}_{(i)}(\beta) \right\}, \tag{3.1}$$

where $W_{n,(i)}^{\beta} = W_n(X_{(i)}^{\beta} - \beta Z_{(i)}^{\beta}, \beta)$, and

$$\widetilde{Z}_{(i)}(\beta) = \frac{\sum_{j \in \mathcal{C}} \rho_j Z_j I(X_j - \beta Z_j \geq X_{(i)}^{\beta} - \beta Z_{(i)}^{\beta})}{\sum_{j \in \mathcal{C}} \rho_j I(X_j - \beta Z_j \geq X_{(i)}^{\beta} - \beta Z_{(i)}^{\beta})} = \frac{\sum_{j \in \widetilde{\mathcal{R}}_{(i)}^{\beta}} \rho_j Z_j}{\sum_{j \in \widetilde{\mathcal{R}}_{(i)}^{\beta}} \rho_j} .$$

Note that under β, the ordered ids are $\{(1)^{\beta}, \cdots, (n_{\mathcal{C}\cup\mathcal{D}})^{\beta}\}$; and under β^+, the ordered ids are $\{(1)^{\beta^+}, \cdots, (n_{\mathcal{C}\cup\mathcal{D}})^{\beta^+}\}$, where $n_{\mathcal{C}\cup\mathcal{D}}$ is the size of $\mathcal{C} \cup \mathcal{D}$. They may change as β moves to β^+. Suppose that for a small change in β, an interchange in ranks occurs only between two neighboring order statistics of the residuals $X_i - \beta Z_i$, $i \in \mathcal{C} \cup \mathcal{D}$. Specifically, if $(k)^{\beta}$ and $(k+1)^{\beta}$ are interchanged, then we have

$$(j)^{\beta^+} = (j)^{\beta}, \text{ for } j \notin \{k, k+1\}, \tag{3.2}$$

and

$$(k+1)^{\beta^+} = (k)^{\beta}, \qquad (k)^{\beta^+} = (k+1)^{\beta}. \tag{3.3}$$

Since $X_{(k+1)}^{\beta} - \beta Z_{(k+1)}^{\beta} \geq X_{(k)}^{\beta} - \beta Z_{(k)}^{\beta}$, by (3.2) and (3.3) we have

$$X_{(k)}^{\beta} - \beta^+ Z_{(k)}^{\beta} = X_{(k+1)}^{\beta^+} - \beta^+ Z_{(k+1)}^{\beta^+} \;\geq\; X_{(k)}^{\beta^+} - \beta^+ Z_{(k)}^{\beta^+} = X_{(k+1)}^{\beta} - \beta^+ Z_{(k+1)}^{\beta},$$

which imply that for this specific $\delta = \beta^+ - \beta$ and the corresponding k,

$$\delta(Z_{(k+1)}^{\beta} - Z_{(k)}^{\beta}) \geq 0. \tag{3.4}$$

Now we consider the generalized Gehan-type weights,

$$W_{n,(i)}^{\beta} = G_{n,(i)}^{\beta} \equiv \frac{\sum_{j\in\widetilde{\mathcal{R}}_{(i)}^{\beta}} \rho_j}{\sum_{j=1}^{n} \rho_j} = \frac{\sum_{j\in\widetilde{\mathcal{R}}_{(i)}^{\beta}} \rho_j}{\sum_{j\in\mathcal{C}\cup\mathcal{D}} \rho_j}. \tag{3.5}$$

The last equality holds since we must have $\rho_j = 0$ for all $j \notin \mathcal{C} \cup \mathcal{D}$ for a case-cohort study. When $\rho_j \equiv 1$ for all $j = 1, \ldots, n$, which implies that the subcohort $\mathcal{C}$ is actually the entire cohort, they reduce to the Gehan weights. To simplify the notation, we omit the second argument in the estimating function $\widetilde{S}_n$ when the above Gehan-type weights are used, i.e., $\widetilde{S}_n(\beta, G_n) \equiv \widetilde{S}_n(\beta)$. It is easily seen that $\widetilde{S}_n(\beta)$ has a simpler form, and we can prove that

$$\widetilde{S}_n(\beta^+) - \widetilde{S}_n(\beta) = \left(Z_{(k+1)}^{\beta} - Z_{(k)}^{\beta}\right)\left(\Delta_{(k+1)}^{\beta}\eta_{(k)}^{\beta}\rho_{(k)}^{\beta} + \Delta_{(k)}^{\beta}\eta_{(k+1)}^{\beta}\rho_{(k+1)}^{\beta}\right) \Big/ \sum_{j\in\mathcal{C}\cup\mathcal{D}} \rho_j\ . \tag{3.6}$$

The detailed derivation is in the Appendix.

When $\delta = \beta^+ - \beta > 0$, i.e., β increases to β^+, from (3.4) we know that

$$Z_{(k+1)}^{\beta} - Z_{(k)}^{\beta} \geq 0.$$

Then from (3.6) we have $\widetilde{S}_n(\beta^+) - \widetilde{S}_n(\beta) \geq 0$. We thus conclude that, with the generalized Gehan-type weights defined in (3.5), the estimating function $\widetilde{S}_n(\beta)$ is always a non-decreasing function of β.

4 Linear Programming

With the generalized Gehen's weights defined in (3.5), from (3.1) we can write $\widetilde{S}_n(\boldsymbol{\beta})$ as

$$\begin{aligned}
\widetilde{S}_n(\boldsymbol{\beta}) \sum_{j\in\mathcal{C}\cup\mathcal{D}} \rho_j &= \sum_{i\in\mathcal{C}\cup\mathcal{D}} \Delta_{(i)}^{\beta}\left(\sum_{j\in\widetilde{\mathcal{R}}_{(i)}^{\beta}} \rho_j \mathbf{Z}_{(i)}^{\beta} - \sum_{j\in\widetilde{\mathcal{R}}_{(i)}^{\beta}} \rho_j \mathbf{Z}_j\right) \\
&= \sum_{i\in\mathcal{C}\cup\mathcal{D}} \Delta_{(i)}^{\beta}\left(\sum_{j\in\mathcal{C}\cup\mathcal{D}} \rho_j\eta_j I\big(X_j - \boldsymbol{\beta}'\mathbf{Z}_j \geq X_{(i)}^{\beta} - \boldsymbol{\beta}'\mathbf{Z}_{(i)}^{\beta}\big)\big(\mathbf{Z}_{(i)}^{\beta} - \mathbf{Z}_j\big)\right) \\
&= \sum_{i\in\mathcal{C}\cup\mathcal{D}} \sum_{j\in\mathcal{C}\cup\mathcal{D}} \Delta_i\rho_j\eta_j I\big(X_j - \boldsymbol{\beta}'\mathbf{Z}_j \geq X_i - \boldsymbol{\beta}'\mathbf{Z}_i\big)\big(\mathbf{Z}_i - \mathbf{Z}_j\big)\ .
\end{aligned} \tag{4.1}$$

We can see that the right hand side of equation (4.1) is the gradient of function

$$\sum_{i\in\mathcal{C}\cup\mathcal{D}} \sum_{j\in\mathcal{C}\cup\mathcal{D}} \Delta_i\rho_j\eta_j \big\{(X_j - \boldsymbol{\beta}'\mathbf{Z}_j) - (X_i - \boldsymbol{\beta}'\mathbf{Z}_i)\big\}^+,$$

where $\{a\}^+ = \max(a, 0)$. Since $\sum_{j \in \mathcal{C} \cup \mathcal{D}} \rho_j$ is a constant and $\widetilde{S}_n(\boldsymbol{\beta})$ is non-decreasing in each component of $\boldsymbol{\beta}$, the above function is convex and can thus be minimized. Similar to Lin et al. (1998) for cohort data, a minimizer of this function is a solution (or an approximate solution) to $\widetilde{S}_n(\boldsymbol{\beta}) = 0$, and can be achieved by solving the following linear programming problem with unknown variables γ_{ij} and $\boldsymbol{\beta}$:

$$
\begin{aligned}
&\min \sum_{i \in \mathcal{C} \cup \mathcal{D}} \sum_{j \in \mathcal{C} \cup \mathcal{D}} \Delta_i \rho_j \eta_j \gamma_{ij}, \quad \text{subject to} \\
&\gamma_{ij} \geq 0, \\
&\gamma_{ij} \geq (X_j - \boldsymbol{\beta}'\mathbf{Z}_j) - (X_i - \boldsymbol{\beta}'\mathbf{Z}_i),
\end{aligned}
$$

for all $i, j \in \mathcal{C} \cup \mathcal{D}$. Again when $\mathcal{C}$ becomes the entire cohort, we have $\rho_j = \eta_j = 1$ for all $j = 1, \ldots n$, and the above optimization problem reduces to that for cohort data in Lin et al. (1998), see also Kalbfleisch and Prentice (2002), Section 7.4.3.

Linear programming is a classical optimization problem with well developed algorithm available, e.g. simplex method. But apparently the above linear programming problem has $n^2_{\mathcal{C} \cup \mathcal{D}} + d$ unknown variables γ_{ij} and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_d)'$, and $n^2_{\mathcal{C} \cup \mathcal{D}}$ linear constrains (not including those non-negative constrains for γ_{ij}). Even for a moderate sample size $n_{\mathcal{C} \cup \mathcal{D}}$, the scale of the optimization problem can easily go beyond tens or hundreds of thousands. Thus whether the problem is numerically solvable will largely depend on what computing facility is available. For the time being, it is not feasible to use a regular PC to solve such a linear programming problem with a sample size in the range of hundreds that is commonly seen in practice.

Since the dimension of equation $\widetilde{S}_n(\boldsymbol{\beta}) = 0$ is usually much lower than the sample size, the linear programming creates too many unknown variables and thus does not seem to be an efficient way of formulating the problem, though such a formulation is mathematically beautiful. Considering the sparse coefficient matrix of the constrains, it may be possible to develop an more computationally efficient algorithm for the above linear programming problem. We do not pursue along this line here. Instead, we propose a Newton-type algorithm to solve $\widetilde{S}_n(\boldsymbol{\beta}) = 0$ directly using the monotone property of the step function $\widetilde{S}_n(\boldsymbol{\beta})$.

5 The Hybrid Newton-Type Method

In this section, we propose a fast algorithm in obtaining a root of the estimating function for a multiple censored regression model. We only consider the generalized Gehan-type weights. Arbitrary weights will be discussed later in Section 7.

To illustrate the idea, we work with one-dimensional covariate first. For any give β, we want to find out the two closest points around β, β^- and β^+ satisfying $\beta^- < \beta < \beta^+$, at which the estimating function $\widetilde{S}_n$ jumps. We know that as β varies, $\widetilde{S}_n(\beta)$ may change only if an interchange (or more interchanges) in ranks occurs between neighboring order statistics of the residuals $X_i - \beta Z_i$, $i = 1, \cdots, n$. We only need to consider neighboring interchanges in ranks to obtain either β^- or β^+. Let $\beta^+ = \beta + \delta^+$ and $\beta^- = \beta + \delta^-$. Then δ^+ is obtained by

$$\begin{aligned} \delta^+ &= \min_k \left\{ \frac{\epsilon^\beta_{(k+1)} - \epsilon^\beta_{(k)}}{Z^\beta_{(k+1)} - Z^\beta_{(k)}} :\ \epsilon^\beta_{(k+1)} > \epsilon^\beta_{(k)},\ Z^\beta_{(k+1)} > Z^\beta_{(k)}, \right. \\ &\qquad\qquad \left. \Delta^\beta_{(k+1)}\eta^\beta_{(k)}\rho^\beta_{(k)} + \Delta^\beta_{(k)}\eta^\beta_{(k+1)}\rho^\beta_{(k+1)} \neq 0 \right\} \\ &= \frac{\epsilon^\beta_{(k^{\delta^+}+1)} - \epsilon^\beta_{(k^{\delta^+})}}{Z^\beta_{(k^{\delta^+}+1)} - Z^\beta_{(k^{\delta^+})}} . \end{aligned} \tag{5.1}$$

Here we denote the corresponding k at which the minimizing is achieved by k^{δ^+}. The last two constrains are determined by equation (3.6), and the first constraint avoids the value of β at which $\widetilde{S}_n(\beta)$ jumps. Similarly δ^- is obtained by

$$\begin{aligned} \delta^- &= \max_k \left\{ \frac{\epsilon^\beta_{(k+1)} - \epsilon^\beta_{(k)}}{Z^\beta_{(k+1)} - Z^\beta_{(k)}} :\ \epsilon^\beta_{(k+1)} > \epsilon^\beta_{(k)},\ Z^\beta_{(k+1)} < Z^\beta_{(k)}, \right. \\ &\qquad\qquad \left. \Delta^\beta_{(k+1)}\eta^\beta_{(k)}\rho^\beta_{(k)} + \Delta^\beta_{(k)}\eta^\beta_{(k+1)}\rho^\beta_{(k+1)} \neq 0 \right\} \\ &= \frac{\epsilon^\beta_{(k^{\delta^-}+1)} - \epsilon^\beta_{(k^{\delta^-})}}{Z^\beta_{(k^{\delta^-}+1)} - Z^\beta_{(k^{\delta^-})}} . \end{aligned} \tag{5.2}$$

Here we denote the corresponding k at which the maximizing is achieved by k^{δ^-}. Note that $\delta^+ > 0$ and $\delta^- < 0$.

From (3.6) we know that

$$
\begin{aligned}
&\widetilde{S}_n(\beta+\delta^+)\\
&=\widetilde{S}_n(\beta)+\left(Z^{\beta}_{(k^{\delta^+}+1)}-Z^{\beta}_{(k^{\delta^+})}\right)\left(\Delta^{\beta}_{(k^{\delta^+}+1)}\eta^{\beta}_{(k^{\delta^+})}\rho^{\beta}_{(k^{\delta^+})}+\Delta^{\beta}_{(k^{\delta^+})}\eta^{\beta}_{(k^{\delta^+}+1)}\rho^{\beta}_{(k^{\delta^+}+1)}\right)\Big/\sum_{j\in\mathcal{C}\cup\mathcal{D}}\rho_j\\
&=\widetilde{S}_n(\beta)+d^+,
\end{aligned}
\tag{5.3}
$$

and

$$
\begin{aligned}
&\widetilde{S}_n(\beta+\delta^-)\\
&=\widetilde{S}_n(\beta)+\left(Z^{\beta}_{(k^{\delta^-}+1)}-Z^{\beta}_{(k^{\delta^-})}\right)\left(\Delta^{\beta}_{(k^{\delta^-}+1)}\eta^{\beta}_{(k^{\delta^-})}\rho^{\beta}_{(k^{\delta^-})}+\Delta^{\beta}_{(k^{\delta^-})}\eta^{\beta}_{(k^{\delta^-}+1)}\rho^{\beta}_{(k^{\delta^-}+1)}\right)\Big/\sum_{j\in\mathcal{C}\cup\mathcal{D}}\rho_j\\
&=\widetilde{S}_n(\beta)+d^-.
\end{aligned}
\tag{5.4}
$$

By (5.1) and (5.2) we have $d^+>0$ and $d^-<0$.

Then we use the slope of the line formed by $(\beta+\delta^-,\widetilde{S}_n(\beta+\delta^-))$ and $(\beta+\delta^+,\widetilde{S}_n(\beta+\delta^+))$ as an approximation of the slope of $\widetilde{S}_n$ at β. The updated β at $(m+1)$-th iteration can then be expressed as

$$
\beta^{(m+1)} \;=\; \beta^{(m)}-\frac{\delta^{+(m)}-\delta^{-(m)}}{d^{+(m)}-d^{-(m)}}\,\widetilde{S}_n\left(\beta^{(m)}\right)\;. \tag{5.5}
$$

Figure 1 illustrates the above iteration that is similar to the discretized Newton method. From (5.1) - (5.5) we know that the slope is always positive. Hence we have that if $\beta^{m)}$ is a point such that $\widetilde{S}_n(\beta^{(m)})>0$, then $\beta^{(m+1)}<\beta^{(m)}$, and vice versa. In a situation that $\delta^{+(m)}-\delta^{-(m)}$ is much bigger than $d^{+(m)}-d^{-(m)}$, i.e. a very flat slope, the iteration may end up with divergence. But we can well control this situation based on the monotonicity of $\widetilde{S}_n(\beta)$. Starting from the second iteration after choosing an initial value of β, $\beta^{(0)}$, if $\beta^{(m+1)}$ obtained from (5.5) satisfies $|\widetilde{S}_n(\beta^{(m+1)})|\geq\max\{|\widetilde{S}_n(\beta^{(m)})|,\ |\widetilde{S}_n(\beta^{(m-1)})|\}$, $m\geq 1$, then instead we update β by

$$
\beta^{(m+1)}=\left\{\begin{array}{l}\beta^{(m)}+\delta^{-(m)}I(\widetilde{S}_n(\beta^{(m)})>0)+\delta^{+(m)}I(\widetilde{S}_n(\beta^{(m)})<0)\\ \qquad\text{or}\\ \beta^{(m)}-\widetilde{S}_n(\beta^{(m)})\end{array}\right. \tag{5.6}
$$

whichever yields smaller $|\widetilde{S}_n(\beta^{(m+1)})|$, or the first if both yields the same $|\widetilde{S}_n(\beta^{(m+1)})|$. Here the first option is to move the next β to the far end of the interval in which $\widetilde{S}_n$ is a constant, while the second option corresponds to the parallel-chord method with constant slope 1, see e.g. Ortega and Rheinboldt (1970). This is why we name the proposed method as the hybrid Newton-type method. The initial value $\beta^{(0)}$ can be chosen from a parametric model. In the analysis conducted in the next section, we choose lognormal error distribution for the linear model (2.1). Sometimes the initial value can be far right or left, then either δ^+ or δ^- does not exist by its definition in either (5.1) or (5.2). We then use the first option in (5.6) to update β. We always use the order of index set of observed data to break ties in ranks if they happen to appear.

For the multivariate case, we propose using a diagonal matrix with diagonal elements determined by (5.5) to replace the Jacobian matrix in the Newton-Raphson method. Suppose $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_d)' \in \mathbb{R}^d$. Let

$$\widetilde{\mathbf{S}}_n(\boldsymbol{\beta}) = (\widetilde{S}_{n1}(\boldsymbol{\beta}), \cdots, \widetilde{S}_{nd}(\boldsymbol{\beta}))',$$

where the ℓth term in $\widetilde{\mathbf{S}}_n(\boldsymbol{\beta})$ is

$$\widetilde{S}_{n\ell}(\boldsymbol{\beta}) \equiv \sum_{i=1}^{n} \int G_n(u, \boldsymbol{\beta}) \{Z_{i\ell} - \widetilde{Z}_{\cdot\ell}(u, \boldsymbol{\beta})\} \mathrm{d}N_i(u + \boldsymbol{\beta}'\mathbf{Z}_i) \tag{5.7}$$

with

$$\widetilde{Z}_{\cdot\ell}(u, \boldsymbol{\beta}) \quad = \quad \frac{\sum_{j=1}^{n} \rho_j Z_{j\ell} Y_j(u + \boldsymbol{\beta}'\mathbf{Z}_j)}{\sum_{j=1}^{n} \rho_j Y_j(u + \boldsymbol{\beta}'\mathbf{Z}_j)}$$

For a given $\boldsymbol{\beta}$, the slope of the ℓ-th component $\widetilde{S}_{n\ell}(\boldsymbol{\beta})$ to β_ℓ at $\boldsymbol{\beta}$ can be calculated as follows: Let $X_{i\ell} = X_i - \beta_1 Z_{i1} \cdots - \beta_{\ell-1} Z_{i,\ell-1} - \beta_{\ell+1} Z_{i,\ell+1} \cdots - \beta_d Z_{id}$ and treat it as the X_i in the univariate case. Then compute the slope as in (5.5) by

$$\alpha_\ell = \frac{d_\ell^+ - d_\ell^-}{\delta_\ell^+ - \delta_\ell^-}$$

with $d_\ell^+, d_\ell^-, \delta_\ell^+$ and δ_ℓ^- defined as in (5.1) - (5.4) for $(\beta_\ell, \widetilde{S}_{n\ell}(\boldsymbol{\beta}))$.

Note that with the Jacobian replaced by such a diagonal matrix, simultaneous updating $\boldsymbol{\beta}$ in an iteration is equivalent to updating β_1 through β_d one at a time. A modification can thus be imposed for divergence control by (5.6). The algorithm stops when the change in each component of β is less than a pre-specified quantity (e.g., 10^{-5}). To reduce the number of unnecessary iterations, we also declare convergence when $n^{-1/2}\widetilde{S}_{n\ell}(\boldsymbol{\beta}^{(m)})$ is less than a small quantity (e.g., 10^{-4}) for all ℓ. The method works well in our numerical examples.

6 Numerical Examples

6.1 Simulations

We report on simulations for multivariate regressions to evaluate the algorithm and the finite sample performance of the proposed estimator. Two different sample size of $n = 500$ and $n = 5000$ are considered. The simulation settings are: $Z_1 \sim N(0, 1)$, $Z_2 \sim$ Bernoulli(p_2), error term ϵ follows the exponential distribution with mean 1, failure time T is from the model $T = \beta_1 Z_1 + \beta_2 Z_2 + \epsilon$, and censoring time C follows $\exp(C) \sim$ Exponential(λ) where λ is chosen so that the number of failures is about 100 under either $n = 500$ or $n = 5000$. We choose $p_2 = 0.2$, $\beta_1 = 1$, and $\beta_2 = -1$. In addition, we define the distribution of a surrogate Z_2^* of Z_2 using $\eta = P(Z_2^* = 1|Z_2 = 1)$ and $\nu = P(Z_2^* = 0|Z_2 = 0)$. We chose (η, ν) : $(\eta, \nu) \in \{(0.7, 0.7), (0.9, 0.9)\}$. Thus $Z^* \sim$ Bernoulli($(1-\nu)(1-p_Z) + \eta p_Z$). The subcohort is either a simple random sample of the cohort or a stratified sample selected by the independent Bernoulli sampling with selection probability $\pi(Z_2^*)$ such that approximately equal numbers of subjects are selected from the two strata: $\{Z_2^* = 1\}$ and $\{Z_2^* = 0\}$. Due to the fact that Z_1 and Z_2 are uncorrelated, we expect similar distribution of Z_1 among the strata $\{Z_2^* = 1\}$ and $\{Z_2^* = 0\}$. Simulation results that compare the case-cohort studies and the full cohort study are given in Table 1. The 90% quantile and range of the number of iterations needed for convergence are also listed in the table. In all cases, the bias of the estimator for $\boldsymbol{\beta}$ is negligible and the variance estimator performs well. Coverages are also satisfactory. Stratification provides better efficiency, and higher correlation between the surrogate Z_2^* and Z_2 leads to slightly higher efficiency. The program is written in R language. For $n = 500$, The convergence criterion is either $|\boldsymbol{\beta}^{(m)}| < 0.00001$ or $|n^{-1/2}\widetilde{\mathbf{S}}_n(\boldsymbol{\beta}^{(m)})| < 0.0001$ where $|\cdot|$ is the maximum norm. For $n = 5000$, The convergence criterion is either $|\boldsymbol{\beta}^{(m)}| < 0.000001$ or $|n^{-1/2}\widetilde{\mathbf{S}}_n(\boldsymbol{\beta}^{(m)})| < 0.00001$. More stringent criteria for convergence seems to have little effect on results except number of iterations to converge.

6.2 Illustration of a Real Data Analysis

The specific data we consider in this article are from the prostate cancer study conducted by investigators at the University of Michigan. The patients had carcinoma of the prostate and were treated with radiation therapy between year 1987 and 2000. The endpoint of interest is clinical recurrence (local recurrence or distant metastasis). For patients who received hormonal therapy as a salvage therapy, we also treat them as failures. Patients who were free of clinical recurrence are considered to be censored at the last date of contact or at the time of death from other causes. There are total 427 patients with 110 failures based on the above definition. The median follow-up time was 47.5 months (ranged from 0.7 month to 144.5 months). Among patients who experienced tumor recurrence, the median time to event was 30 months (ranged from 0.7 months to 102 months).

The baseline covariates included in the model are the baseline prostate specific antigen (PSA) value, Tumor stage, and Gleason score. The baseline PSA (bPSA) values have a median value of 7.9 (ranged from 0.4 to 228.5). We transform bPSA by log(1+bPSA). Tumor stage is a categorical variable with 3 levels: T1, T2, and T3-T4. There are 107 patients with T1, 262 with T2, and 58 with T3-T4. Gleason score ranged from 2 to 10 and is treated as a continuous variable.

A censored linear regression model is fitted using this data and the results are tabulated in Table 2. The full cohort method uses all subjects' information while the simple random sample (SRS) method takes a subcohort of size around 100 and the stratification (Strat) method takes about equal number of subjects from T1, T2 and T3-T4 groups. The algorithm takes 22 iterations for the full cohort analysis to converge, and 34 for SRS and 27 for Strat to converge. All methods yield the same conclusion and their point estimates are similar to each other. All covariates have significant effects on the time to tumor recurrence.

7 Discussion

For arbitrary weight function W_n^{β}, a similar idea to Jin, Lin, Wei, and Ying (2003) applies. Define a weighted estimating function for $\boldsymbol{\beta}$ by

$$\sum_{i \in \mathcal{C} \cup \mathcal{D}} \left[\Delta_{(i)}^{\beta} \frac{W_{n,(i)}^{\tilde{\beta}}}{G_{n,(i)}^{\tilde{\beta}}} G_{n,(i)}^{\beta} \left\{ Z_{(i)}^{\beta} - \widetilde{Z}_{(i)}^{\beta} \right\} \right],$$

where $\tilde{\boldsymbol{\beta}}$ can be the generalized Gehan-type estimator. Again, it can be proved that the above estimating function is componentwise monotone non-decreasing. Thus the corresponding estimator of $\boldsymbol{\beta}$ can be obtained by the proposed hybrid Newton-type method.

Appendix

Proof of (3.6)

To evaluate the difference between $\widetilde{S}_n(\beta, W_n)$ and $\widetilde{S}_n(\beta^+, W_n)$, it helps to see how $\widetilde{\mathcal{R}}_{(i)}^{\beta+}$ changes when β changes to β^+. To simplify notation, we replace $n_{\mathcal{C}\cup\mathcal{D}}$ by n and still keep its meaning as the number of completely observed subjects.

Apparently,

$$\begin{aligned}
\widetilde{\mathcal{R}}_{(j)}^{\beta+} &= \widetilde{\mathcal{R}}_{(j)}^{\beta}, \text{ for } j \notin \{k, k+1\}, \qquad (7.1)\\
\widetilde{\mathcal{R}}_{(k)}^{\beta+} &= \{\eta_{(k)}^{\beta+}, \eta_{(k+1)}^{\beta+}, \eta_{(k+2)}^{\beta+}, \cdots, \eta_{(n)}^{\beta+}\}\\
&= \{\eta_{(k+1)}^{\beta}, \eta_{(k)}^{\beta}, \eta_{(k+2)}^{\beta}, \cdots, \eta_{(n)}^{\beta}\}\\
&= \widetilde{\mathcal{R}}_{(k)}^{\beta} = \{\widetilde{\mathcal{R}}_{(k+2)}^{\beta}, \eta_{(k)}^{\beta}, \eta_{(k+1)}^{\beta}\},\\
\widetilde{\mathcal{R}}_{(k+1)}^{\beta+} &= \{\eta_{(k+1)}^{\beta+}, \eta_{(k+2)}^{\beta+}, \cdots, \eta_{(n)}^{\beta+}\} = \{\widetilde{\mathcal{R}}_{(k+2)}^{\beta}, \eta_{(k)}^{\beta}\}, \qquad (7.2)
\end{aligned}$$

where

$$\widetilde{\mathcal{R}}_{(k+2)}^{\beta} = \{\eta_{(k+2)}^{\beta}, \cdots, \eta_{(n)}^{\beta}\}.$$

To simplify the calculation, we use

$$W_n = G_n \cdot \sum_{j=1}^{n} \rho_j = \sum_{j \in \widetilde{\mathcal{R}}_{(k)}^{\beta}} \rho_j \ ,$$

which will not affect the conclusion since $\sum_{j=1}^{n} \rho_j$ is a constant.

Now write $\widetilde{S}_n(\beta, W_n)$ and $\widetilde{S}_n(\beta^+, W_n)$ as

$$\widetilde{S}_n(\beta, W_n) = \sum_{i=1}^{n} \Delta_{(i)}^{\beta} \Bigg\{ \sum_{j \in \widetilde{\mathcal{R}}_{(i)}^{\beta}} \rho_j Z_{(i)}^{\beta} - \sum_{j \in \widetilde{\mathcal{R}}_{(i)}^{\beta}} \rho_j Z_j \Bigg\} \qquad (7.3)$$

and

$$\widetilde{S}_n(\beta^+, W_n) = \sum_{i=1}^{n} \Delta_{(i)}^{\beta+} \Bigg\{ \sum_{j \in \widetilde{\mathcal{R}}_{(i)}^{\beta+}} \rho_j Z_{(i)}^{\beta+} - \sum_{j \in \widetilde{\mathcal{R}}_{(i)}^{\beta+}} \rho_j Z_j \Bigg\}. \qquad (7.4)$$

Denote

$$C = \sum_{i \notin \{k, k+1\}} \Delta_{(i)}^{\beta} \Bigg\{ \sum_{j \in \widetilde{\mathcal{R}}_{(i)}^{\beta}} \rho_j Z_{(i)}^{\beta} - \sum_{j \in \widetilde{\mathcal{R}}_{(i)}^{\beta}} \rho_j Z_j \Bigg\} \qquad (7.5)$$

Then by utilizing the changes in risk sets and labels as listed in (3.2) - (3.3) and (7.1) - (7.2), we can write $\widetilde{S}_n(\beta, W_n)$ as

$$\begin{aligned}\widetilde{S}_n(\beta, W_n) &= C + \Delta^{\beta}_{(k)}\left\{ \sum_{j\in\widetilde{\mathcal{R}}^{\beta}_{(k+2)}} \rho_j(Z^{\beta}_{(k)} - Z_j) + \eta^{\beta}_{(k+1)}\rho^{\beta}_{(k+1)}\left(Z^{\beta}_{(k)} - Z^{\beta}_{(k+1)}\right)\right\} \\ &\quad + \Delta^{\beta}_{(k+1)} \sum_{j\in\widetilde{\mathcal{R}}^{\beta}_{(k+2)}} \rho_j\left(Z^{\beta}_{(k+1)} - Z_j\right) \ . \qquad (7.6)\end{aligned}$$

Similarly,

$$\begin{aligned}\widetilde{S}_n(\beta^+, W_n) &= C + \Delta^{\beta+}_{(k)}\left\{ \sum_{j\in\widetilde{\mathcal{R}}^{\beta+}_{(k+2)}} \rho_j(Z^{\beta+}_{(k)} - Z_j) + \eta^{\beta+}_{(k+1)}\rho^{\beta+}_{(k+1)}\left(Z^{\beta+}_{(k)} - Z^{\beta+}_{(k+1)}\right)\right\} \\ &\quad + \Delta^{\beta+}_{(k+1)} \sum_{j\in\widetilde{\mathcal{R}}^{\beta+}_{(k+2)}} \rho_j\left(Z^{\beta+}_{(k+1)} - Z_j\right) \\ &= C + \Delta^{\beta}_{(k+1)}\left\{ \sum_{j\in\widetilde{\mathcal{R}}^{\beta}_{(k+2)}} \rho_j(Z^{\beta}_{(k+1)} - Z_j) + \eta^{\beta}_{(k)}\rho^{\beta}_{(k)}\left(Z^{\beta}_{(k+1)} - Z^{\beta}_{(k)}\right)\right\} \\ &\quad + \Delta^{\beta}_{(k)} \sum_{j\in\widetilde{\mathcal{R}}^{\beta}_{(k+2)}} \rho_j\left(Z^{\beta}_{(k)} - Z_j\right) \ . \qquad (7.7)\end{aligned}$$

Hence subtract (7.6) from (7.7), we obtain

$$\begin{aligned}&\widetilde{S}_n(\beta^+, W_n) - \widetilde{S}_n(\beta, W_n) \\ &\quad = \Delta^{\beta}_{(k+1)}\eta^{\beta}_{(k)}\rho^{\beta}_{(k)}\left(Z^{\beta}_{(k+1)} - Z^{\beta}_{(k)}\right) - \Delta^{\beta}_{(k)}\eta^{\beta}_{(k+1)}\rho^{\beta}_{(k+1)}\left(Z^{\beta}_{(k)} - Z^{\beta}_{(k+1)}\right) \\ &\quad = \left(Z^{\beta}_{(k+1)} - Z^{\beta}_{(k)}\right)\left(\Delta^{\beta}_{(k+1)}\eta^{\beta}_{(k)}\rho^{\beta}_{(k)} + \Delta^{\beta}_{(k)}\eta^{\beta}_{(k+1)}\rho^{\beta}_{(k+1)}\right) \ .\end{aligned}$$

REFERENCES

Barlow, W. E., Ichikawa, L., Rosner, D., and Izumi, S. (1999), Analysis of Case-cohort Designs. *Journal of Clinical Epidemiology*, 52, 1165-1172.

Borgan, Ø., Langholz, B., Samuelsen, S.O., Goldstein, L., and Pogoda, J. (2000), Exposure stratified case-cohort designs. *Lifetime Data Analysis*, 6, 39-58.

Fygenson, M. and Ritov, Y. (1994). Monotone estimating equations for censored data. *Annals of Statistics*, 22, 732-746.

Huang, Y. (2002), Calibration regression of censored lifetime medical cost. *Journal of the American Statistical Association*, 97, 318-327.

Jin, Z., Lin, D. Y., Wei, L. J., and Ying, Z. (2003), Rank-based inference for the accelerated failure time model. *Biometrika*, 90, 341-353

Kalbfleisch, J. D. and Prentice R. L. (2002). *The Statistical Analysis of Failure Time Data*, 2nd Ed., John Wiley & Sons, Inc., New Jersey.

Lin, D. Y. and Geyer, C. J. (1992). Computational methods for semiparametric linear regression with censored data. *Journal of Computational and Graphical Statistics*, 1, 77-90.

Lin, D. Y., Wei, L. J., and Ying, Z. (1998). Accelerated failure time models for counting processes. *Biometrika*, 85, 605-618.

Nan, B., Yu, M., and Kalbfleisch, J. D. (2004), Censored Linear Regression for Case-Cohort Studies. Technical Report. Biostatistics Department, University of Michigan. Submitted.

Ortega, J. M. and Rheinboldt, W. C. (1970), *Iterative Solution of nonlinear Equations in Several Variables*, Academic Press, Inc., New York.

Prentice, R. L. (1986), A case-cohort design for epidemiologic cohort studies and disease prevention trials, *Biometrika*, 73, 1-11.

Ritov, Y. (1990), Estimation in a Linear Regression Model with Censored Data, *Annals of Statistics*, 18, 303-328.

Self, S. G. and Prentice, R. L. (1988), Asymptotic distribution theory and efficiency results for case-cohort studies, *Annals of Statistics*, 16, 64-81.

Tsiatis, A. A. (1990), Estimating regression parameters using linear rank tests for censored data, *Annals of Statistics*, 18, 354-372.

Ying, Z. (1993), A large sample study of rank estimation for censored regression data, *Annals of Statistics*, 21, 76-99.

Table 1: Simulation Results Based on 200 Data Sets under the Model $\log T = \beta_1 Z_1 + \beta_2 Z_2 + \epsilon$ ($\beta_1 = 1, \beta_2 = -1, p_Z = 0.2$)

	Full	SRS	Strat1†	Strat2‡
	Sample size = 500, Failure≈ 100, Subcohort=100			
Mean of $\hat{\boldsymbol{\beta}}$	(1.002, -0.996)	(1.002, -1.020)	(1.015, -0.997)	(1.018, -0.993)
SE of $\hat{\boldsymbol{\beta}}$	(0.063, 0.118)	(0.091, 0.212)	(0.107, 0.185)	(0.116, 0.181)
Mean. $(\hat{se}(\hat{\beta}_1), \hat{se}(\hat{\beta}_2))$	(0.065, 0.118)	(0.101, 0.218)	(0.112, 0.192)	(0.117, 0.183)
95% CI cover β_1.	95.5%	91%	96.5%	95.5%
95% CI cover β_2.	93.5%	92%	95 %	94 %
iterations§	14(2-25)	39(3-51)	14(3-22)	16(4-22)
	Sample size = 5000, Failure≈ 100, Subcohort=250			
Mean of $\hat{\boldsymbol{\beta}}$	(1.007, -1.004)	(1.018, -1.004)	(1.012, -0.983)	(1.021, -0.999)
SE of $\hat{\boldsymbol{\beta}}$	(0.039, 0.058)	(0.088, 0.130)	(0.085, 0.147)	(0.089, 0.129)
Mean. $(\hat{se}(\hat{\beta}_1), \hat{se}(\hat{\beta}_2))$	(0.040, 0.056)	(0.090, 0.162)	(0.088, 0.154)	(0.103, 0.131)
95% CI cover β_1.	96%	92.5%	94.5%	95%
95% CI cover β_2.	94%	94.5%	93.5%	94.5%
iterations§	18(3-25)	70(2 - 120)	21(5 - 34)	25(6-87)

§ 90% quantile and range of required iterations for convergence;
† Strat1 is stratified with $\eta = \nu = 0.7$;
‡ Strat2 is stratified with $\eta = \nu = 0.9$;

Table 2: Results for the prostate cancer study.

	Parameter estimates (S.E.) under different sampling			
	Stage I	Stage II	log(1+bPSA)	Gleason Score
Full	1.420	0.945	-0.579	-0.285
	(0.254)	(0.207)	(0.127)	(0.068)
SRS†	1.395	0.920	-0.663	-0.347
	(0.354)	(0.306)	(0.193)	(0.084)
Strat‡	1.736	1.090	-0.511	-0.238
	(0.160)	(0.230)	(0.295)	(0.135)

† SRS with 100 subjects in subcohort;
‡ Strat is stratified with 43 subjects in the Tumor stage 1 group, 29 subjects in the Tumor stage 2 group, and 29 in the Tumor stage 3 or 4 group.

Figure 1: Numerical approximation of the slope used in the hybrid Newton method