Lifetime Data Analysis, 4, 355–391 (1998)

Semiparametric Efficient Estimation in the Generalized Odds-Rate Class of Regression Models for Right-Censored Time-to-Event Data

DANIEL O. SCHARFSTEIN dscharf@jhsph.edu
Department of Biostatistics, Johns Hopkins School of Hygiene and Public Health, Baltimore, MD 21205

ANASTASIOS A. TSIATIS
Department of Statistics, North Carolina State University, Raleigh, NC 27695

PETER B. GILBERT
Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115

Received September 12, 1996; Revised March 16, 1998; Accepted May 11, 1998

Abstract. The generalized odds-rate class of regression models for time to event data is indexed by a non-negative constant ρ and assumes that

$$g_\rho(S(t|Z)) = \alpha(t) + \beta' Z$$

where $g_\rho(s) = \log(\rho^{-1}(s^{-\rho} - 1))$ for $\rho > 0$, $g_0(s) = \log(-\log s)$, $S(t|Z)$ is the survival function of the time to event for an individual with qx1 covariate vector Z, β is a qx1 vector of unknown regression parameters, and $\alpha(t)$ is some arbitrary increasing function of t. When $\rho = 0$, this model is equivalent to the proportional hazards model and when $\rho = 1$, this model reduces to the proportional odds model. In the presence of right censoring, we construct estimators for β and $\exp(\alpha(t))$ and show that they are consistent and asymptotically normal. In addition, we show that the estimator for β is semiparametric efficient in the sense that it attains the semiparametric variance bound.

Keywords: Nonparametric maximum likelihood, proportional hazards model, proportional odds model, survival analysis

1. Introduction

In the analysis of clinical trials, regression models are often used to assess the relationship between a time-to-event outcome and covariates. The most widely used of these models is the proportional hazards model introduced by Cox (1972). More recently, Bennett (1983) presented the proportional odds regression model. A large class of models which includes both the proportional hazards and proportional odds models was discussed by Harrington and Fleming (1982), Clayton and Cuzick (1986) and Dabrowska and Doksum (1988a). This class is referred to as the generalized odds-rate class of regression models and is indexed by a non-negative constant ρ. If we let T denote the time to event and Z be a corresponding qx1 vector of covariates, then a proportional ρ-odds model within this class assumes that

$$g_\rho(S(t|Z)) = \alpha(t) + \beta' Z \tag{1}$$

where $g_\rho(s) = \log(\rho^{-1}(s^{-\rho} - 1))$ for $\rho > 0$, $g_0(s) = \log(-\log s)$, $S(t|Z)$ is the survival function of T given Z, β is a qx1 vector of unknown regression parameters, and $\alpha(t)$ is some arbitrary increasing function of t. Note that when $\rho = 0$, (1) is equivalent to the proportional hazards model and when $\rho = 1$, (1) reduces to the proportional odds model.

The generalized odds-rate class of regression models is a subset of the class of semiparametric linear transformation models. This latter class is comprised of models in which an unknown transformation of the time to event is assumed to be linearly related to covariates plus an independent random error with a completely specified distribution. Specifically, we can rewrite (1) as

$$\alpha(T) = -\beta' Z + \epsilon_\rho \tag{2}$$

where $\exp(\epsilon_\rho)$ is distributed according to a Pareto(ρ) distribution for $\rho > 0$ and an exponential(1) distribution for $\rho = 0$.

In the presence of right censoring, we are interested in making inferences about the regression parameters. For the more general class of semiparametric linear transformation models, at least two different approaches have been proposed to estimate these parameters. These approaches can be applied to the generalized odds-rate class. Dabrowska and Doksum (1988b) used a simulation-based approach to approximate the partial likelihood and the maximum partial likelihood estimator. Assuming that the censoring distribution does not depend on covariates, Cheng, Wei, and Ying (1995, 1997) adopted a generalized estimating approach to estimate the regression parameters and were able to establish consistency and asymptotic normality of their estimator. For the generalized odds-rate class, Harrington and Fleming (1982) proposed the G^ρ statistic for efficiently testing the proportionality parameter in the two-sample, right-censored data problem. Clayton and Cuzick (1986) showed how to compute a maximum marginal likelihood estimator for β by using a quasi-EM algorithm. Dabrowska and Doksum (1988a) considered estimation of the proportionality parameter in the two-sample, uncensored problem. For the proportional odds model, Bennett (1983) approximated the likelihood by introducing new nuisance parameters for each distinct failure time and then applied standard maximum likelihood theory to obtain a parameter estimate for β. Pettitt (1983, 1984) proposed a weighted least squares estimator which was based on maximizing an approximation to the marginal likelihood of the ranks of the censored and uncensored failure times. Parzen and Harrington (1993) developed an adaptive spline procedure with a small number of knots to estimate $\exp(\alpha(t))$, the baseline odds of failing by time t, and then applied standard likelihood techniques to estimate β. For the two-sample problem, Hsieh (1995) used empirical process approximations to formulate a non-linear regression equation which yields a generalized least squares estimator for β. Murphy, Rossini, and van der Vaart (1997) used profile likelihood techniques to construct a semiparametric efficient estimator for β and showed that their estimator was consistent and asymptotically normal. For the proportional hazards model, Cox (1975) proposed the partial likelihood estimator for the regression parameters. The asymptotic properties of this estimator were established by Tsiatis (1981) and Andersen and Gill (1982).

The goal of this paper is to find a semiparametric efficient estimator for the regression parameters in the proportional ρ-odds model when the time-to-event outcome is subject to right censoring. To do this, we first find the semiparametric variance bound (Section 3). Then, in Section 4, we define our estimator, which is similar to the one proposed by Bennett (1983). In Section 5, we show that our estimator is consistent, asymptotically normal, and attains the efficiency bound. In the process, we provide a consistent and asymptotically normal estimator for $\exp(\alpha(t))$. For the most part, our proofs mimic the techniques used by Murphy (1994, 1995) to establish the asymptotic theory for the frailty model. Our proofs differ in the asymptotic normality section where we introduce results from empirical process theory (van der Vaart and Wellner, 1996). While the theoretical arguments used in this paper are not new, it does present and validate estimation procedures for a rich class of models which serve as an alternative to the often overused proportional hazards model.

Our results are all conditional on ρ being fixed and known. Assuming ρ is known, we devote Section 6 to the results of a simulation study designed to test how well our estimation procedure works in small to moderate samples and under different degrees of censoring. Selection of ρ is a crucial step in model fitting. However, it is an unresolved issue as to whether the variability of β should be affected by the selection of ρ (Hinkley and Runger, 1984). In an example in Section 7, we discuss two methods for estimating ρ, profile likelihood estimation and a graphical procedure, but present results for β conditional on this transformation. The alternative approach to model selection (Bickel and Doksum, 1981) requires that the variability of the parameter estimates reflect estimation of the transformation. The theory necessary to handle this approach is beyond the scope of this paper. We conclude with a summary in Section 8.

2. The Proportional ρ-Odds Model

Let Z be a $q \times 1$ vector of covariates. Denote the time to event and the time of censoring by the positive, bounded random variables, T and C, respectively. Without loss of generality, we assume that T and C are bounded from above by 1. It is assumed that T and C are conditionally independent given the covariate Z. The observable time until death or censoring will be denoted by the bivariate random vector (X, Δ), where $X = \min(T, C)$ and $\Delta = 1$ if $T \leq C$ and 0 otherwise.

The distribution of the survival time T is related to the covariate Z according to the proportional ρ-odds regression model (1). We restrict attention to positive ρ since the theory of semiparametric efficient estimation in the proportional hazards model has been well established. From (1), we can write

$$S(t|Z) = (1 + \rho A(t)\exp(\beta' Z))^{-1/\rho}$$

and

$$\lambda(t|Z) = \frac{a(t)\exp(\beta' Z)}{1 + \rho A(t)\exp(\beta' Z)}$$

where $\lambda(t|Z) = -\frac{\partial \log S(t|z)}{\partial t}$, $a(t) = \exp(\alpha(t))\frac{d\alpha(t)}{dt}$ is any arbitrary, non-negative function, $A(t) = \int_0^t a(u)du = \exp(\alpha(t))$, and $r(t) = \log(a(t))$. Let β_0 and $\alpha_0(t)$ represent the true values of β and $\alpha(t)$, respectively. We assume that β_0 belongs to a compact set, say $[-b, b]^q$, $A_0(t) = \exp(\alpha_0(t))$ is an absolutely continuous, bounded, increasing function, and $|Z|$ is bounded by a constant c.

Let n be the number of individuals in the study. Suppose that associated with each individual is the random vector (X_i, Δ_i, Z_i), $i = 1, \ldots, n$, which are assumed to be independent and identically distributed. Since we are dealing with i.i.d. data, we will first focus attention on the individual censored data likelihood. Suppressing the individual subscript, we can write this likelihood as

$$\mathcal{L}(\beta, A) = [\lambda(X|Z)]^{\Delta} S(X|Z) = \frac{(a(X)\exp(\beta' Z))^{\Delta}}{(1 + \rho A(X)\exp(\beta' Z))^{1/\rho+\Delta}}$$

The log likelihood is

$$\ell(\beta, A) = \Delta(\log(a(X)) + \beta' Z) - (1/\rho + \Delta)\log(1 + \rho A(X)\exp(\beta' Z))$$

3. Semiparametric Variance Bound

The semiparametric variance bound is defined to be the supremum of the Cramer-Rao bounds for β over all regular parametric submodels (see Newey, 1990). In the above setting, a parametric submodel corresponds to a parameterization of $a(u)$, say $a(u, \eta)$, where $a(u, \eta_0) = a_0(u) = \exp(\alpha_0(t))\frac{d\alpha_0(t)}{dt}$ for some η_0. The parameters of the submodel are $\theta = (\beta', \eta')'$. So the log likelihood for a parametric submodel is

$$\ell(\beta, \eta) = \Delta(\log(a(X, \eta)) + \beta' Z) - (1/\rho + \Delta)\log(1 + \rho A(X, \eta)\exp(\beta' Z))$$

where $A(x, \eta) = \int_0^x a(u, \eta)du$. The score for β is

$$\begin{aligned} \frac{\partial \ell}{\partial \beta} &= \frac{\Delta Z}{1 + \rho A(X, \eta)\exp(\beta' Z)} - \frac{Z A(X, \eta)\exp(\beta' Z)}{1 + \rho A(X, \eta)\exp(\beta' Z)} \\ &= \int_0^1 \frac{Z}{1 + \rho A(u, \eta)\exp(\beta' Z)} dM(u, \rho, \beta, \eta) \end{aligned} \tag{3}$$

where $M(t, \rho, \beta, \eta) = N(t) - \int_0^t \frac{a(u,\eta)\exp(\beta' Z)}{1+\rho A(u,\eta)\exp(\beta' Z)} Y(u)du$ is the $\mathcal{F}_t$-counting process martingale, $N(t) = 1(X \le t, \Delta = 1)$, $Y(u) = 1(X \ge u)$, and $\mathcal{F}_t$ is the smallest sigma-algebra generated by $\{N(u), Y(u), 0 \le u \le t\}$. The score for η is

$$\begin{aligned} \frac{\partial \ell}{\partial \eta} &= \Delta\left(\frac{a_\eta(X, \eta)}{a(X, \eta)} - \frac{\rho A_\eta(X, \eta)\exp(\beta' Z)}{1 + \rho A(X, \eta)\exp(\beta' Z)}\right) - \frac{A_\eta(X, \eta)\exp(\beta' Z)}{1 + \rho A(X, \eta)\exp(\beta' Z)} \\ &= \int_0^1 \left(\frac{a_\eta(u, \eta)}{a(u, \eta)} - \frac{\rho A_\eta(u, \eta)\exp(\beta' Z)}{1 + \rho A(u, \eta)\exp(\beta' Z)}\right) dM(u, \rho, \beta, \eta) \end{aligned} \tag{4}$$

where $a_\eta(x, \eta) = \frac{\partial a(x,\eta)}{\partial \eta}$ and $A_\eta(x, \eta) = \int_0^x a_\eta(u, \eta)du$. Let S_β and S_η denote the scores for β and η evaluated at the truth, respectively.

Formally, we define the tangent set in the nonparametric direction, Λ, to be the mean square closure of the set of all random vectors AS_η, where S_η is the score for η in some regular parametric submodel and A is a conformable constant matrix with q rows. That is,

$$\Lambda = \{\ell \in \mathrm{R}^q : E[\|\ell\|^2] < \infty,\ \exists A_j S_{\eta_j} \text{ with } \lim_{j\to\infty} E[\|\ell - A_j S_{\eta_j}\|^2] = 0\},$$

where S_{η_j} is the nuisance score vector evaluated at the truth from the jth parametric submodel, A_j is a q-row conformable matrix of constants, and $\|\ell\|^2 = \ell'\ell$. Since $\frac{a_\eta(u,\eta)}{a(u,\eta)}$ can be any function of u, a plausible conjecture for Λ is

$$\begin{aligned}\Lambda \;=\; \{&f(X,\Delta,Z) : f(X,\Delta,Z)\\ &= \int_0^1 \left(w(u) - \frac{\rho\int_0^u a_0(v)\exp(\beta_0' Z) w(v)dv}{1+\rho A_0(u)\exp(\beta_0' Z)} \right) dM(u,\rho,\beta_0,\eta_0),\\ &w(u) \text{ is any } q-\text{dimensional function of } u,\ E[\|f(X,\Delta,Z)\|^2] < \infty\}\end{aligned}$$

To verify this conjecture, we need to show that there exists a parametric submodel with $S_\eta = f(X,\Delta,Z)$ for any $f(X,\Delta,Z) \in \Lambda$. Given the relationship between $w(u)$ and $f(X,\Delta,Z)$, a parametric submodel with $a(u,\eta) = a_0(u)(1+\eta' w(u))$, with $\eta_0 = 0$, has this property.

We consider Λ to be a subset of a Hilbert space of qx1 random vectors H, with inner product $E[H_1' H_2]$ and $E[H'H] < \infty$. We define the efficient score for β as $S_{eff} = S_\beta - \Pi[S_\beta|\Lambda]$, where $\Pi[\cdot|\cdot]$ is the projection operator. Since Λ can be shown to be linear, the projection of S_β on Λ, $\Pi[S_\beta|\Lambda]$, exists and is the unique element of Λ which satisfies $E[(S_\beta - \Pi[S_\beta|\Lambda])'\ell] = 0$ for all $\ell \in \Lambda$. So, provided that the $E[S_{eff}S_{eff}']$ is nonsingular, the semiparametric variance bound, V, is $(E[S_{eff}S_{eff}'])^{-1}$.

To project S_β onto Λ, we need to find the vector $w(u)$ such that

$$\begin{aligned}E\Bigg[\int_0^1 \Bigg(&\frac{Z}{1+\rho A_0(u)\exp(\beta_0' Z)} - w(u)\\ &+ \frac{\rho\int_0^u a_0(v)\exp(\beta_0' Z) w(v)dv}{1+\rho A_0(u)\exp(\beta_0' Z)}\Bigg)' dM(u,\rho,\beta_0,\eta_0)\\ &* \int_0^1 \left(w^*(u) - \frac{\rho\int_0^u a_0(v)\exp(\beta_0' Z) w^*(v)dv}{1+\rho A_0(u)\exp(\beta_0' Z)} \right) dM(u,\beta_0,\eta_0)\Bigg] = 0 \ \ \forall\, w^*\end{aligned}$$

Through algebra, we can show that the vector $w(u)$ which satisfies the above equation is a solution to the following integral equation:

$$w(u) - \int_0^1 K(u,v)w(v)dv = f(u), \quad u \in [0,1] \tag{5}$$

where

$$\begin{aligned}K(u,v) \;&=\; \frac{E[\frac{\rho(1+\rho\Delta)a_0(u)\exp(\beta_0' Z)Y(u)a_0(v)\exp(\beta_0' Z)Y(v)}{(1+\rho A_0(X)\exp(\beta_0' Z))^2}]}{E[\frac{(1+\rho\Delta)a_0(u)\exp(\beta_0' Z)Y(u)}{1+\rho A_0(X)\exp(\beta_0' Z)}]}\\ f(u) \;&=\; \frac{E[\frac{(1+\rho\Delta)Za_0(u)\exp(\beta_0' Z)Y(u)}{(1+\rho A_0(X)\exp(\beta_0' Z))^2}]}{E[\frac{(1+\rho\Delta)a_0(u)\exp(\beta_0' Z)Y(u)}{1+\rho A_0(X)\exp(\beta_0' Z)}]}\end{aligned}$$

By showing that $\max_{u\in[0,1]}\int_0^1 |K(u,v)|dv < 1$, we know that there exists a solution to this integral equation which can be found by successive approximation (Kress, 1989). So, we write

$$\max_{u\in[0,1]}\int_0^1 |K(u,v)|dv = \max_{u\in[0,1]} \frac{E[\frac{(1+\rho\Delta)a_0(u)\exp(\beta_0'Z)Y(u)}{1+\rho A_0(X)\exp(\beta_0'Z)} * \frac{\rho A_0(X)\exp(\beta_0'Z)}{1+\rho A_0(X)\exp(\beta_0'Z)}]}{E[\frac{(1+\rho\Delta)a_0(u)\exp(\beta_0'Z)Y(u)}{1+\rho A_0(X)\exp(\beta_0'Z)}]}$$

This expression is less than 1 since $\frac{\rho A_0(X)\exp(\beta_0'Z)}{1+\rho A_0(X)\exp(\beta_0'Z)} < 1$. Denote the successive approximation solution by $w_{eff}(u)$. Then, the efficient score is

$$\begin{aligned} S_{eff} \;=\; & \int_0^1 \left(\frac{Z}{1+\rho A_0(u)\exp(\beta_0'Z)} - w_{eff}(u) \right. \\ & \left. + \frac{\rho\int_0^u a_0(v)\exp(\beta_0'Z)w_{eff}(v)dv}{1+\rho A_0(u)\exp(\beta_0'Z)} \right) dM(u,\rho,\beta_0,\eta_0) \end{aligned}$$

and the variance bound is $V = E[S_{eff}S_{eff}']^{-1}$. In fact, we can show that

$$E[S_{eff}S_{eff}'] = E\left[(1+\rho\Delta)\frac{\int_0^X a_0(u)\exp(\beta_0'Z)(Z-w_{eff}(u))Z'du}{(1+\rho A_0(X)\exp(\beta_0'Z))^2}\right]$$

4. Estimation

To estimate the regression parameters and the baseline odds of failing by time t, we use full nonparametric maximum likelihood. We assume that there are no tied death times and the number of deaths, $k(n)$, increases with the sample size. This assumption is made for ease of presentation, but our results can be easily adapted to accommodate tied death times. For simplicity of notation, it will be useful to reorder the indices of the data such that $X_1,\ldots,X_{k(n)}$ represents the increasingly ordered failure times and $X_{k(n)+1},\ldots,X_n$ represents the non-decreasingly ordered censored observations. To obtain our estimates, we maximize the following extended empirical likelihood:

$$\prod_{i=1}^n \prod_{0\le t\le 1}\left(\frac{Y_i(t)\Delta A(t)\exp(\beta' Z_i)}{1+\rho A(t)\exp(\beta' Z_i)}\right)^{\Delta N_i(t)} \exp\left(-\int_0^1 \frac{Y_i(u)\exp(\beta' Z_i)}{1+\rho A(u)\exp(\beta' Z_i)}dA(u)\right) \tag{6}$$

where $\Delta A(t)$ represents the jump of A at time t, and $N_i(t)$ and $Y_i(t)$ are the failure counting process and at risk process for the ith individual, respectively. The natural logarithm of (6) is given by

$$\begin{aligned} nL_n(\beta, A) \;=\; & \sum_{i=1}^n \int_0^1 \log\left(\frac{Y_i(u)\Delta A(u)\exp(\beta' Z_i)}{1+\rho A(u)\exp(\beta' Z_i)}\right) dN_i(u) \\ & - \int_0^1 \frac{Y_i(u)\exp(\beta' Z_i)}{1+\rho A(u)\exp(\beta' Z_i)}dA(u) \end{aligned} \tag{7}$$

We see that the maximizer $\hat{A}_n$ of (7) must be a step function which takes positive steps at each of the jump times of the N_i's (i.e., the death times). Restricting ourselves to functions of this form, we can show that $\hat{A}_n$ exists and is finite.

THEOREM 1 *The maximizer of $L_n(\beta, A)$, $(\beta, A) = (\hat{\beta}_n, \hat{A}_n)$ exists and is finite.*

Proof: Note that the log empirical likelihood L_n is a continuous function of β and the jump sizes of A. That is, L_n is a continuous function on the convex, compact set $[-b, b]^q X[0, U]^{k(n)}$, where U is finite. On this set, L_n achieves its maximum. To show that a maximum exists on the set $[-b, b]^q X[0, \infty)^{k(n)}$, we show that there exists a U such that for all $(\beta, A) \in \{[-b, b]^q X[0, \infty)^{k(n)}\} \setminus \{[-b, b]^q X[0, U]^{k(n)}\}$ there is a value $(\beta, A) \in [-b, b]^q X[0, U]^{k(n)}$ which has a larger value of L_n. Consider a proof by contradiction. That is, suppose there does not exist such a U. Therefore, for all U there exists $(\beta^U, A^U) \in \{[-b, b]^q X[0, \infty)^{k(n)}\} \setminus \{[-b, b]^q X[0, U]^{k(n)}\}$ which maximizes L_n. But, we can show that $L_n(\beta^U, A^U)$ can be made arbitrarily small by increasing U, which is a contradiction. To see this, let $a_1, \ldots, a_{k(n)}$ denote the jump sizes at the death times. Then,

$$L_n(\beta, A) = \frac{1}{n}\sum_{i=1}^{k(n)} \log\left(\frac{a_i \exp(\beta' Z_i)}{1 + \rho \sum_{j \in S(X_i)} a_j \exp(\beta' Z_i)}\right)$$
$$- \frac{1}{n\rho}\sum_{i=1}^{n} \log\left(1 + \rho \sum_{j \in S(X_i)} a_j \exp(\beta' Z_i)\right)$$

where $S(u) = \{j : X_j \leq u, j = 1, \ldots, k(n)\}$. Note that $L_n(\beta, A)$ is bounded from above by $\text{sign}(\rho - 1)\log(\rho) - \frac{1}{n\rho}\sum_{i=1}^{k(n)} \log(1 + \rho a_i \exp(\beta' Z_i))$. If $(\beta, A) \in \{[-b, b]^q X[0, \infty)^{k(n)}\} \setminus \{[-b, b]^q X[0, U]^{k(n)}\}$, then there exists $1 \leq j \leq k(n)$ such that $a_j > U$. Therefore, $L_n(\beta^U, A^U) < \text{sign}(\rho - 1)\log(\rho) - \frac{1}{n\rho}\log(1 + \rho U \exp(-qbc))$ and $L_n(\beta^U, A^U)$ decreases toward negative infinity as U increases toward positive infinity. ■

Since we know that $(\hat{\beta}_n, \hat{A}_n)$ exists and is finite, we know that the maximum will occur when the derivative of L_n with respect to the jump sizes of A is equal to zero. This leads to the following equation for $\hat{A}_n$:

$$\hat{A}_n(t) = \int_0^t \left(\frac{1}{n}\sum_{i=1}^{n} \frac{(1 + \rho\Delta_i) Y_i(u) \exp(\hat{\beta}_n' Z_i)}{1 + \rho \hat{A}_n(X_i) \exp(\hat{\beta}_n' Z_i)}\right)^{-1} d\bar{N}_n(u) \tag{8}$$

where $\bar{N}_n(u) = \frac{1}{n}\sum_{i=1}^{n} N_i(u)$.

5. Asymptotics

5.1. Almost Sure Consistency

Since we are interested in almost sure (a.s.) consistency, we work with fixed realizations of the data, ω, which are assumed to lie in a set of probability one. This set, Φ, is the countable intersection of sets, Φ_i, each of probability one. Each Φ_i is a set on which the strong law of large numbers holds for some average. Our consistency proof follows Murphy's (1994) proof of a.s. consistency in a proportional hazards model with a random effect. The proof requires the definition of the following quantity which helps to mediate between $\hat{A}_n(t)$ and $A_0(t)$:

$$\bar{A}_n(t) = \int_0^t \left(\frac{1}{n} \sum_{i=1}^{n} \frac{(1+\rho\Delta_i) Y_i(u) \exp(\beta_0' Z_i)}{1+\rho A_0(X_i)\exp(\beta_0' Z_i)} \right)^{-1} d\bar{N}_n(u) \tag{9}$$

Note that $\bar{A}_n$ is a step function with jumps at each of the death times and converges uniformly to A_0 (see Lemma 2 of the Appendix).

THEOREM 2 *Assume* $Pr[Y(t) = 1]$ *is continuous in* t *and* $\inf_{t\in[0,1]} E[Y(t)|Z] > 0$. *Then*

$$\sup_{t\in[0,1]} |\hat{A}_n(t) - A_0(t)| \to 0 \textit{ a.s. and } \|\hat{\beta}_n - \beta_0\|_2 \to 0 \textit{ a.s.}$$

Proof: To begin, fix $\omega \in \Phi$. In Lemma 1 of the Appendix, we show that $\{\hat{A}_n\}$ does not diverge. We know that every bounded sequence in R^k has a convergent subsequence. Thus, there exists a β and a sequence $\{\hat{\beta}_{n_k}\}$ such that $\hat{\beta}_{n_k} \to \beta$. By Helly's theorem (Ash, 1972), we know that there exists a function A and a subsequence $\{\hat{A}_{m_k}\}$ of $\{\hat{A}_{n_k}\}$ such that $\hat{A}_{m_k}(t) \to A(t)$ for all $t \in [0,1]$ at which A is continuous. Since every subsequence of a convergent subsequence in R^k must converge to the same limit, we know that $\{(\hat{\beta}_{m_k}, \hat{A}_{m_k})\}$ must converge to (β, A). We demonstrate in Lemma 2 of the Appendix that A is continuous at the continuity points of A_0. Now, we know that $L_{m_k}(\hat{\beta}_{m_k}, \hat{A}_{m_k}) - L_{m_k}(\beta_0, \bar{A}_{m_k}) \geq 0$ for all finite m_k. Furthermore, we know that

$$\begin{aligned} 0 &\leq L_{m_k}(\hat{\beta}_{m_k}, \hat{A}_{m_k}) - L_{m_k}(\beta_0, \bar{A}_{m_k}) \\ &= \frac{1}{m_k} \sum_{i=1}^{m_k} \int_0^1 \log(\chi_{m_k,i}(u)) \left\{ dN_i(u) - \frac{Y_i(u)\exp(\beta_0' Z_i)}{1+\rho\bar{A}_{m_k}(u)\exp(\beta_0' Z_i)} d\bar{A}_{m_k}(u) \right\} \\ &\quad + \frac{1}{m_k} \sum_{i=1}^{m_k} \int_0^1 \{\log(\chi_{m_k,i}(u)) - \{\chi_{m_k,i}(u) - 1\}\} \frac{Y_i(u)\exp(\beta_0' Z_i)}{1+\rho\bar{A}_{m_k}(u)\exp(\beta_0' Z_i)} d\bar{A}_{m_k}(u) \end{aligned}$$

where

$$\chi_{m_k,i}(u) = \frac{\frac{Y_i(u)\Delta\hat{A}_{m_k}(u)\exp(\hat{\beta}_{m_k}' Z_i)}{1+\rho\hat{A}_{m_k}(u)\exp(\hat{\beta}_{m_k}' Z_i)}}{\frac{Y_i(u)\Delta\bar{A}_{m_k}(u)\exp(\beta_0' Z_i)}{1+\rho\bar{A}_{m_k}(u)\exp(\beta_0' Z_i)}}$$

First, we note that the second term is less than or equal to zero since for $x > 0$, $\log(x) - (x-1) \leq 0$. Using the results and techniques of Lemma 2 of the Appendix, the first term can be shown to converge to zero and the second term converges to

$$E\left[\int_0^1 \left\{\log\left(\frac{\frac{Y(u)\exp(\beta' Z)}{1+\rho A(u)\exp(\beta' Z)}}{\frac{Y(u)\exp(\beta_0' Z)}{1+\rho A_0(u)\exp(\beta_0' Z)}}\gamma(u)\right)\right\} - \left\{\left\{\frac{\frac{Y(u)\exp(\beta' Z)}{1+\rho A(u)\exp(\beta' Z)}}{\frac{Y(u)\exp(\beta_0' Z)}{1+\rho A_0(u)\exp(\beta_0' Z)}}\gamma(u)\right\} - 1\right\}\right\} \frac{Y(u)\exp(\beta_0' Z)}{1+\rho A_0(u)\exp(\beta_0' Z)} dA_0(u)\right] \tag{10}$$

where

$$\gamma(u) = \frac{E[\frac{(1+\rho\Delta)Y(u)\exp(\beta_0' Z)}{1+\rho A_0(X)\exp(\beta_0' Z)}]}{E[\frac{(1+\rho\Delta)Y(u)\exp(\beta' Z)}{1+\rho A(X)\exp(\beta' Z)}]} \tag{11}$$

Note that (10) is equal to minus the Kullback-Leibler information, $E[\ell(\beta, A)] - E[\ell(\beta_0, A_0)]$. Because of the above inequality, we know that the Kullback-Leibler information must equal zero. By Jensen's inequality, we know that $E[\ell(\beta, A)]$ is uniquely maximized at (β_0, A_0). Furthermore, we can show that $E[\ell(\beta, A)]$ is strictly concave in β and r ($r(u) = \log(a(u))$) up to a set of measure zero. Since there is a one-to-one relationship between (β, r) and (β, A), we know that $E[\ell(\beta, A)] - E[\ell(\beta_0, A_0)] = 0$ if and only if $(\beta, A) = (\beta_0, A_0)$ almost everywhere. Therefore, $(\hat{\beta}_{m_k}, \hat{A}_{m_k})$ converges to (β_0, A_0). By Helly's theorem, we know that $(\hat{\beta}_n, \hat{A}_n)$ must also converge to (β_0, A_0) This proof can be conducted for all $\omega \in \Phi$. Therefore $(\hat{\beta}_n, \hat{A}_n)$ converges to (β_0, A_0) a.s. This result can be strengthened to uniform convergence by the Glivenko-Cantelli theorem. ∎

5.2. *Asymptotic Normality*

To establish the asymptotic distribution of our estimators $(\hat{\beta}_n, \hat{A}_n)$, we follow the function analytic approach of Murphy (1995). Instead of calculating score equations as the derivative of L_n with respect to β and the jump sizes of A, we work with one-dimensional submodels through the estimators and differentiate at the estimators. That is, we set $A_d(t) = \int_0^t (1 + dh_1(u)) d\hat{A}_n(u)$ and $\beta_d = dh_2 + \hat{\beta}_n$, where h_1 is a function and h_2 is a q-dimensional vector. Then, then we differentiate with respect to d and evaluate at $d = 0$ to get $S_n(\hat{\beta}_n, \hat{A}_n)(h_1, h_2)$. If $(\hat{\beta}_n, \hat{A}_n)$ maximizes L_n, then $S_n(\hat{\beta}_n, \hat{A}_n)(h_1, h_2) = 0$ for all (h_1, h_2). The form of S_n is given by $S_n = S_{n1} + S_{n2}$, where

$$\begin{aligned} S_{n1}(\beta, A)(h_1) &= \frac{1}{n}\sum_{i=1}^n \int_0^1 h_1(u) dN_i(u) \\ &\quad - \frac{1}{n}\sum_{i=1}^n \frac{(1+\rho\Delta_i)\exp(\beta' Z_i)\int_0^1 Y_i(u) h_1(u) dA(u)}{1+\rho A(X_i)\exp(\beta' Z_i)} \end{aligned}$$

and

$$S_{n2}(\beta, A)(h_2) = \frac{1}{n}\sum_{i=1}^{n}\int_0^1 h_2' Z_i dN_i(u) - \frac{1}{n}\sum_{i=1}^{n}\frac{(1+\rho\Delta_i)h_2' Z_i A(X_i)\exp(\beta' Z_i)}{1+\rho A(X_i)\exp(\beta' Z_i)}$$

Let $BV[0,1]$ denote the space of bounded variation functions defined on $[0,1]$. We assume that the class of h to be the space $H = BV[0,1] x R^q$. With $h \in H$, we define the norm on H to be $\|h\|_H = \|h_1\|_v + |h_2|_1$, where $\|h_1\|_v$ is the absolute value of $h_1(0)$ plus the total variation of h_1 on the interval $[0,1]$ and $|h_2|_1$ is the L_1-norm of h_2. Define $H_p = \{h \in H : \|h\|_H = \|h_1\|_v + |h_2|_1 \leq p\}$. If $p = \infty$, then the inequality is strict. Define $\langle\beta, A\rangle(h) = \int_0^1 h_1(u)dA(u) + h_2'\beta$. Then, (β, A) indexes the space of functionals $\Psi = \{\langle\beta, A\rangle : \sup_{h\in H_p}|\langle\beta, A\rangle(h)| < \infty\}$. Now $\Psi \subset \ell^\infty(H_p)$, where $\ell^\infty(H_p)$ is the space of bounded real-valued functions on H_p under the supremum norm $\|U\| = \sup_{h\in H_p}|U(h)|$. The score function S_n is a random map from Ψ to $\ell^\infty(H_p)$ for all finite p. Convergence in probability (denoted by $\mathcal{P}^*$) and weak convergence will be in terms of outer measure. Outer measure allows us to deal with random quantities which may not be measurable. A random variable is $o_{\mathcal{P}^*}(\cdot)$ if it is bounded by a measurable function which is $o_{\mathcal{P}}(\cdot)$. A similar definition holds for $O_{\mathcal{P}^*}(\cdot)$.

THEOREM 3 *Assume $Pr[Y(t) = 1]$ is continuous in t and $\inf_{t\in(0,1]} E[Y(t)|Z] > 0$. Then*

$$\langle\sqrt{n}(\hat\beta_n - \beta_0), \sqrt{n}(\hat A_n - A_o)\rangle \Longrightarrow G$$

in $\ell^\infty(H_p)$; G is a tight Gaussian process in $\ell^\infty(H_p)$ with mean zero and covariance process

$$Cov(G(h), G(h^*)) = \int_0^1 h_1(u)\sigma_{(1)}^{-1}(h^*)(u)dA_0(u) + h_2'\sigma_{(2)}^{-1}(h^*)$$

where $\sigma = (\sigma_1, \sigma_2)$ is a continuous linear operator from H_∞ to H_∞, with inverse $\sigma^{-1} = (\sigma_{(1)}^{-1}, \sigma_{(2)}^{-1})$. The form of σ is as follows:

$$\begin{aligned}
\sigma_1(h)(u) &= E\left[\frac{(1+\rho\Delta)Y(u)\exp(\beta_0' Z)Z'}{(1+\rho A_0(X)\exp(\beta_0' Z))^2}\right]h_2 \\
&\quad + h_1(u)E\left[\frac{(1+\rho\Delta)Y(u)\exp(\beta_0' Z)}{1+\rho A_0(X)\exp(\beta_0' Z)}\right] \\
&\quad - E\left[\frac{(1+\rho\Delta)Y(u)\exp(\beta_0' Z)\int_0^X h_1(v)\exp(\beta_0' Z)dA_0(v)}{(1+\rho A_0(X)\exp(\beta_0' Z))^2}\right]
\end{aligned}$$

and

$$\begin{aligned}
\sigma_2(h) &= E\left[\frac{(1+\rho\Delta)ZZ'A_0(X)\exp(\beta_0' Z)}{(1+\rho A_0(X)\exp(\beta_0' Z))^2}\right]h_2 \\
&\quad + E\left[\frac{(1+\rho\Delta)Z\exp(\beta_0' Z)\int_0^X h_1(v)dA_0(v)}{(1+\rho A_0(X)\exp(\beta_0' Z))^2}\right]
\end{aligned}$$

The proof of Theorem 3 is guided by the following theorem from van der Vaart (1995). In this theorem, the parameter space Ψ is a subset of $\ell^\infty(H_p)$ and the score function is a random map $S_n\colon \Psi \to \ell^\infty(H_p)$. The true parameter value is ψ_0 and a maximum likelihood estimator is $\hat{\psi}_n$. The asymptotic version of S_n is S. We have $S_n(\hat{\psi}_n) = 0$, $S(\psi_0) = 0$, and $\hat{\psi}_n - \psi_0 = o_{\mathcal{P}^*}(1)$ as elements in $\ell^\infty(H_p)$. The notation "*lin*" before a set denotes the set of all finite linear combinations of the elements of the set.

THEOREM 4 *Assume the following:*

1. *(Asymptotic distribution of the score function)* $\sqrt{n}(S_n(\psi_0) - S(\psi_0)) \Longrightarrow W$, *where* W *is a tight Gaussian process on* $\ell^\infty(H_p)$*;*
2. *(Fréchet differentiability of the asymptotic score)* $\sqrt{n}(S(\hat{\psi}_n) - S(\psi_0)) = -\sqrt{n}\dot{S}(\psi_0)(\hat{\psi}_n - \psi_0) + o_{\mathcal{P}^*}(1 + \sqrt{n}\|\hat{\psi}_n - \psi_0\|)$, *where* $\dot{S}(\psi_0) : lin\{\psi - \psi_0 : \psi \in \Psi\} \to \ell^\infty(H_p)$ *is a continuous linear operator;*
3. *(Invertibility)* $\dot{S}(\psi_0)$ *is continuously invertible on its range;*
4. *(Approximation condition)* $\|\sqrt{n}((S_n - S)(\hat{\psi}_n) - (S_n - S)(\psi_0))\| = o_{\mathcal{P}^*}(1 + \sqrt{n}\|\hat{\psi}_n - \psi_0\|)$

Then, $\sqrt{n}(\hat{\psi}_n - \psi_0) \Longrightarrow \dot{S}(\psi_0)^{-1}W$.

Proof: **(Theorem 3)** To prove Theorem 3, set $\hat{\psi}_n = \langle \hat{\beta}_n, \hat{A}_n \rangle$, $\psi_0 = \langle \beta_0, A_0 \rangle$, and let $S(\psi) = S(\beta, A)$, where $S(\beta, A)$ is the expectation of $S_n(\beta, A)$. That is $S = S_1 + S_2$, where

$$S_1(\beta, A)(h_1) = E\left[\int_0^1 h_1(u)dN(u)\right] - E\left[\frac{(1+\rho\Delta)\exp(\beta' Z)\int_0^X h_1(u)dA(u)}{1+\rho A(X)\exp(\beta' Z)}\right]$$

and

$$S_2(\beta, A)(h_2) = E\left[\int_0^1 h_2' Z dN(u)\right] - E\left[\frac{(1+\rho\Delta)A(X)\exp(\beta' Z)h_2' Z}{1+\rho A(X)\exp(\beta' Z)}\right]$$

For $\psi - \psi_0 \in \ell^\infty(H_p)$, it is useful to place bounds on $\|\psi - \psi_0\|$. In particular, we can show that

$$p\|\beta - \beta_0\|_1 \vee p\|A - A_0\|_\infty \le \|\psi - \psi_0\| \le p\|\beta - \beta_0\|_1 \vee 2p\|A - A_0\|_\infty$$

In the proof that follows, we will often use these bounds in place of $\|\psi - \psi_0\|$.

With the assumptions of Theorem 3, let's validate each of the conditions of Theorem 4. First we want to establish condition 1 for all finite p. We show that the class of score functions $\Psi^* \equiv \{\Psi^*_{(A_0,\beta_0)}h : h \in H_p\}$ is Donsker, where the score operator Ψ^* is given by $\Psi^*_{(A,\beta)}h = \dot{l}_a h_1 + h_2' \frac{\partial \ell}{\partial \beta}$, with $\frac{\partial \ell}{\partial \beta}$ defined in (3) and the operator $\dot{l}_a$ defined by (4) with $\dot{l}_a h_1 = \dot{l}_a \frac{a_\eta}{a} = \frac{\partial \ell}{\partial \eta}$. Boundedness of Z implies that S_β is a uniformly bounded function, which implies that $\{h_2' S_\beta : h_2 \in R^q, |h_2|_1 \le p\}$ is Donsker (see Example 2.10.10 of van der Vaart and Wellner, 1996). Since the sum of bounded Donsker classes is Donsker, the class $\{\dot{l}_{a_0} h_1 : h_1 \in BV[0,1], \|h_1\|_v \le p\}$ is Donsker if the following two classes

$$\mathcal{F}_1 = \left\{\Delta h_1 : h_1 \in BV_{[0,1]}, \|h_1\|_v \le p\right\}, \tag{12}$$

$$\mathcal{F}_2 = \left\{ f_{h_1}(X, Z, \Delta) = \frac{(1+\Delta\rho)\int_0^X h_1 dA_0 \exp(\beta_0' Z)}{1+\rho A_0(X)\exp(\beta_0' Z)} : h_1 \in BV_{[0,1]}, \|h_1\|_v \le p \right\} \tag{13}$$

are Donsker with $\sup\{|E[f]| : f \in \mathcal{F}_i\} < \infty$, $i = 1, 2$. The class $\mathcal{F}_1$ is uniformly bounded, and is Donsker since h_1 varies over bounded variation functions (see Example 2.5.4 of van der Vaart and Wellner, 1996). The class $\mathcal{F}_2$ equals a uniformly bounded function times the class $\{f_{h_1}(X) = \int_0^X h_1 dA_0 : h_1 \in BV[0,1], \|h_1\|_v \le p\}$, and this latter class is Donsker because A_0 is a monotone function (see Example 2.10.27, van der Vaart and Wellner, 1996). Also $\sup\{|E[f]| : f \in \mathcal{F}_2\} < \infty$ because h_1 varies over a Donsker class. Thus, we conclude that Ψ^* is Donsker, so that the first condition holds.

To establish condition 2, it suffices to show that $\|S(\psi) - S(\psi_0) - \dot{S}(\psi_0)(\psi - \psi_0)\|$ is $o(\|\psi - \psi_0\|)$ as $\|\psi - \psi_0\| \to 0$. To do this, we write $S(\beta, A)$ linearly in $d(A - A_0)$ and $\beta - \beta_0$ plus error terms. Specifically, note that

$$\begin{aligned} S_1(\beta, A)(h) = {} & E\left[\frac{(1+\rho\Delta)\exp(\beta_0' Z)\int_0^X h_1(u)\exp(\beta_0' Z)dA_0(u)\rho\int_0^X d(A - A_0)}{(1+\rho A_0(X)\exp(\beta_0' Z))^2}\right] \\ & - E\left[\frac{(1+\rho\Delta)\exp(\beta_0' Z)\int_0^X h_1(u)d(A - A_0)}{1+\rho A_0(X)\exp(\beta_0' Z)}\right] \\ & - (\beta - \beta_0)' E\left[\frac{(1+\rho\Delta)Z\exp(\beta_0' Z)\int_0^X h_1(u)dA_0}{(1+\rho A_0(X)\exp(\beta_0' Z))^2}\right] \\ & + error_1(\beta, A)(h) \end{aligned}$$

and

$$\begin{aligned} S_2(\beta, A)(h) = {} & -E\left[\frac{(1+\rho\Delta)h_2' Z\exp(\beta_0' Z)\int_0^X d(A - A_0)}{(1+\rho A_0(X)\exp(\beta_0' Z))^2}\right] \\ & - (\beta - \beta_0)' E\left[\frac{(1+\rho\Delta)Zh_2' Z A_0(X)\exp(\beta_0' Z)}{(1+\rho A_0(X)\exp(\beta_0' Z))^2}\right] \\ & + error_2(\beta, A)(h) \end{aligned}$$

The error terms can be very easily shown to satisfy

$$\sup_{h \in H_p} \frac{|error_i(\beta, A)(h)|}{\|\beta - \beta_0\|_1 \vee \|A - A_0\|_\infty} \to 0$$

as $\|\beta - \beta_0\|_1 \vee \|A - A_0\|_\infty \to 0$. This follows from the boundedness of N, Y, β, and A.

$$\frac{\|S(\psi) - S(\psi_0) - \dot{S}(\psi_0)(\psi - \psi_0)\|}{\|\psi - \psi_0\|} \le \frac{\|error_1(\beta, A)(h)\| + \|error_2(\beta, A)(h)\|}{p\|\beta - \beta_0\|_1 \vee p\|A - A_0\|_\infty}$$

As $\|\psi - \psi_0\| \to 0$, we know that $p\|\beta - \beta_0\|_1 \vee p\|A - A_0\|_\infty \to 0$. Hence, we can conclude that $\|S(\psi) - S(\psi_0) - \dot{S}(\psi_0)(\psi - \psi_0)\|$ is $o(\|\psi - \psi_0\|)$ as $\|\psi - \psi_0\| \to 0$. As a consequence, note that

$$\dot{S}(\psi_0)(\hat{\psi}_n - \psi_0)(h) = -\int_0^1 \sigma_1(h)d(\hat{A}_n - A_0) + (\hat{\beta}_n - \beta_0)'\sigma_2(h).$$

For condition 3, we need to prove that $\dot{S}(\psi_0)$ is continuously invertible. Intuitively, it is clear that the invertibility of $\dot{S}(\psi_0)$ is closely connected to the invertibility of $\sigma = (\sigma_1, \sigma_2)$. Because of the relationship between the variance of the score and σ, we refer to σ as the Fisher information. Note that σ is a linear operator from $L_2(dA_0)xR^q$ into itself. The Fisher information is defined in an almost-everywhere sense (dA_0), whereas we will need invertibility everywhere due to the discreteness of $\hat{A}_n$. For $h \in L_2(dA_0)xR^q$, we first show that the Fisher information is one-to-one. That is, we will demonstrate

$$\int_0^1 \sigma_1(h)(u)h_1(u)dA_0(u) + h_2'\sigma_2(h) = 0$$

implies that $h_2 = 0$ and $h_1 = 0$ almost everywhere (dA_0). Using the definitions for $\sigma_1(h)(u)$ and $\sigma_2(h)$, we know that

$$E\left[\left\{h_1(u) - \frac{\rho\int_0^u h_1(v)\exp(\beta_0'Z)dA_0(v) - h_2'Z}{1 + \rho A_0(u)\exp(\beta_0'Z)}\right\}^2 \frac{\exp(\beta_0'Z)}{1 + \rho A_0(u)\exp(\beta_0'Z)} E[Y(u)|Z]\right]$$

must equal zero, a.e (dA_0). Since $\frac{\exp(\beta_0'Z)}{1+\rho A_0(u)\exp(\beta_0'Z)}E[Y(u)|Z]$ is a strictly positive random variable, we know that for almost all $\omega \in \Omega$,

$$h_1(u) - \frac{\rho\int_0^u h_1(v)\exp(\beta_0'Z(\omega))dA_0(v) - h_2'Z(\omega)}{1 + \rho A_0(u)\exp(\beta_0'Z(\omega))}$$

must equal zero, a.e. (dA_0). This implies that for almost all $\omega \in \Omega$,

$$h_1(u) + \int_0^u \rho(h_1(u) - h_1(v))\exp(\beta_0'Z(\omega))dA_0(v) = -h_2'Z(\omega)$$

a.e. (dA_0). From this, we see that h_2 must equal zero. With $h_2 = 0$, it is easy to show that $h_1(u)$ must be zero a.e. (dA_0).

Now, we want to use this fact to show that σ, as a continuous linear operator from H_∞ to H_∞ has a continuous inverse. Since H_∞ is a Banach space, we know that if σ is invertible, then the inverse will be continuous (see Banach inverse theorem on page 149 of Luenberger, 1969). One way to show that σ is invertible is to demonstrate that σ is one-to-one and that it can be written as the difference between a bounded, linear operator with a bounded inverse and a compact, linear operator. This follows from Corollary 3.8 and Theorem 3.4 of Kress (1989). To show that σ is one-to-one, we set $\sigma = 0$ and show that $h_2 = 0$ and $h_1(u) = 0$

for all u. If $\sigma_1(h)(u) = 0$ for all u and $\sigma_2(h) = 0$, then we know, from the arguments above, that $h_2 = 0$ and $h_1(u) = 0$ a.e. (dA_0). Let $h = (h_1, 0)$, where $h_1(u) = 0$ a.e. (dA_0). Then,

$$\begin{aligned}\sigma_1(h)(u) \;&=\; -E\left[\frac{(1+\rho\Delta)Y(u)\exp(\beta_0'Z)\rho\int_0^X h_1(v)\exp(\beta_0'Z)dA_0(v)}{(1+\rho A_0(X)\exp(\beta_0'Z))^2}\right]\\ &\quad + E\left[\frac{(1+\rho\Delta)Y(u)\exp(\beta_0'Z)h_1(u)}{1+\rho A_0(X)\exp(\beta_0'Z)}\right] = 0\end{aligned}$$

for all u. By the definition of h_1, the first term must equal zero. The second term can be rewritten as

$$E\left[\frac{\exp(\beta_0'Z)}{1+\rho A_0(u)\exp(\beta_0'Z)}E[Y(u)|Z]\right]h_1(u) = 0$$

Since $E\frac{\exp(\beta_0'Z)}{1+\rho A_0(X)\exp(\beta_0'Z)}E[Y(u)|Z]]$ is greater than 0 for all u, we know that $h_1(u)$ must be identically equal to zero for all u.

Now, we want to show that σ can be written as the difference between a continuously invertible linear operator and a compact, linear operator. The first linear operator is

$$\begin{aligned}\Sigma(h) \;&=\; \left(h_1(u)E\left[\frac{(1+\rho\Delta)Y(u)\exp(\beta_0'Z)}{1+\rho A_0(X)\exp(\beta_0'Z)}\right],\right.\\ &\qquad \left. E\left[\frac{(1+\rho\Delta)ZZ'A_0(X)\exp(\beta_0'Z)}{(1+\rho A_0(X)\exp(\beta_0'Z))^2}\right]h_2\right)\end{aligned}$$

and the second term is $\Sigma(h) - \sigma(h)$. The inverse of Σ is

$$\begin{aligned}\Sigma^{-1}(h)(u) \;&=\; \left(h_1(u)E\left[\frac{(1+\rho\Delta)Y(u)\exp(\beta_0'Z)}{1+\rho A_0(X)\exp(\beta_0'Z)}\right]^{-1},\right.\\ &\qquad \left. E\left[\frac{(1+\rho\Delta)ZZ'A_0(X)\exp(\beta_0'Z)}{(1+\rho A_0(X)\exp(\beta_0'Z))^2}\right]^{-1}h_2\right)\end{aligned}$$

From the assumptions of this theorem, it follows that Σ^{-1} is a bounded linear operator.

To show that $\Sigma(h) - \sigma(h)$ is compact, we let $\{h_n\}$ be a sequence in H_p. By the definition of a compact operator (see Kress, 1989), we must prove that there exists a convergent subsequence of $\Sigma(h_n) - \sigma(h_n)$. Since h_{1n} is of bounded variation, we can write h_{1n} as the difference of increasing functions (see Lemma 2.3.3 of Ash, 1972). Both of these increasing functions are bounded in absolute value by $2p$. This means that Helly's theorem can be used to find a pointwise convergent subsequence. Let $\{h_{n_k}\}$ denote the convergent subsequence and let h^* denote its limit. We must prove that $\Sigma(h_{n_k}) - \sigma(h_{n_k})$ converges to

$\Sigma(h^*) - \sigma(h^*)$ in $\|h\|_H$ norm. To establish this result, note that $\Sigma(h) - \sigma(h)$ is equal to

$$\Bigg(-E\left[\frac{(1+\rho\Delta)Y(u)\exp(\beta_0'Z)Z'}{(1+\rho A_0(X)\exp(\beta_0'Z))^2}\right]h_2 \\ +\ E\left[\frac{(1+\rho\Delta)Y(u)\exp(\beta_0'Z)\rho\int_0^X h_1(v)\exp(\beta_0'Z)dA_0(v)}{(1+\rho A_0(X)\exp(\beta_0'Z))^2}\right], \\ -\ E\left[\frac{(1+\rho\Delta)Z\exp(\beta_0'Z)\int_0^X h_1(v)dA_0(v)}{(1+\rho A_0(X)\exp(\beta_0'Z))^2}\right]\Bigg)$$

Now, $\|\Sigma(h_{n_k}) - \sigma(h_{n_k}) - \Sigma(h^*) + \sigma(h^*)\|_H$ is less than or equal to

$$\Bigg\|-E\left[\frac{(1+\rho\Delta)Y(u)\exp(\beta_0'Z)Z'}{(1+\rho A_0(X)\exp(\beta_0'Z))^2}\right](h_{2n_k} - h_2^*) \\ + E\left[\frac{(1+\rho\Delta)Y(u)\exp(\beta_0'Z)\rho\int_0^X(h_{1n_k}(v) - h_1^*(v))\exp(\beta_0'Z)dA_0(v)}{(1+\rho A_0(X)\exp(\beta_0'Z))^2}\right]\Bigg\|_v \\ + \left\|-E\left[\frac{(1+\rho\Delta)Z\exp(\beta_0'Z)\int_0^X(h_{1n_k}(v) - h_1^*(v))dA_0(v)}{(1+\rho A_0(X)\exp(\beta_0'Z))^2}\right]\right\|_\infty$$

This quantity is less than

$$(1+\rho)c\exp(qbc)\|h_{2n} - h_2^*\|_1 \\ + (1+\rho)\exp(qbc)(\rho\exp(qbc) + c)\int_0^1 |h_{1n}(v) - h_1^*(v)|dA_0(v)$$

Using the dominated convergence theorem, we can show that this sum converges to zero. Therefore, we conclude that $\|\Sigma(h_n) - \sigma(h_n) - \Sigma(h^*) + \sigma(h^*)\|_H$ converges to zero and that $\Sigma(h) - \sigma(h)$ is a compact operator from H_p onto its range for all finite p.

Using this result, we want to show that for all finite p, $\dot{S}(\psi_0)$ is continuously invertible on its range. It is sufficient to demonstrate that for all finite p, there exists a $M_p > 0$ such that

$$\inf_{\phi\in lin\{\psi-\psi_0:\psi\in\Psi\}} \frac{\sup_{h\in H_p}|\int_0^1\sigma_1(h)d\phi_1 + \sigma_2(h)\phi_2|}{\|\phi\|} \tag{14}$$

greater than or equal to M_p (see Theorem 2 on page 153 of Kantorovich and Akilov, 1982). Because σ is a continuously invertible from H_∞ to H_∞, we know that there exists $r_p > 0$ such that $\sigma^{-1}(H_{r_p}) \subset H_p$. Therefore, we know that (14) is greater than or equal

$$\inf_{\phi\in lin\{\psi-\psi_0:\psi\in\Psi\}} \frac{\sup_{h\in\sigma^{-1}(H_{r_p})}|\int_0^1\sigma_1(h)d\phi_1 + \sigma_2(h)\phi_2|}{\|\phi\|}$$

which is equivalent to

$$\inf_{\phi \in lin\{\psi - \psi_0 : \psi \in \Psi\}} \frac{\sup_{h \in H_{r_p}} |\int_0^1 h_1 d\phi_1 + h_2 \phi_2|}{\|\phi\|} \tag{15}$$

Now, this quantity must be greater than or equal to $M_p = \frac{r_p}{2p}$. To see this, consider an arbitrary element $\phi \in \{\psi - \psi_0 : \psi \in \Psi\}$. That is, $\phi = \sum_{i=1}^k c_i(\psi_i - \psi_0)$ for some k and some $(c_1, \dots, c_k)$. Therefore, $\sup_{h \in H_{r_p}} |\int_0^1 h_1 d\phi + h_2 \phi_2|$ is greater than or equal to $r_p \| \sum_{i=1}^k c_i(\beta_i - \beta_0)\|_1 \vee r_p \| \sum_{i=1}^k c_i(A_i - A_0)\|_\infty$, and $\|\phi\|$ is less than or equal to $p\| \sum_{i=1}^k c_i(\beta_i - \beta_0)\|_1 \vee 2p \| \sum_{i=1}^k c_i(A_i - A_0)\|_\infty$. Thus, for all $\phi \in \{\psi - \psi_0 : \psi \in \Psi\}$, we know that the ratio in (15) is greater than M_p, which implies that (14) must be greater than M_p. Hence, we have shown that $\dot{S}(\psi_0)$ is continuously invertible on its range.

To verify the approximation condition of Theorem 4, it suffices to show (by Lemma 1 of van der Vaart, 1995) that

$$\left\{\Psi^*_{(A,\beta)} h - \Psi^*_{(A_0,\beta_0)} h : h \in H_p, \|\psi - \psi_0\| < \delta\right\} \tag{16}$$

is Donsker for some $\delta > 0$ and

$$\sup_{h \in H_p} E[(\Psi^*_{(A,\beta)} h - \Psi^*_{(A_0,\beta_0)} h)^2] \to 0 \tag{17}$$

as $(A, \beta) \to (A_0, \beta_0)$.

By breaking the class of (16) into two classes, we proceed by showing that the classes $\mathcal{F}_3$ and $\mathcal{F}_4$ defined below are Donsker with uniformly bounded envelopes:

$$\begin{aligned} \mathcal{F}_3 = \Bigg\{ & h_2' Z \left[\frac{\Delta + A(X)\exp(\beta' Z)}{1 + \rho A(X) \exp(\beta' Z)} \right. \\ & \left. - \frac{\Delta + A_0(X)\exp(\beta_0' Z)}{1 + \rho A_0(X)\exp(\beta_0' Z)} \right] : h_2 \in R^q, |h_2|_1 \le p, \\ & \beta \in [-\delta, \delta]^q, A \text{ nonnegative, increasing with } A(1) \le 2A_0(1) \Bigg\} \end{aligned} \tag{18}$$

$$\begin{aligned} \mathcal{F}_4 = \Bigg\{ & (1 + \Delta\rho) \left[\frac{\int_0^X h_1 dA \exp(\beta' Z)}{1 + \rho A(X)\exp(\beta' Z)} \right. \\ & \left. - \frac{\int_0^X h_1 dA_0 \exp(\beta_0' Z)}{1 + \rho A_0(X) \exp(\beta_0' Z)} \right] : h_1 \in BV[0, 1], \\ & \|h_1\|_v \le p, \beta \in [-\delta, \delta]^q, A \text{ nonnegative, increasing with } A(1) \le 2A_0(1) \Bigg\}. \end{aligned} \tag{19}$$

We use two results from empirical process theory, that classes of smooth functions are Donsker (Theorem 2.7.1, van der Vaart and Wellner, 1996) and classes of Lipschitz transformations of Donsker classes with integrable envelope functions are Donsker (Theorem 2.10.6, van der Vaart and Wellner, 1996). Differentiability of $g_\beta(Z) = \exp(\beta' Z)$

in Z and boundedness of the derivative (because $|Z| < \infty$) implies $\{g_\beta(Z) : \beta \in [-\delta, \delta]^q\}$ is Donsker for any finite δ. Since the class of nonnegative increasing functions A on $[0, 1]$ with $A(1) \leq 2A_0(1)$ is Donsker by monotonicity, and is uniformly bounded, it follows that the denominator $\{1+\rho A(X)\exp(\beta' Z)\}$ with (β, A) varying over the index class is Donsker for any finite δ. Since this denominator is bounded away from zero, another application of Theorem 2.10.6 yields that $\mathcal{F}_3$ is Donsker.

Further, Theorem 2.10.6 implies the class BV_M of all real-valued functions on $[0, 1]$ that are uniformly bounded by a constant M and are of variation bounded by M is Donsker. Since $\| \int_0^X h_1 dA\|_v \leq \|h_1\|_\infty * \|A\|_v < \infty$, the class $\{\int_0^X h_1 dA : h_1 \in BV[0, 1], \|h_1\|_v \leq p\}$ with A varying over the set in (19) is Donsker. Then $\mathcal{F}_4$ is Donsker since it is comprised of ratios of bounded Donsker classes. Boundedness of A_0 and Z give $\sup\{E[|f|; f \in \mathcal{F}_i\} < \infty, i = 3, 4$, so that (16) holds. Condition (17) holds by the dominated convergence theorem. Thus, we have shown that the approximation condition holds.

Putting these four results together, we know that for all finite p

$$\begin{aligned} -\dot{S}(\psi_0)\sqrt{n}(\hat{\psi}_n - \psi_0)(h) &= \int_0^1 \sigma_1(h)(u) d\sqrt{n}(\hat{A}_n - A_0) + \sqrt{n}(\hat{\beta}_n - \beta_0)'\sigma_2(h) \\ &= \sqrt{n}(S_n(\hat{\psi}_n) - S(\psi_0))(h) + o_{\mathcal{P}^*}(1) \\ &= \sqrt{n}(S_n(\psi_0) - S(\psi_0))(h) + o_{\mathcal{P}^*}(1) \end{aligned}$$

uniformly in $h \in H_p$. So, $\sqrt{n}(\hat{\psi}_n - \psi_0) \Longrightarrow \dot{S}(\psi_0)^{-1}W$. But, what is $\dot{S}(\psi_0)^{-1}W$? By the continuous invertibility of σ, we know that, for finite p, there exists a finite r_p for which $\sigma^{-1}(g) \in H_{r_p}$ if $g \in H_p$. To get a weak convergence result for $\hat{\psi}_n$ at $g \in H_p$, we put $h = \sigma^{-1}(g)$ in the above equation to get

$$\int_0^1 g_1(u) d\sqrt{n}(\hat{A}_n - A_0) + \sqrt{n}(\hat{\beta}_n - \beta_0)' g_2 = \sqrt{n}(S_n(\psi_0) - S(\psi_0))(\sigma^{-1}(g)) + o_{\mathcal{P}^*}(1)$$

uniformly in $g \in H_p$. The right hand side converges to $W(\sigma^{-1}(g))$, which has mean zero and variance $\int_0^1 g_1\sigma_{(1)}^{-1}(g)dA_0 + \sigma_{(2)}^{-1}(g)'g_2$. Since this holds uniformly in $g \in H_p$, we can conclude that

$$\langle\sqrt{n}(\hat{\beta}_n - \beta_0), \sqrt{n}(\hat{A}_n - A_0)\rangle \Longrightarrow G$$

where G is a tight Gaussian process on $\ell^\infty(H_p)$ with mean zero and covariance process

$$Cov[G(g), G(g^*)] = \int_0^1 g_1\sigma_{(1)}^{-1}(g^*)dA_0 + \sigma_{(2)}^{-1}(g^*)'g_2$$

Hence, $\dot{S}(\psi_0)^{-1}W = G$. ∎

In the above framework, specific choices of g correspond to the quantities of interest. For example, if we let $g_1(u) = 0$ for all u and $g_2 = e_i$, the ith unit vector, then we get the ith element of $\hat{\beta}_n$. If we let $g_1(u) = 1(u \leq t)$ and $g_2 = 0$, then we get our estimator for $A_0(t)$. The above theorem shows that these quantities are asymptotically normal. Theorem 2 verifies the consistency of these estimators.

To estimate the asymptotic variance of $\langle\sqrt{n}(\hat\beta_n - \beta_0), \sqrt{n}(\hat A_n - A_0)\rangle(g)$, we use the following procedure. First, estimate σ with $\hat\sigma$. Next, solve for h such that $g_1 = \hat\sigma_1(h)$ and $g_2 = \hat\sigma_2(h)$. Finally, estimate the asymptotic variance as $\int_0^1 g_1 h_1 d\hat A_n + h_2' g_2$. The natural estimator for $\sigma(h)$ is

$$
\begin{aligned}
\hat\sigma_1(h)(u) &= \frac{1}{n}\sum_{i=1}^n \frac{(1+\rho\Delta_i)Y_i(u)\exp(\hat\beta_n' Z_i)Z_i'}{(1+\rho\hat A_n(X_i)\exp(\hat\beta_n' Z_i))^2} h_2 \\
&\quad - \frac{1}{n}\sum_{i=1}^n \frac{(1+\rho\Delta_i)Y_i(u)\exp(\hat\beta_n' Z_i)\rho\int_0^{X_i} h_1(v)\exp(\hat\beta_n' Z_i)d\hat A_n(v)}{(1+\rho\hat A_n(X_i)\exp(\hat\beta_n' Z_i))^2} \\
&\quad + h_1(u)\frac{1}{n}\sum_{i=1}^n \frac{(1+\rho\Delta_i)Y_i(u)\exp(\hat\beta_n' Z_i)}{1+\rho\hat A_n(X_i)\exp(\hat\beta_n' Z_i)}
\end{aligned}
$$

and

$$
\begin{aligned}
\hat\sigma_2(h) &= \frac{1}{n}\sum_{i=1}^n \frac{(1+\rho\Delta_i)Z_i Z_i' \hat A_n(X_i)\exp(\hat\beta_n' Z_i)}{(1+\rho\hat A_n(X_i)\exp(\hat\beta_n' Z_i))^2} h_2 \\
&\quad + \frac{1}{n}\sum_{i=1}^n \frac{(1+\rho\Delta_i)Z_i\exp(\hat\beta_n' Z_i)\int_0^{X_i} h_1(v)d\hat A_n(v)}{(1+\rho\hat A_n(X_i)\exp(\hat\beta_n' Z_i))^2}
\end{aligned}
$$

Due to the continuous invertibility of σ, this estimation procedure is equivalent to the following naive approach of estimation of the asymptotic variance. First, form the second derivative matrix of L_n by taking derivatives with respect to β and the jump sizes of A. Next, invert this large matrix, multiply by negative one and form the vector $(g_1(X_1), \ldots, g_1(X_{k(n)}), g_2)$. Finally, pre and post multiply the large inverted matrix to get the estimate for the asymptotic variance. This is essentially the standard likelihood approach to the finite parameter problem.

THEOREM 5 *Assume $Pr[Y(t) = 1]$ is continuous in t and* $\inf_{t\in(0,1]} E[Y(t)|Z] > 0$. *For $g \in H_p$, the solution $h = \hat\sigma^{-1}(g)$ exists with probability going to one as n increases. In addition, $\int_0^1 g_1 h_1 d\hat A_n + h_2' g_2$ converges in probability to $\int_0^1 g_1 \sigma_{(1)}^{-1}(g)dA_0 + \sigma_{(2)}^{-1}(g)' g_2$.*

Proof: In Theorem 3 of Murphy (1995), the author outlines the method of proof. Suppose we could show

$$\sup_{\|h\|=1}\|\hat\sigma(h) - \sigma(h)\|_H \xrightarrow{P} 0 \tag{20}$$

Then, because σ is continuously invertible, $\hat\sigma$ is continuously invertible on a set of probability converging to 1. The range of $\hat\sigma$ is all of H_∞, which follows by the same proof that $Range(\sigma) = H_\infty$, as in Rudin (1973, pp 99–103). Since $\hat\sigma$ converges to σ by (20) and σ is continuously invertible on H_∞, there exists a finite q for which $\hat\sigma^{-1}(H_q) \subset H_p$ on a set of probability converging to 1. Therefore

$$
\begin{aligned}
\|\hat\sigma^{-1}(g) - \sigma^{-1}(g)\|_H &\le \sup_{h\in H_q}\|\hat\sigma^{-1}(q) - \sigma^{-1}(q)\|_H \\
&\le \sup_{h\in H_p}\|\sigma^{-1}(\sigma(h)) - \sigma^{-1}(\hat\sigma(h))\|_H
\end{aligned}
$$

$$\leq \ \sup_{h\in H_p} \frac{\|\sigma^{-1}(h)\|_H}{\|h\|_H} \sup_{h\in H_p} \|\sigma(h) - \hat{\sigma}(h)\|_H$$

Since σ^{-1} is a bounded operator, once 20 is shown, the proof is complete that $\hat{\sigma}^{-1}$ converges to σ^{-1} uniformly over H.

The most difficult part of this is showing that the variation of

$$\frac{1}{n}\sum_{i=1}^{n} \frac{(1+\rho\Delta_i)Y_i(u)\exp(\hat{\beta}_n' Z_i)\rho\int_0^{X_i} h_1(v)\exp(\hat{\beta}_n' Z_i)d\hat{A}_n(v)}{(1+\rho\hat{A}_n(X_i)\exp(\hat{\beta}_n' Z_i))^2} - E\left[\frac{(1+\rho\Delta)Y(u)\exp(\beta_0' Z)\rho\int_0^{X} h_1(v)\exp(\beta_0' Z)dA_0(v)}{(1+\rho A_0(X)\exp(\beta_0' Z))^2}\right]$$

goes to zero. We can bound this variation by the variation of

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{(1+\rho\Delta_i)\rho\int_0^{X_i} h_1(v)d\hat{A}_n(v)}{(1+\rho\hat{A}_n(X_i)\exp(\hat{\beta}_n' Z_i))^2}(\exp(\hat{\beta}_n' Z_i))^2 - \frac{(1+\rho\Delta_i)\rho\int_0^{X_i} h_1(v)dA_0(v)}{(1+\rho A_0(X_i)\exp(\beta_0' Z_i))^2} E\left[(\exp(\beta_0' Z_i))^2 | N_i, Y_i\right]\right) Y_i(u)$$

plus the variation of

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{(1+\rho\Delta_i)\rho\int_0^{X_i} h_1(v)dA_0(v)}{(1+\rho A_0(X_i)\exp(\beta_0' Z_i))^2} E\left[(\exp(\beta_0' Z_i))^2 | N_i, Y_i\right] Y_i(u) - E[\frac{(1+\rho\Delta)\rho\int_0^{X} h_1(v)dA_0(v)}{(1+\rho A_0(X)\exp(\beta_0' Z))^2}(\exp(\beta_0' Z))^2 Y(u)]\right)$$

Uniform consistency of $(\hat{\beta}_n, \hat{A}_n)$ implies the first term converges to zero. The variation of the second term is bounded above by $(1+\rho)\rho$ times

$$\int_0^1 \|h_1\|_v \left\| \frac{1}{n}\sum_{i=1}^{n} \frac{E\left[(\exp(\beta_0' Z_i))^2 | N_i, Y_i\right] Y_i(u) Y_i(\cdot)}{(1+\rho A_0(X_i)\exp(\beta_0' Z_i))^2} - E\left[\frac{(\exp(\beta_0' Z))^2 Y(u) Y(\cdot)}{(1+\rho A_0(X)\exp(\beta_0' Z))^2}\right]\right\|_v dA_0(u)$$

The integrand converges to zero by the strong law of large numbers in Banach spaces. Thus the term converges to zero by the dominated convergence theorem, using $\|h_1\|_v \leq 1$.

■

Finally, we want to bring the problem full circle and demonstrate that the asymptotic variance of $\sqrt{n}(\hat{\beta}_n - \beta_0)$ is equal to the semiparametric variance bound calculated to Section 3.2. By the Cramer-Wold device (see Serfling, 1980), it suffices to demonstrate that the asymptotic variance of $\lambda'\sqrt{n}(\hat{\beta}_n - \beta_0)$ is equal to $\lambda' V\lambda$, where λ is any vector in

R^q. Thus, we have to show that $\sigma_{(2)}^{-1}(g) = Vg_2$, where $g_1(u) = 0$ for all u and $g_2 = \lambda$. To find $\sigma_{(2)}^{-1}(g)$, we must find an h such that

$$\begin{aligned}\sigma_1(h)(u) \;=\; & E\left[\frac{(1+\rho\Delta)Y(u)\exp(\beta_0'Z)Z'}{(1+\rho A_0(X)\exp(\beta_0'Z))^2}\right]h_2 \\ & + h_1(u)E\left[\frac{(1+\rho\Delta)Y(u)\exp(\beta_0'Z)}{1+\rho A_0(X)\exp(\beta_0'Z)}\right] \\ & - E\left[\frac{(1+\rho\Delta)Y(u)\exp(\beta_0'Z)\rho\int_0^X h_1(v)\exp(\beta_0'Z)dA_0(v)}{(1+\rho A_0(X)\exp(\beta_0'Z))^2}\right] = 0 \quad (21)\end{aligned}$$

and

$$\begin{aligned}\sigma_2(h) \;=\; & E\left[\frac{(1+\rho\Delta)ZZ'A_0(X)\exp(\beta_0'Z)}{(1+\rho A_0(X)\exp(\beta_0'Z))^2}\right]h_2 \\ & + E\left[\frac{(1+\rho\Delta)Z\exp(\beta_0'Z)\int_0^X h_1(v)dA_0(v)}{(1+\rho A_0(X)\exp(\beta_0'Z))^2}\right] = \lambda \qquad (22)\end{aligned}$$

Consider the solution, $h_2 = V\lambda$ and $h_1(u) = -w_{eff}(u)'W^{-1}(UV - I)\lambda$, where

$$\begin{aligned}U \;&=\; E\left[(1+\rho\Delta)\frac{A_0(X)\exp(\beta_0'Z)ZZ'}{(1+\rho A_0(X)\exp(\beta_0'Z))^2}\right] \\ W \;&=\; E\left[(1+\rho\Delta)\frac{\int_0^X \exp(\beta_0'Z)w_{eff}(u)Z'dA_0(u)}{(1+\rho A_0(X)\exp(\beta_0'Z))^2}\right]\end{aligned}$$

Note that $V = (U - W)^{-1}$. It is clear that this solution satisfies (22). To show that it satisfies (21), we must show that

$$\begin{aligned}& E\left[\frac{(1+\rho\Delta)Y(u)\exp(\beta_0'Z)Z'}{(1+\rho A_0(X)\exp(\beta_0'Z))^2}\right]V\lambda \\ & -w_{eff}(u)'E\left[\frac{(1+\rho\Delta)Y(u)\exp(\beta_0'Z)}{1+\rho A_0(X)\exp(\beta_0'Z)}\right]W^{-1}(UV-I)\lambda \\ & +E\left[\frac{(1+\rho\Delta)Y(u)\exp(\beta_0'Z)\rho\int_0^X w_{eff}(v)'\exp(\beta_0'Z)dA_0(v)}{(1+\rho A_0(X)\exp(\beta_0'Z))^2}\right]W^{-1}(UV-I)\lambda = 0\end{aligned}$$

for all u. Since $V = W^{-1}(UV - I)$, we know that this will hold if

$$\begin{aligned}& E\left[\frac{(1+\rho\Delta)Y(u)\exp(\beta_0'Z)Z'}{(1+\rho A_0(X)\exp(\beta_0'Z))^2}\right] \\ & +E\left[\frac{(1+\rho\Delta)Y(u)\exp(\beta_0'Z)\rho\int_0^X w_{eff}(v)'\exp(\beta_0'Z)dA_0(v)}{(1+\rho A_0(X)\exp(\beta_0'Z))^2}\right] \\ & -w_{eff}(u)'E\left[\frac{(1+\rho\Delta)Y(u)\exp(\beta_0'Z)}{1+\rho A_0(X)\exp(\beta_0'Z)}\right] = 0\end{aligned}$$

for all u. This is satisfied because $w_{eff}(u)$ is the solution to the integral equation (5).

6. Simulation Study

We performed a simulation study to assess the accuracy of our estimation approach. In particular, we were interested in determining how well our method performs in small to moderate sample sizes and how it is affected by the amount of censoring and different values of ρ. Towards this end, we generated data according to a ρ-proportional odds model with two covariates. The first covariate was uniformly distributed on $[0, 1]$ and the second covariate was Bernoulli (0.5). We considered three values of ρ, 0.5, 1.0, and 2.0, and assumed that $A_0(t) = \exp\{t\} - 1$. In addition, we assumed that censoring times were uniformly distributed on the interval $[0, t_c]$, where t_c was chosen to yield censoring levels of 10%, 20%, 40%, and 60%. We considered true values for the parameters to be either 0 or 1. We simulated data in sets of 100 subjects and for each combination of simulation parameters, we performed 500 simulations. The models that we fit to the data corresponded to the true value of ρ. In Table 1, we report the means, average standard errors, and Monte Carlo standard errors for the parameter estimates from these simulations.

In general, we see that there is very little bias associated with our procedure. However, there are five situations (indicated by $*$) where there is a significant difference at the 0.05 level between the average of the parameter estimates and the true value of the parameter. In addition, we note that the variability of the parameter estimates increases with the level of censoring and the value of ρ. Finally, the normal approximation appears to work well for this sample size as evidenced by the closeness of the Monte Carlo standard errors and the average standard error estimates.

7. Veteran's Administration Lung Cancer Trial

In the Veteran's Administration lung cancer trial (Prentice, 1973; Kalbfleisch and Prentice, 1980), males with inoperative lung cancer were randomized to either a standard or experimental chemotherapy. The primary endpoint of interested was time until death, as measured in days. For this example, we will only consider the subset of 97 patients who received no prior therapy. We are interested in the relationship between the time to death and two covariates, performance status at randomization (PS) and histological type of tumor (squamous, small, adeno, or large). These data were analyzed using a proportional odds model by Bennett (1983), Pettitt (1984), Parzen and Harrington (1993), Cheng et al. (1995), and Murphy et al. (1996). All of these authors fit a model with PS and indicator functions for tumor type. Table 2 presents the results of fitting this model using our method as well as the results obtained by the other approaches. Table 3 displays the results of fitting various models within the generalized odds-rate class. Specifically, we fit models with ρ's ranging from 0.0 to 2.0. Figure 1 presents our estimates of the baseline survival function for each of the fitted models. Note that all of these models paint a similar picture. Namely, performance status at randomization is positively associated with survival for patients with all tumor types. Furthermore, there is not a significant difference between the survival rates

Table 1. Results from the Simulation Study (Asterisk indicates significant difference (0.05 level) between mean of parameter estimates and the truth).

			$\hat{\beta}_1$			$\hat{\beta}_2$		
ρ	(β_1, β_2)	Censoring	Mean	Avg. s.e.	M.C. s.e.	Mean	Avg. s.e.	M.C. s.e.
0.5	(1,1)	20%	1.0340	0.5337	0.5165	1.0177	0.3158	0.3245
		40%	1.0615*	0.5729	0.5717	0.9966	0.3382	0.3467
		60%	1.0565	0.6567	0.6849	1.0246	0.3902	0.3924
	(1,0)	20%	1.0305	0.5360	0.5372	0.0186	0.3035	0.3121
		40%	1.0390	0.5760	0.5962	0.0029	0.3272	0.3329
		60%	0.9719	0.6583	0.7081	0.0267	0.3731	0.3862
	(0,1)	20%	−0.0251	0.5271	0.5332	0.9912	0.3170	0.3267
		40%	0.0187	0.5673	0.5905	0.9914	0.3384	0.3291
		60%	0.0523	0.6463	0.6574	1.0041	0.3898	0.3989
	(0,0)	20%	0.0305	0.5276	0.5103	0.0084	0.3053	0.2966
		40%	−0.0119	0.5678	0.5972	−0.0296*	0.3278	0.3353
		60%	−0.0085	0.6496	0.6590	−0.0240	0.3730	0.3923
1.0	(1,1)	20%	0.9952	0.6362	0.6351	1.0149	0.3742	0.3612
		40%	0.9713	0.6588	0.6724	1.0023	0.3890	0.3942
		60%	1.0438	0.7309	0.7414	1.0189	0.4282	0.4445
	(1,0)	20%	0.9777	0.6309	0.6213	−0.0011	0.3625	0.3569
		40%	1.0472	0.6648	0.7075	−0.0027	0.3779	0.3868
		60%	0.9380	0.7234	0.7724	0.0122	0.4119	0.4110
	(0,1)	20%	0.0528	0.6252	0.6528	1.0344*	0.3749	0.3783
		40%	0.0253	0.6609	0.6686	0.9839	0.3895	0.4047
		60%	0.0175	0.7220	0.7311	1.0142	0.4280	0.4237
	(0,0)	20%	0.0244	0.6244	0.6275	0.0025	0.3626	0.3625
		40%	0.0038	0.6554	0.6439	0.0080	0.3800	0.3807
		60%	−0.0002	0.7189	0.7213	−0.0284	0.4149	0.4355
2.0	(1,1)	20%	0.9880	0.7960	0.7290	0.9735	0.4638	0.4622
		40%	1.0614*	0.8109	0.8119	0.9973	0.4741	0.4809
		60%	1.0824*	0.8526	0.9054	1.0257	0.5005	0.5293
	(1,0)	20%	0.9334	0.7906	0.8388	−0.0064	0.4543	0.4604
		40%	1.0358	0.8138	0.8421	0.0228	0.4653	0.4831
		60%	1.0139	0.8563	0.8268	0.0355	0.4882	0.5349
	(0,1)	20%	−0.0340	0.7859	0.8110	0.9760	0.4646	0.4533
		40%	0.0192	0.8033	0.8101	1.0512*	0.4779	0.4923
		60%	−0.0490	0.8543	0.9084	0.9803	0.4985	0.5232
	(0,0)	20%	0.0491	0.7788	0.8151	0.0043	0.4544	0.4704
		40%	0.0097	0.8045	0.8229	0.0017	0.4648	0.4546
		60%	0.0029	0.8521	0.8927	0.0016	0.4909	0.4761

Table 2. Veteran's Administration Lung Cancer Data: Comparison of model estimates for proportional odds model by method.

	PS		Squamous		Small		Adeno	
Method	Estimate	s.e.	Estimate	s.e.	Estimate	s.e.	Estimate	s.e.
Scharfstein	−0.053	0.010	−0.180	0.588	1.383	0.524	1.314	0.554
Murphy	−0.055	0.010	−0.217	0.589	1.440	0.525	1.339	0.556
Cheng	−0.055	0.010	−0.006	0.572	1.496	0.498	1.556	0.414
Parzen	−0.053	0.010	−0.173	0.620	1.380	0.482	1.31	0.453
Pettitt	−0.055	0.009	−0.177	0.593	1.440	0.520	1.300	0.554
Bennett	−0.053	0.010	−0.181	0.588	1.380	0.524	1.31	0.554

Table 3. Veteran's Administration Lung Cancer Data: Comparison of model estimates for varying ρ's.

	PS		Squamous		Small		Adeno		
ρ	Estimate	s.e.	Estimate	s.e.	Estimate	s.e.	Estimate	s.e.	Log Likelihood
0.00	−0.024	0.006	−0.214	0.347	0.548	0.321	0.851	0.348	−375.44
0.50	−0.040	0.008	−0.240	0.479	1.100	0.440	1.119	0.462	−371.61
0.85	−0.050	0.010	−0.205	0.558	1.309	0.505	1.258	0.528	−371.16
1.00	−0.053	0.010	−0.180	0.588	1.383	0.524	1.314	0.554	−371.25
1.50	−0.064	0.012	−0.073	0.675	1.605	0.596	1.497	0.636	−372.69
2.00	−0.072	0.014	0.046	0.749	1.813	0.661	1.678	0.712	−373.71

of patients with squamous and large tumors. However, patients with large tumor types tend to have significantly better survival prognosis than those with either small or adeno types.

Selection of ρ is a crucial step in model fitting. Cheng, Wei, and Ying (1997) proposed a graphical approach to selection of ρ. They noted that if the model is correct, then the distribution of $\hat{\alpha}(T) + \hat{\beta}'Z$ should be distributed according to a Pareto(ρ) for $\rho > 0$ and according to an exponential(1) distribution if $\rho = 0$. For given ρ, this distribution is best approximated by the Kaplan-Meier estimator based on the sample $\{(\hat{\alpha}(X_i) + \hat{\beta}'Z_i, \Delta_i), i = 1, \ldots, n\}$. Figure 2 presents P-P plots comparing the error distribution and the Kaplan-Meier estimator for varying values of ρ ranging from 0.0 to 2.0. If ρ is correct then the P-P plot should form a 45 degree line through the origin. In Figure 2, we presents estimates of the area between the P-P plot and the straight line. With this criteria, $\rho = 1.0$ seems to provide the best model fit. An alternative approach is to jointly maximize the profile likelihood with respect to ρ. Figure 3 presents the log profile likelihood for varying values of ρ between 0 and 5. We see that it is maximized around 0.85. The last column of Table 3 displays the actual value of the log profile likelihoods for the six values of ρ considered. So, it appears that there are values of $\rho > 0$ that fit better than the proportional hazards model.

8. Summary

In this paper, we used full nonparametric maximum likelihood to construct estimators for the regression parameters and baseline odds function in the generalized odds-rate model. Using theory developed by Murphy (1994), Murphy (1995) and van der Vaart and Wellner (1996), we showed that these estimators are consistent and asymptotically normal. Furthermore,

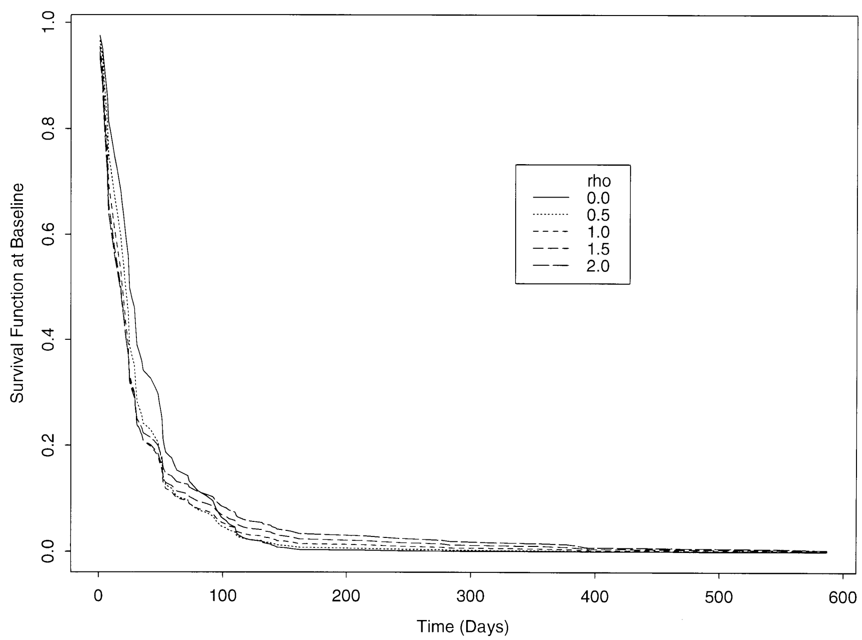

Figure 1. Baseline survival functions for varying ρ's.

the estimate of the regression parameters was shown to attain the semiparametric efficiency bound. The estimators were also shown to perform well in small to moderate samples. For future research, we intend to pursue the issue of model selection in greater depth.

Appendix

LEMMA 1 *Assume $Pr[Y(t) = 1]$ is continuous in t. Then for each $\omega \in \Phi$, $\hat{A}_n(1)$ is bounded by a finite constant for all $n \geq 1$.*

Proof: To begin, fix $\omega \in \Phi$. To prove that $\hat{A}_n(1)$ is bounded by a finite constant for all $n \geq 1$, consider a proof by contradiction. That is, suppose that $\lim_{n\to\infty} \hat{A}_n(1) = \infty$. Note that $L_n(\hat{\beta}_n, \hat{A}_n) - L_n(\beta_0, \bar{A}_n)$ must be non-negative for all n. The goal is to show that this difference will become negative as $n \to \infty$, which is a contradiction. In the following, any terms which are bounded away from positive infinity will be represented by $O(1)$. Now, we can write this difference as

$$\int_0^1 \log\left(\frac{\Delta\hat{A}_n(u)}{\Delta\bar{A}_n(u)}\right) d\bar{N}_n(u) - \frac{1}{n}\sum_{i=1}^{n}(1/\rho + \Delta_i)\log(1 + \rho\hat{A}_n(X_i)\exp(\hat{\beta}_n' Z_i))$$
$$+\frac{1}{n}\sum_{i=1}^{n}\Delta_i(\hat{\beta}_n - \beta_0)' Z_i + \frac{1}{n}\sum_{i=1}^{n}(1/\rho + \Delta_i)\log(1 + \rho\bar{A}_n(X_i)\exp(\beta_0' Z_i))$$

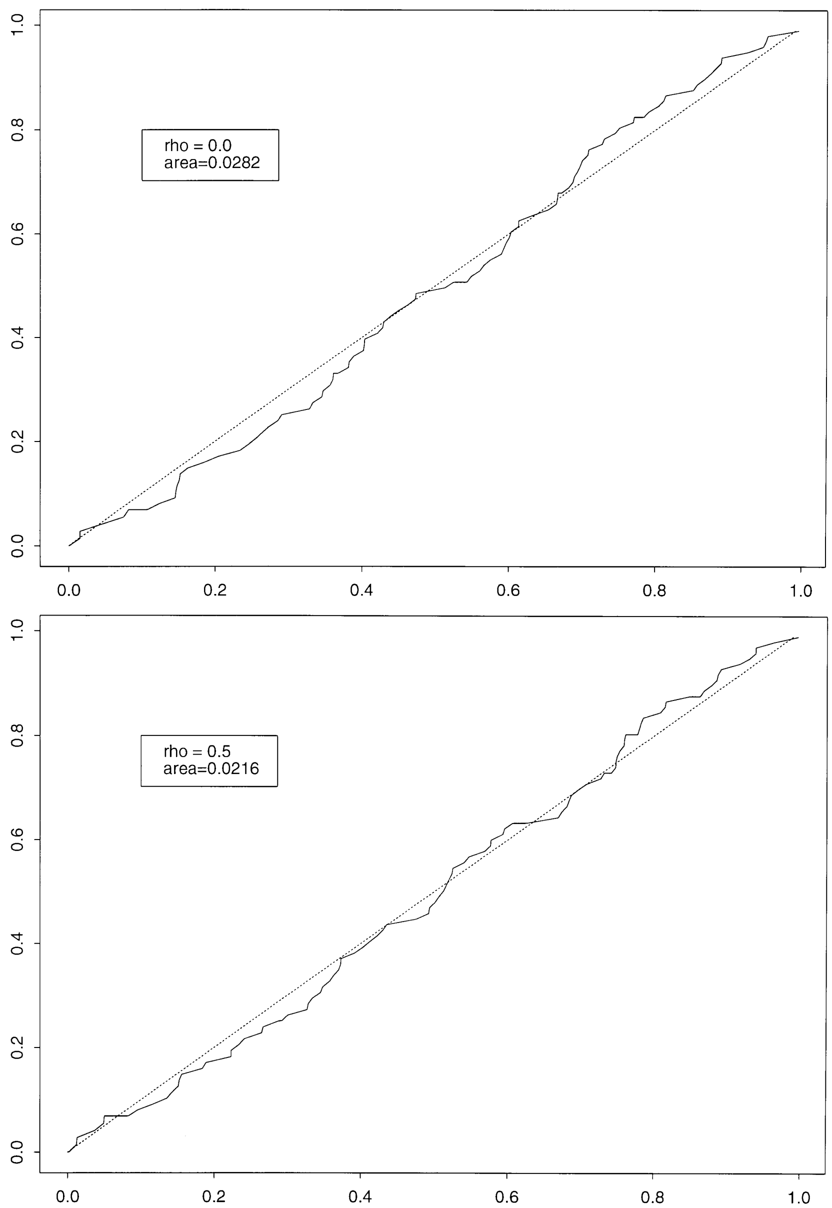

Figure 2. Goodness of fit P-P plots for various values of ρ's.

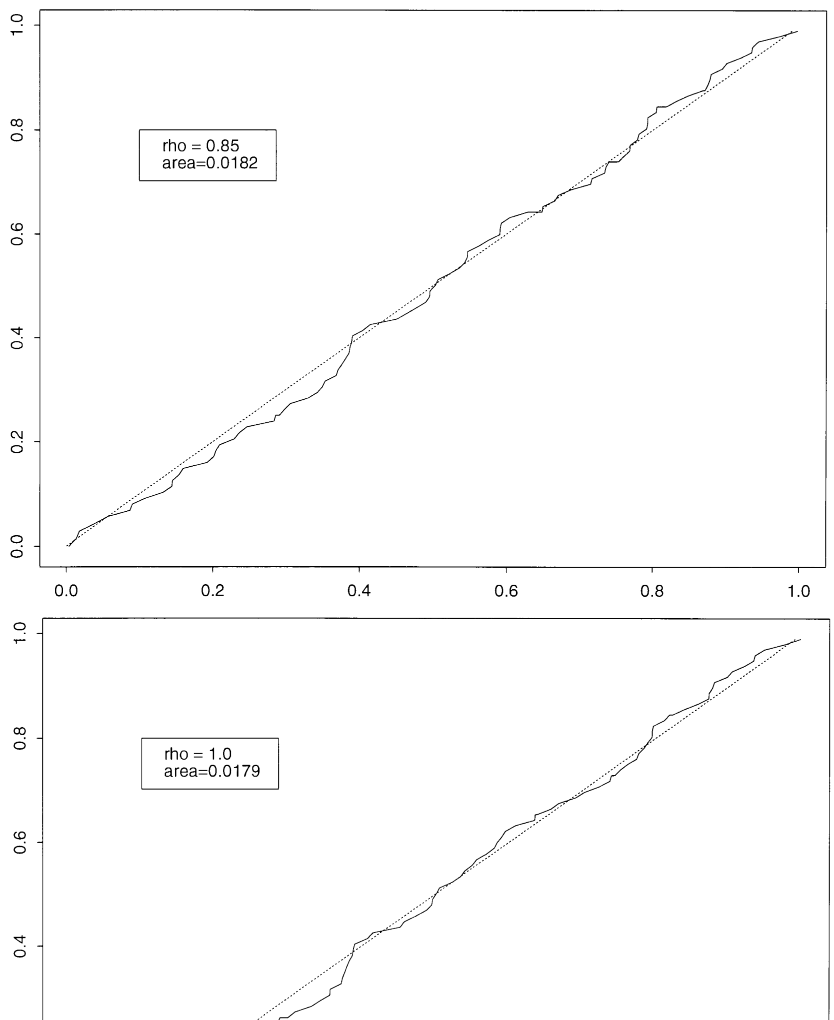

Figure 2. Continued.

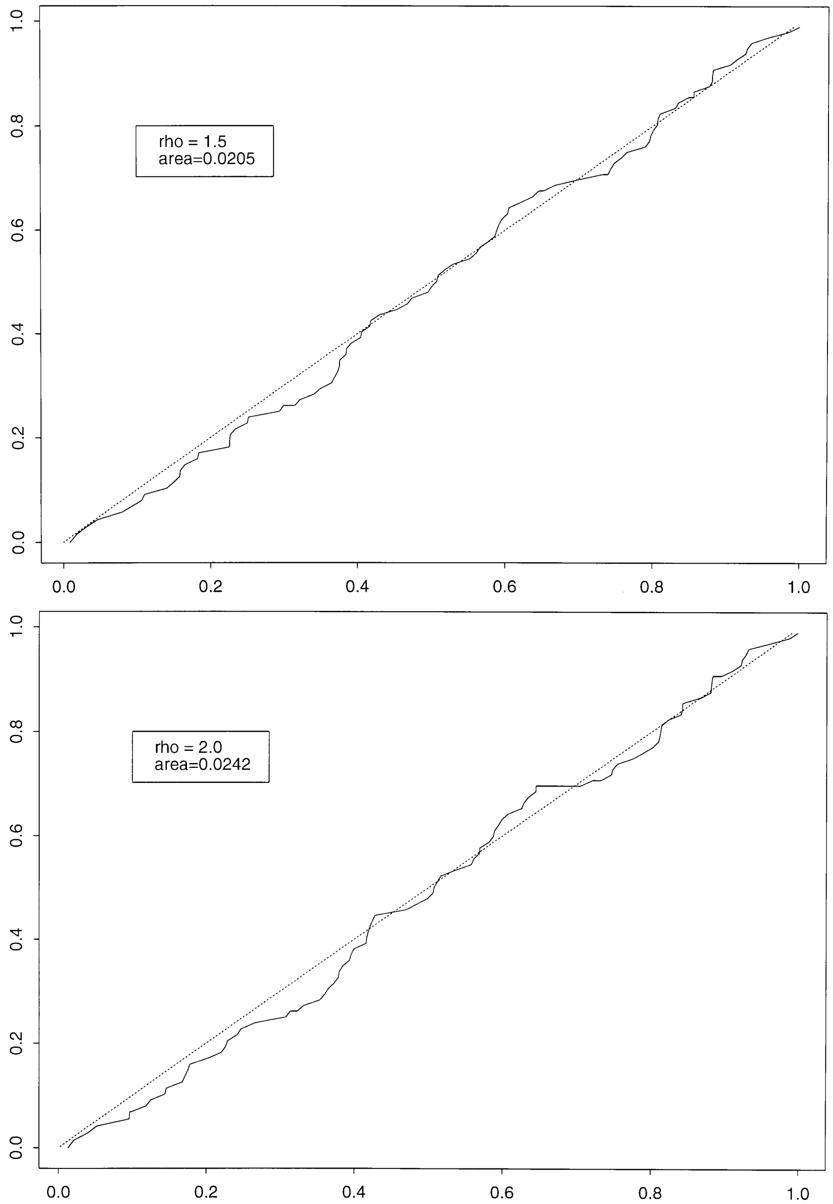

Figure 2. Continued.

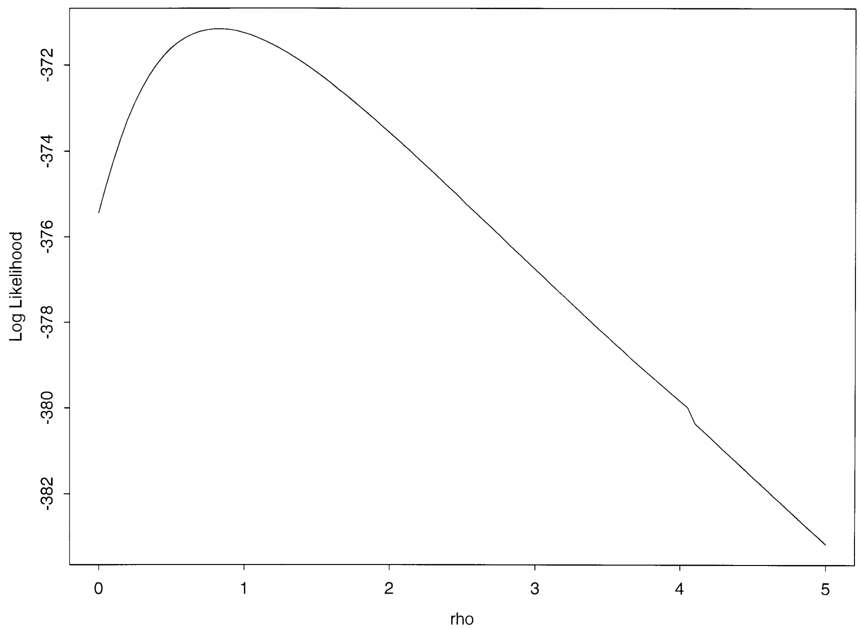

Figure 3. Log likelihood for varying ρ's.

The last two terms above are bounded away from positive infinity. Using (8) and (9), we can plug in and re-express the first term above. Thus, we know that $L_n(\hat{\beta}_n, \hat{A}_n) - L_n(\beta_0, \bar{A}_n)$ is bounded from above by

$$O(1) + \int_0^1 \log\left(\frac{1}{n}\sum_{j=1}^{n}\frac{(1+\rho\Delta_j)Y_j(u)\exp(\beta_0' Z_j)}{1+\rho A_0(X_j)\exp(\beta_0' Z_j)}\right) d\bar{N}_n(u)$$
$$- \int_0^1 \log\left(\frac{1}{n}\sum_{j=1}^{n}\frac{(1+\rho\Delta_j)Y_j(u)\exp(\hat{\beta}_n' Z_j)}{1+\rho \hat{A}_n(X_j)\exp(\hat{\beta}_n' Z_j)}\right) d\bar{N}_n(u)$$
$$-\frac{1}{n}\sum_{i=1}^{n}(1/\rho + \Delta_i)\log(1+\rho\hat{A}_n(X_i)\exp(\hat{\beta}_n' Z_i))$$

The second term is bounded away from positive infinity. So, $L_n(\hat{\beta}_n, \hat{A}_n) - L_n(\beta_0, \bar{A}_n)$ is less or equal to

$$O(1) - \int_0^1 \log\left(\frac{1}{n}\sum_{j=1}^{n}\frac{(1+\rho\Delta_j)Y_j(u)\exp(\hat{\beta}_n' Z_j)}{1+\rho \hat{A}_n(X_j)\exp(\hat{\beta}_n' Z_j)}\right) d\bar{N}_n(u)$$
$$-\frac{1}{n}\sum_{i=1}^{n}(1/\rho + \Delta_i)\log(1+\rho\hat{A}_n(X_i)\exp(\hat{\beta}_n' Z_i)) \tag{A.1}$$

Intuitively, we see that as $\hat{A}_n$ diverges to infinity, the second term diverges to positive infinity and the third term diverges to negative infinity. We wish to show that the sum of these two terms diverges to negative infinity. To show this, we partition the interval $[0, 1]$ according to a nonnegative, strictly decreasing subsequence $1 = s_0 > s_1 > \cdots \geq 0$. Letting $c^* = e^{-qbc}$, we can show that the second term in (A.1) is bounded from above by

$$-\frac{1}{n}\sum_{i=1}^{n} Y_i(s_1)\int_0^1 \log\left(\frac{1}{n}\sum_{j=1}^{n}\frac{(1+\rho\Delta_j)c^*Y_j(u)}{1+\rho c^*\hat{A}_n(1)}\right) dN_i(u)$$

$$-\sum_{p=1}^{P}\frac{1}{n}\sum_{i=1}^{n} Y_i(s_{p+1})(1-Y_i(s_p))\int_0^1 \log\left(\frac{1}{n}\sum_{j=1}^{n}\frac{(1+\rho\Delta_j)c^*Y_j(u)}{1+\rho c^*\hat{A}_n(s_p)}\right) dN_i(u)$$

$$-\frac{1}{n}\sum_{i=1}^{n}(1-Y_i(s_{P+1}))\int_0^1 \log\left(\frac{1}{n}\sum_{j=1}^{n}\frac{(1+\rho\Delta_j)c^*Y_j(u)}{1+\rho c^*\hat{A}_n(X_j)}\right) dN_i(u)$$

The third term in (A.1) is bounded from above by

$$-\frac{1}{n}\sum_{i=1}^{n}(1/\rho+\Delta_i)Y_i(1)\log(1+\rho c^*\hat{A}_n(1))$$

$$-\sum_{p=1}^{P}\frac{1}{n}\sum_{i=1}^{n}(1/\rho+\Delta_i)Y_i(s_p)(1-Y_i(s_{p-1}))\log(1+\rho c^*\hat{A}_n(s_p))$$

$$-\frac{1}{n}\sum_{i=1}^{n}(1/\rho+\Delta_i)(1-Y_i(s_P))\log(1+\rho c^*\hat{A}_n(X_i))$$

Combining the upper bounds for the second and third terms in (A.1), we get an overall upper bound of

$$\begin{aligned}
O(1) &- \frac{1}{n}\sum_{i=1}^{n}(1/\rho+\Delta_i)(1-Y_i(s_P))\log(1+\rho c^*\hat{A}_n(X_i)) \\
&-\frac{1}{n}\sum_{i=1}^{n} Y_i(s_1)\int_0^1 \log\left(\frac{1}{n}\sum_{j=1}^{n}(1+\rho\Delta_j)c^*Y_j(u)\right) dN_i(u) \\
&-\sum_{p=1}^{P}\frac{1}{n}\sum_{i=1}^{n} Y_i(s_{p+1})(1-Y_i(s_p))\int_0^1 \log\left(\frac{1}{n}\sum_{j=1}^{n}(1+\rho\Delta_j)c^*Y_j(u)\right) dN_i(u) \\
&-\log(1+\rho c^*\hat{A}_n(1))\left\{\frac{1}{n}\sum_{i=1}^{n}(1/\rho+\Delta_i)Y_i(1)-\frac{1}{n}\sum_{i=1}^{n}\Delta_i Y_i(s_1)\right\}
\end{aligned}$$

$$
\begin{aligned}
&- \sum_{p=1}^{P} \log(1 + \rho c^* \hat{A}_n(s_p)) \left\{ \frac{1}{n} \sum_{i=1}^{n} (1/\rho + \Delta_i) Y_i(s_p)(1 - Y_i(s_{p-1})\right. \\
&\qquad\qquad\qquad\qquad\qquad\qquad \left. - \frac{1}{n} \sum_{i=1}^{n} \Delta_i Y_i(s_{p+1})(1 - Y_i(s_p)) \right\} \\
&- \frac{1}{n} \sum_{i=1}^{n} (1 - Y_i(s_{P+1})) \int_0^1 \log\left(\frac{1}{n} \sum_{j=1}^{n} \frac{(1 + \rho\Delta_j) c^* Y_j(u)}{1 + \rho c^* \hat{A}_n(X_j)} \right) dN_i(u)
\end{aligned}
$$

Note that we can ignore the second term above because it is negative for all n. The third term is bounded away from positive infinity. To see this, note that this term is less than or equal to

$$
- \left\{ \log(c^*) + \log\left(\frac{1}{n} \sum_{j=1}^{n} Y_j(1) \right) \right\} \frac{1}{n} \sum_{i=1}^{n} \Delta_i Y_i(s_1)
$$

This term converges to $-\{\log(c^*) + \log(E[Y(1)])\} E[\Delta Y(s_1)]$, which is a bounded quantity. For finite P, the fourth term is also bounded away from positive infinity. To establish this, we note that the fourth term is less than or equal to

$$
- \sum_{p=1}^{P} \left\{ \log(c^*) + \log\left(\frac{1}{n} \sum_{j=1}^{n} Y_j(s_p) \right) \right\} \frac{1}{n} \sum_{i=1}^{n} \Delta_i Y_i(s_{p+1})(1 - Y_i(s_p))
$$

which converges to the bounded quantity $- \sum_{p=1}^{P} \{\log(c^*) + \log(E[Y(s_p)])\} E[\Delta Y(s_{p+1})(1 - Y(s_p))]$. If P is infinite, then we will have a problem. However, we can show how to choose a finite sequence $\{s_p\}$ such that the sum of the last three terms diverges to negative infinity. The sequence is constructed by the following recursive algorithm. Choose $U > 1$ and $s_0 = 1$. Let s_1 be the smallest value in the interval $[0, s_0)$ for which

$$
E[((U\rho)^{-1} + \Delta) Y(1)] \geq E[\Delta Y(s_1)]
$$

If $s_1 = 0$ then stop. If $s_1 > 0$, then continuity of $Pr[Y(t) = 1]$ implies equality above. Then, continue according to the following procedure.

1. Set $p = 1$.

2. Given s_p, let s_{p+1} be the smallest value in the interval $[0, s_p)$ for which

$$
E[((U\rho)^{-1} + \Delta) Y(s_p)(1 - Y(s_{p-1})] \geq E[\Delta Y(s_{p+1})(1 - Y(s_p))]
$$

3. If $s_{p+1} = 0$, then stop. If $s_{p+1} > 0$, continuity of $Pr[Y(t) = 1]$ implies equality above. Then, increment p by 1 and return to step 2.

This algorithm must converge in a finite number of steps. Consider a proof by contradiction. For finite K, we know that

$$E[((U\rho)^{-1}+\Delta)Y(1)]+\sum_{p=1}^{K}E[((U\rho)^{-1}+\Delta)Y(s_p)(1-Y(s_{p-1}))]$$

is equal to

$$E[\Delta Y(s_1)]+\sum_{p=1}^{K}E[\Delta Y(s_{p+1})(1-Y(s_p))]$$

This implies that

$$E[((U\rho)^{-1}+\Delta)Y(s_K)]=E[\Delta Y(s_{K+1})] \tag{A.2}$$

Since $\{s_p\}$ is strictly decreasing and positive, it converges to a value s^0. Taking the limit of (A.2) as K goes to infinity, we find that

$$E[((U\rho)^{-1}+\Delta)Y(s^0+)]=E[\Delta Y(s^0+)],$$

which is impossible. Therefore, a finite P exists with $s_{P+1}=0$. By construction of the algorithm, we can now see that the sum of the fifth, sixth, and seventh terms of the upper bound diverge to negative infinity. In particular, we see that the coefficients of $\log(1+\rho c^*\hat{A}_n(s_p))$ are strictly positive for large n. This implies the the sum of the fifth and sixth terms diverge to negative infinity. In addition, the seventh term is identically equal to zero. Hence, we have established a contradiction to the original divergence premise and we know that $\hat{A}_n$ does not diverge. ■

LEMMA 2 *Suppose that* $\inf_{t\in(0,1]}E[Y(t)|Z]>0$. *Then*

$$\sup_{t\in[0,1]}|\bar{A}_n(t)-A_0(t)|\to 0 \textit{ a.s.},$$

and for each $\omega\in\Phi$,

- *A is absolutely continuous.*
- $\sup_{t\in[0,1]}|\frac{d\hat{A}_{m_k}}{d\bar{A}_{m_k}}(t)-\gamma(t)|\to 0$
- $\sup_{t\in[0,1]}|\hat{A}_{m_k}(t)-\int_0^t\gamma\,dA_0|\to 0$

where $\gamma(u)$ is defined by (11)

Proof: To prove the first result, note that

$$A_0(t)=\int_0^t\left(E\left[\frac{(1+\rho\Delta)Y(u)\exp(\beta_0'Z)}{1+\rho A_0(X)\exp(\beta_0'Z)}\right]\right)^{-1}dE[N(u)]$$

where $E[N(u)]=E[\int_0^u\frac{Y(v)\exp(\beta_0'Z)}{1+\rho A_0(u)\exp(\beta_0'Z)}dA_0(u)]$. Then,

$$\begin{aligned}\bar{A}_n(t)-A_0(t) \;=\; &\int_0^t\left(\frac{1}{n}\sum_{i=1}^{n}\frac{(1+\rho\Delta_i)Y_i(u)\exp(\beta_0'Z_i)}{1+\rho A_0(X_i)\exp(\beta_0'Z_i)}\right)^{-1}d\bar{N}_n(u)\\&-\int_0^t\left(E\left[\frac{(1+\rho\Delta)Y(u)\exp(\beta_0'Z)}{1+\rho A_0(X)\exp(\beta_0'Z)}\right]\right)^{-1}dE[N(u)]\end{aligned}$$

By Lemma A.1 of Tsiatis (1981), we know that

$$\sup_{t\in[0,1]}\left|\frac{1}{n}\sum_{i=1}^{n}\frac{(1+\rho\Delta_i)Y_i(u)\exp(\beta_0'Z_i)}{1+\rho A_0(X_i)\exp(\beta_0'Z_i)}-E\left[\frac{(1+\rho\Delta)Y(u)\exp(\beta_0'Z)}{1+\rho A_0(X)\exp(\beta_0'Z)}\right]\right|\to 0\text{ a.s.}$$

Using Lemma A.2 of Tsiatis (1981), we can then establish that $\sup_{t\in[0,1]}|\bar{A}_n(t)-A_0(t)|\to 0$ almost surely.

For the next three results, we fix $\omega\in\Phi$. The following argument shows that A must be absolutely continuous. If we let f be any non-negative, bounded, continuous function, then

$$\begin{aligned}\int_0^1 f(u)dA(u) &= \int_0^1 f(u)d(A(u)-\hat{A}_{m_k}(u)) \\ &\quad+\int_0^1 f(u)\left(\frac{1}{m_k}\sum_{i=1}^{m_k}\frac{(1+\rho\Delta_i)Y_i(u)\exp(\hat{\beta}_{m_k}'Z_i)}{1+\rho\hat{A}_{m_k}(X_i)\exp(\hat{\beta}_{m_k}'Z_i)}\right)^{-1}d\bar{N}_{m_k}(u) \\ &\le \int_0^1 f(u)d(A(u)-\hat{A}_{m_k}(u)) \\ &\quad+(\exp(qbc)+a)\int_0^1 f(u)\left(\frac{1}{m_k}\sum_{i=1}^{m_k}Y_i(u)\right)^{-1}d\bar{N}_{m_k}(u)\end{aligned}$$

where $\lim_{m_k\to\infty}\hat{A}_{m_k}(1)<a$. Consider each of the two terms on the right hand side of the above inequality. The first term converges to zero by the Helly-Bray lemma (page 180 of Loeve, 1963). Using Lemmas A.1 and A.2 of Tsiatis (1981), we know that the second term converges to

$$(\exp(qbc)+a)\int_0^1 f(u)\,(E[Y(u)])^{-1}\,E\left[\frac{Y(u)\exp(\beta_0'Z)}{1+\rho A_0(u)\exp(\beta_0'Z)}\right]dA_0(u)$$

Since $E[Y(u)]$ is assumed to be bounded away from zero for all $u\in(0,1]$, we know that

$$\int_0^1 f(u)dA(u)\ \le\ (\exp(qbc)+a)\int_0^1 f(u)\,(E[Y(u)])^{-1}\,E\left[\frac{Y(u)\exp(\beta_0'Z)}{1+\rho A_0(u)\exp(\beta_0'Z)}\right]dA_0(u)$$

By choosing f appropriately, this inequality implies that A must be continuous at the continuity points of A_0. Since A_0 is assumed to be absolutely continuous, then so is A.

To prove the third result, consider

$$\frac{d\hat{A}_{m_k}}{d\bar{A}_{m_k}}(t) = \frac{\frac{1}{m_k}\sum_{i=1}^{m_k} \frac{(1+\rho\Delta_i)Y_i(t)\exp(\beta_0' Z_i)}{1+\rho A_0(X_i)\exp(\beta_0' Z_i)}}{\frac{1}{m_k}\sum_{i=1}^{m_k} \frac{(1+\rho\Delta_i)Y_i(t)\exp(\hat{\beta}_{m_k}' Z_i)}{1+\rho \hat{A}_{m_k}(X_i)\exp(\hat{\beta}_{m_k}' Z_i)}} \tag{A.3}$$

By the strong law of large numbers, the numerator converges to its expectation in supremum norm. That is,

$$\sup_{t\in[0,1]}\left|\frac{1}{m_k}\sum_{i=1}^{m_k}\frac{(1+\rho\Delta_i)Y_i(t)\exp(\beta_0' Z_i)}{1+\rho A_0(X_i)\exp(\beta_0' Z_i)} - E\left[\frac{(1+\rho\Delta)Y(t)\exp(\beta_0' Z)}{1+\rho A_0(X)\exp(\beta_0' Z)}\right]\right| \to 0$$

The denominator should converge to $E[\frac{(1+\rho\Delta)Y(t)\exp(\beta' Z)}{1+\rho A(X)\exp(\beta' Z)}]$ in supremum norm. To show this, we see that

$$\begin{aligned}
&\left|\frac{1}{m_k}\sum_{i=1}^{m_k}\frac{(1+\rho\Delta_i)Y_i(t)\exp(\hat{\beta}_{m_k}' Z_i)}{1+\rho \hat{A}_{m_k}(X_i)\exp(\hat{\beta}_{m_k}' Z_i)} - E\left[\frac{(1+\rho\Delta)Y(t)\exp(\beta' Z)}{1+\rho A(X)\exp(\beta' Z)}\right]\right| \\
&\quad\le \left|\frac{1}{m_k}\sum_{i=1}^{m_k}\frac{(1+\rho\Delta_i)Y_i(t)\exp(\hat{\beta}_{m_k}' Z_i)}{1+\rho \hat{A}_{m_k}(X_i)\exp(\hat{\beta}_{m_k}' Z_i)} - \frac{1}{m_k}\sum_{i=1}^{m_k}\frac{(1+\rho\Delta_i)Y_i(t)\exp(\beta' Z_i)}{1+\rho \hat{A}_{m_k}(X_i)\exp(\hat{\beta}_{m_k}' Z_i)}\right| \\
&\qquad + \left|\frac{1}{m_k}\sum_{i=1}^{m_k}\frac{(1+\rho\Delta_i)Y_i(t)\exp(\beta' Z_i)}{1+\rho \hat{A}_{m_k}(X_i)\exp(\hat{\beta}_{m_k}' Z_i)} - \frac{1}{m_k}\sum_{i=1}^{m_k}\frac{(1+\rho\Delta_i)Y_i(t)\exp(\beta' Z_i)}{1+\rho \hat{A}_{m_k}(X_i)\exp(\beta' Z_i)}\right| \\
&\qquad + \left|\frac{1}{m_k}\sum_{i=1}^{m_k}\frac{(1+\rho\Delta_i)Y_i(t)\exp(\beta' Z_i)}{1+\rho \hat{A}_{m_k}(X_i)\exp(\beta' Z_i)} - \frac{1}{m_k}\sum_{i=1}^{m_k}\frac{(1+\rho\Delta_i)Y_i(t)\exp(\beta' Z_i)}{1+\rho A(X_i)\exp(\beta' Z_i)}\right| \\
&\qquad + \left|\frac{1}{m_k}\sum_{i=1}^{m_k}\frac{(1+\rho\Delta_i)Y_i(t)\exp(\beta' Z_i)}{1+\rho A(X_i)\exp(\beta' Z_i)} - E\left[\frac{(1+\rho\Delta)Y(t)\exp(\beta' Z)}{1+\rho A(X)\exp(\beta' Z)}\right]\right| \\
&\quad\le (1+\rho)(1+\rho a)\exp(qbc)|1-\exp(-\|\hat{\beta}_{m_k}-\beta\|_\infty c)| \\
&\qquad + (1+\rho)\rho\exp(2qbc)\sup_{u\in[0,1]}|\hat{A}_{m_k}(u)-A(u)| \\
&\qquad + \left|\frac{1}{m_k}\sum_{i=1}^{m_k}\frac{(1+\rho\Delta_i)Y_i(t)\exp(\beta' Z_i)}{1+\rho A(X_i)\exp(\beta' Z_i)} - E\left[\frac{(1+\rho\Delta)Y(t)\exp(\beta' Z)}{1+\rho A(X)\exp(\beta' Z)}\right]\right|
\end{aligned}$$

The first two terms on the right hand side of the final inequality above converge to zero. At first thought it appears that the law of large numbers can be used to show that the third term converges to zero in supremum norm. The problem arises that β and A change with ω. Furthermore, the sets of probability one over which the strong law of large numbers applies depend on β and A. Since there are an uncountable number of ω's, there may be an uncountable number of (β, A)'s. This uncountable number of limiting quantities may lead to an uncountable number of sets of probability one, whose intersection might have probability zero. In this event, the strong law of large numbers may break down. To avoid this issue, we note that the space of absolutely continuous, bounded, increasing functions $\{A(t)\}$ is separable under the supremum norm. That is, the space has a countably dense subset. Let $\{\mathcal{G}_l\}$ represent this subset. Now, include in the intersections of sets forming Φ, sets for which

$$\sup_{t\in[0,1]}\left|\frac{1}{n}\sum_{i=1}^{n}\frac{(1+\rho\Delta_i)Y_i(t)\exp(\xi'Z_i)}{1+\rho\mathcal{G}_l(X_i)\exp(\xi'Z_i)}-E\left[\frac{(1+\rho\Delta)Y(t)\exp(\xi'Z)}{1+\rho\mathcal{G}_l(X)\exp(\xi'Z)}\right]\right|$$

converge to zero for each rational ξ and $l \geq 1$. Now,

$$\begin{aligned}
&\left|\frac{1}{m_k}\sum_{i=1}^{m_k}\frac{(1+\rho\Delta_i)Y_i(t)\exp(\beta'Z_i)}{1+\rho A(X_i)\exp(\beta'Z_i)}-E\left[\frac{(1+\rho\Delta)Y(t)\exp(\beta'Z)}{1+\rho A(X)\exp(\beta'Z)}\right]\right|\\
&\quad\leq\left|\frac{1}{m_k}\sum_{i=1}^{m_k}\frac{(1+\rho\Delta_i)Y_i(t)\exp(\beta'Z_i)}{1+\rho A(X_i)\exp(\beta'Z_i)}-\frac{1}{m_k}\sum_{i=1}^{m_k}\frac{(1+\rho\Delta_i)Y_i(t)\exp(\xi'Z_i)}{1+\rho A(X_i)\exp(\beta'Z_i)}\right|\\
&\qquad+\left|\frac{1}{m_k}\sum_{i=1}^{m_k}\frac{(1+\rho\Delta_i)Y_i(t)\exp(\xi'Z_i)}{1+\rho A(X_i)\exp(\beta'Z_i)}-\frac{1}{m_k}\sum_{i=1}^{m_k}\frac{(1+\rho\Delta_i)Y_i(t)\exp(\xi'Z_i)}{1+\rho A(X_i)\exp(\xi'Z_i)}\right|\\
&\qquad+\left|\frac{1}{m_k}\sum_{i=1}^{m_k}\frac{(1+\rho\Delta_i)Y_i(t)\exp(\xi'Z_i)}{1+\rho A(X_i)\exp(\xi'Z_i)}-\frac{1}{m_k}\sum_{i=1}^{m_k}\frac{(1+\rho\Delta_i)Y_i(t)\exp(\xi'Z_i)}{1+\rho\mathcal{G}_l(X_i)\exp(\xi'Z_i)}\right|\\
&\qquad+\left|\frac{1}{m_k}\sum_{i=1}^{m_k}\frac{(1+\rho\Delta_i)Y_i(t)\exp(\xi'Z_i)}{1+\rho\mathcal{G}_l(X_i)\exp(\xi'Z_i)}-E\left[\frac{(1+\rho\Delta)Y(t)\exp(\xi'Z)}{1+\rho\mathcal{G}_l(X)\exp(\xi'Z)}\right]\right|\\
&\qquad+\left|E\left[\frac{(1+\rho\Delta)Y(t)\exp(\xi'Z)}{1+\rho\mathcal{G}_l(X)\exp(\xi'Z)}\right]-E\left[\frac{(1+\rho\Delta)Y(t)\exp(\beta'Z)}{1+\rho\mathcal{G}_l(X)\exp(\xi'Z)}\right]\right|\\
&\qquad+\left|E\left[\frac{(1+\rho\Delta)Y(t)\exp(\beta'Z)}{1+\rho\mathcal{G}_l(X)\exp(\xi'Z)}\right]-E\left[\frac{(1+\rho\Delta)Y(t)\exp(\beta'Z)}{1+\rho\mathcal{G}_l(X)\exp(\beta'Z)}\right]\right|\\
&\qquad+\left|E\left[\frac{(1+\rho\Delta)Y(t)\exp(\beta'Z)}{1+\rho\mathcal{G}_l(X)\exp(\beta'Z)}\right]-E\left[\frac{(1+\rho\Delta)Y(t)\exp(\beta'Z)}{1+\rho A(X)\exp(\beta'Z)}\right]\right|\\
&\quad\leq 2(1+\rho)(1+\rho a)e^{qbc}|1-e^{-\|\beta-\xi\|_\infty c}|+2(1+\rho)\rho e^{2qbc}\sup_{u\in[0,1]}|A(u)-\mathcal{G}_l(u)|\\
&\qquad+\left|\frac{1}{m_k}\sum_{i=1}^{m_k}\frac{(1+\rho\Delta_i)Y_i(t)\exp(\xi'Z_i)}{1+\rho\mathcal{G}_l(X_i)\exp(\xi'Z_i)}-E\left[\frac{(1+\rho\Delta)Y(t)\exp(\xi'Z)}{1+\rho\mathcal{G}_l(X)\exp(\xi'Z)}\right]\right|
\end{aligned}$$

As $m_k \to \infty$, the three terms on the right hand side of the final inequality above can be made as small as possible via a proper choice of ξ and l. Hence, we have shown that

$$\sup_{t\in[0,1]}\left|\frac{1}{m_k}\sum_{i=1}^{m_k}\frac{(1+\rho\Delta_i)Y_i(t)\exp(\beta'Z_i)}{1+\rho A(X_i)\exp(\beta'Z_i)}-E\left[\frac{(1+\rho\Delta)Y(t)\exp(\beta'Z)}{1+\rho A(X)\exp(\beta'Z)}\right]\right|\to 0$$

With this result, we can now conclude that the denominator of (A.3) converges in supremum norm to $E[\frac{(1+\rho\Delta)Y(t)\exp(\beta' Z)}{1+\rho A(X)\exp(\beta' Z)}]$. Combining the results for the numerator and denominator, we can conclude that $\sup_{t\in[0,1]} |\frac{d\hat{A}_{m_k}}{d\bar{A}_{m_k}}(t) - \gamma(t)| \to 0$.

To prove the final result, note that

$$\begin{aligned}
&\left|\hat{A}_{m_k}(t) - \int_0^t \gamma(u) dA_0(u)\right| \\
&\quad = \left|\int_0^t \left(\frac{1}{m_k}\sum_{i=1}^{m_k} \frac{(1+\rho\Delta_i)Y_i(u)\exp(\hat{\beta}'_{m_k} Z_i)}{1+\rho\hat{A}_{m_k}(X_i)\exp(\hat{\beta}'_{m_k} Z_i)}\right)^{-1} d\bar{N}_{m_k}(u) - \int_0^t \gamma(u) dA_0(u)\right| \\
&\quad = \left|\int_0^t \left(\frac{1}{m_k}\sum_{i=1}^{m_k} \frac{(1+\rho\Delta_i)Y_i(u)\exp(\hat{\beta}'_{m_k} Z_i)}{1+\rho\hat{A}_{m_k}(X_i)\exp(\hat{\beta}'_{m_k} Z_i)}\right)^{-1} d\bar{N}_{m_k}(u)\right. \\
&\qquad \left. - \int_0^t \left(E\left[\frac{(1+\rho\Delta)Y(u)\exp(\beta' Z)}{1+\rho A(X)\exp(\beta' Z)}\right]\right)^{-1} dE[N(u)]\right| \\
&\quad \le \left|\int_0^t \left(\frac{1}{m_k}\sum_{i=1}^{m_k} \frac{(1+\rho\Delta_i)Y_i(u)\exp(\hat{\beta}'_{m_k} Z_i)}{1+\rho\hat{A}_{m_k}(X_i)\exp(\hat{\beta}'_{m_k} Z_i)}\right)^{-1}\right. \\
&\qquad \left. - \left(E\left[\frac{(1+\rho\Delta)Y(u)\exp(\beta' Z)}{1+\rho A(X)\exp(\beta' Z)}\right]\right)^{-1} d\bar{N}_{m_k}(u)\right| \\
&\qquad + \left|\int_0^t \left(E\left[\frac{(1+\rho\Delta)Y(u)\exp(\beta' Z)}{1+\rho A(X)\exp(\beta' Z)}\right]\right)^{-1} d(\bar{N}_{m_k}(u) - E[N(u)])\right| \\
&\quad \le \sup_{u\in[0,t]} \left|\left(\frac{1}{m_k}\sum_{i=1}^{m_k} \frac{(1+\rho\Delta_i)Y_i(u)\exp(\hat{\beta}'_{m_k} Z_i)}{1+\rho\hat{A}_{m_k}(X_i)\exp(\hat{\beta}'_{m_k} Z_i)}\right)^{-1}\right. \\
&\qquad \left. - \left(E\left[\frac{(1+\rho\Delta)Y(u)\exp(\beta' Z)}{1+\rho A(X)\exp(\beta' Z)}\right]\right)^{-1}\right| \\
&\qquad + \left|\int_0^t \left(E\left[\frac{(1+\rho\Delta)Y(u)\exp(\beta' Z)}{1+\rho A(X)\exp(\beta' Z)}\right]\right)^{-1} d(\bar{N}_{m_k}(u) - E[N(u)])\right|
\end{aligned}$$

The first term on the right hand side of the inequality converges to zero because the denominator of (A.3) was shown to converge to $E[\frac{(1+\rho\Delta)Y(t)\exp(\beta' Z)}{1+\rho A(X)\exp(\beta' Z)}]$ in supremum norm. The second term converges to zero by the Helly-Bray lemma. Pointwise convergence can be strengthened to uniform convergence by applying the same monotonicity argument used in the proof of the Glivenko-Cantelli Theorem (page 96 of Shorack and Wellner, 1986).

■

Acknowledgments

This research was sponsored in part by National Institute of Health Grants R01-AI-31789 from NIAID, R01-CA-51962 from NCI, R01-AI-32475 from NIAID, R01-MH-56639 from NIMH, and R01-DA-10184 from NIDA . The authors are grateful to Drs. James Robins and Aad van der Vaart for their help in the preparation of this manuscript.

References

P. K. Andersen and R. D. Gill, “Cox’s regression model for counting processes: A large sample study,” *The Annals of Statistics* vol. 10 pp. 1100–1120, 1982.

R. B. Ash, *Real Analysis and Probability*, Academic Press: San Diego, CA, 1972.

S. Bennett, “Analysis of survival data by the proportional odds model,” *Statistics in Medicine* vol. 2 pp. 273–277, 1983.

P. J. Bickel and K. A. Doksum, “An analysis of transformations revisted,” *Journal of the American Statistical Association* vol. 76 pp. 296-311, 1981.

P. Billingsley, *Probability and Measure*, John Wiley & Sons: New York, NY, 1986.

S. C. Cheng, L. J. Wei, and Z. Ying, “Analysis of transformation models with censored data,” *Biometrika* vol. 82 pp. 835–845, 1995.

S. C. Cheng, L. J. Wei, and Z. Ying, “Predicting survival probabilities with semiparametric transformation models,” *Journal of the American Statistical Association* vol. 92 pp. 227–235, 1997.

D. Clayton and J. Cuzick, “The semi-parametric pareto model for regression analysis of survival times,” *Proceedings of International Statistical Institute*, Amsterdam, 1986.

D. R. Cox, “Regression models and life-tables,” *Journal of the Royal Statistical Society, Series B* vol. 34 pp. 187–220, 1972.

D. R. Cox, “Partial likelihood,” *Biometrika* vol. 62 pp. 269–276, 1975.

D. M. Dabrowska and K. A. Doksum, “Estimation and testing in a two-sample generalized odds-rate model,” *Journal of the American Statistical Association* vol. 83 pp. 744–749, 1988a.

D. M. Dabrowska and K. A. Doksum, “Partial likelihood in transformation models with censored data,” *Scandinavian Journal of Statistics* vol. 15 pp. 1–23, 1988b.

D. P. Harrington and T. R. Fleming, “A class of rank test procedures for censored survival data,” *Biometrika* vol. 69 pp. 553–566, 1982.

F. Hsieh, “The empirical process approach for semiparametric two-sample models with heterogeneous treatment effect,” *Journal of the Royal Statistical Society, Series B* vol. 4 pp. 735–748, 1995.

D. V. Hinkley and G. Runger, “The analysis of transformed data,” *Journal of the American Statistical Association* vol. 79 pp. 302–309, 1984.

J. D. Kalbfleisch and R. L. Prentice, *The Statistical Analysis of Failure Time Data*, John Wiley & Sons: New York, NY, 1980.

L. V. Kantorovich and G. P. Akilov, *Functional Analysis*, Pergamon Press: Oxford, 1992.

R. Kress, *Linear Integral Equations*, Springer-Varlag: Berlin, 1989.

M. Loeve, *Probability Theory*, D. Van Nordstrand Company: Princeton, NJ, 1963.

D. G. Luenberger, *Optimization by Vector Space Methods*, John Wiley & Sons: New York, NY, 1969.

W. K. Newey, “Semiparametric efficiency bounds,” *Journal of Applied Econometrics* vol. 5 pp. 99–135, 1990.

S. A. Murphy, “Consistency in a proportional hazards model incorporating a random effect,” *The Annals of Statistics* vol. 22 pp. 712–731, 1994.

S. A. Murphy, “Asymptotic theory for the frailty model,” *The Annals of Statistics* vol. 23 pp. 182–198, 1995.

S. A. Murphy, A. J. Rossini, and A. W. van der Vaart, “Maximum likelihood estimation in the proportional odds model,” *Journal of the American Statistical Association* vol. 92 pp. 968–976, 1997.

M. I. Parzen and D. P. Harrington, “Proportional odds regression with right-censored data using adaptive integrated splines,” Technical Report, Harvard School of Public Health, 1993.

A. N. Pettitt, “Approximate methods using ranks for regression with censored data,” *Biometrika* vol. 70 pp. 121–132, 1983.

A. N. Pettitt, "Proportional odds models for survival data and estimates using ranks," *Applied Statistics* vol. 33 pp. 169–175, 1984.

D. Pollard, *Convergence of Stochastic Processes*, Springer-Verlag: New York, NY, 1984.

D. Pollard, *Empirical Processes: Theory and Applications*, Institute of Mathematical Sciences: Hayward, CA, 1990.

R. L. Prentice, "Exponential survivals with censoring and explanatory variables," *Biometrika* vol. 60 pp. 279–288, 1973.

J. M. Robins, A. Rotnitzky, and L. P. Zhao, "Estimation of regression coefficients when some regressors are not always observed," *Journal of the American Statistical Association* vol. 89 pp. 846–866, 1994.

W. Rudin, *Functional Analysis*, McGraw-Hill: New York, NY, 1973.

R. J. Serfling, *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons: New York, NY, 1980.

G. R. Shorack and J. A. Wellner, *Empirical Processes with Applications to Statistics*, John Wiley & Sons: New York, NY, 1986

A. A. Tsiatis, "A large sample study of Cox's regression model," *The Annals of Statistics* vol. 8 pp. 93–108, 1981.

A. W. van der Vaart, "Efficiency of infinite dimensional M-estimators," *Statistica Neerlandica* vol. 49 pp. 9–30, 1995.

A. W. van der Vaart and J. A. Wellner, *Weak Convergence and Empirical Processes*, Springer-Verlag: New York, NY, 1996.