mathematics

MDPI

Article

Asymptotic Properties for Cumulative Probability Models for Continuous Outcomes

Chun Li [1,*], Yuqi Tian [2], Donglin Zeng [3] and Bryan E. Shepherd [2]

[1] Division of Biostatistics, Department of Population and Public Health Sciences, University of Southern California, Los Angeles, CA 90033, USA
[2] Department of Biostatistics, Vanderbilt University, Nashville, TN 37203, USA; yuqi.tian@vanderbilt.edu (Y.T.); bryan.shepherd@vanderbilt.edu (B.E.S.)
[3] Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA; dzeng@umich.edu
* Correspondence: cli77199@usc.edu

Abstract: Regression models for continuous outcomes frequently require a transformation of the outcome, which is often specified a priori or estimated from a parametric family. Cumulative probability models (CPMs) nonparametrically estimate the transformation by treating the continuous outcome as if it is ordered categorically. They thus represent a flexible analysis approach for continuous outcomes. However, it is difficult to establish asymptotic properties for CPMs due to the potentially unbounded range of the transformation. Here we show asymptotic properties for CPMs when applied to slightly modified data where bounds, one lower and one upper, are chosen and the outcomes outside the bounds are set as two ordinal categories. We prove the uniform consistency of the estimated regression coefficients and of the estimated transformation function between the bounds. We also describe their joint asymptotic distribution, and show that the estimated regression coefficients attain the semiparametric efficiency bound. We show with simulations that results from this approach and those from using the CPM on the original data are very similar when a small fraction of the data are modified. We reanalyze a dataset of HIV-positive patients with CPMs to illustrate and compare the approaches.

Keywords: cumulative probability model; semiparametric transformation model; uniform consistency; asymptotic distribution

MSC: 62G99

Citation: Li, C.; Tian, Y.; Zeng, D.; Shepherd, B.E. Asymptotic Properties for Cumulative Probability Models for Continuous Outcomes. *Mathematics* **2023**, *11*, 4896. https://doi.org/10.3390/math11244896

Academic Editors: Antonio Musolesi and Alexandra Soberon

Received: 31 October 2023
Revised: 28 November 2023
Accepted: 4 December 2023
Published: 7 December 2023

1. Introduction

Regression analyses of continuous outcomes often require a transformation of the outcome to meet modeling assumptions. In practice, convenient but ad hoc transformations such as a logarithm or square root are often used on right-skewed outcomes; an alternative is to use the Box–Cox family [1] of transformations, which is effectively a family of power functions plus the logarithm transformation. Because the correct transformation for the continuous outcome is often unknown and it may not fall in a prespecified family, it is desirable to estimate the transformation in a flexible way. Semiparametric transformation models have been introduced to address this issue [2,3]. These models involve a latent intermediate variable and two model components: one connecting the latent variable to the outcome variable through an unknown transformation and the other connecting the latent variable to the input variables as in traditional regression models with unknown beta coefficients.

Early parameter estimation for semiparametric transformation models was based on the marginal likelihood of the vector of outcome ranks [2–4]. Although this marginal likelihood can be simplified to the partial likelihood in Cox proportional hazards models [5], it cannot be simplified for other transformation models, and various approximations had

to be used. As the marginal likelihood only involves the beta coefficients, additional ad hoc procedures were developed to estimate the transformation [2,3].

Later developments for fitting semiparametric linear transformation models primarily focused on right-censored data, initially relying on estimating equations [6,7]. Zeng and Lin [8] developed nonparametric maximum likelihood estimators (NPMLEs) based on likelihoods for time-to-event data and showed the consistency and asymptotic distribution of their estimators. NPMLEs are desirable because they are fully efficient. For continuous outcomes, more recent developments have used B-splines and Bernstein polynomials to flexibly model the transformation [9,10], but these estimators of the transformation are not fully nonparametric.

With continuous outcomes, one way to nonparametrically estimate the transformation is to treat the outcome as if it is ordinal—without any categorization—and fit to cumulative probability models (CPMs; also called cumulative link models) [11]. Liu et al. [11] showed that the CPM's multinomial likelihood for continuous outcomes is equivalent to the nonparametric likelihood for semiparametric transformation models. This result led to new NPMLEs for semiparametric transformation models for continuous outcomes using computationally simple ordinal regression methods. They showed with simulations that CPMs perform well in a wide range of scenarios. The method has since been used in applications to analyze various outcomes [12–19].

However, there is no established asymptotic theory for this new NPMLE approach for continuous outcomes. One main hurdle is that the unknown transformation of the continuous outcome variable can have an unbounded range of values, which makes it hard to establish asymptotic properties across the whole range. The approaches that were used to prove asymptotic properties for the NPMLE of the baseline cumulative hazard function for time-to-event outcomes cannot be applied directly to study transformation models for a continuous outcome [8] because the latter has no bounds on its range and no clear definition of a baseline hazard function.

To address this issue, we establish several asymptotic properties in this paper for CPMs when they are applied to continuous outcomes with slight modification. Briefly, a lower bound L and an upper bound U for the outcome are chosen prior to analysis, the outcomes below L are set as the lowest category and those above U as the highest category, and then a CPM is fitted to the modified data. We prove that, in this approach, the nonparametric estimate of the transformation function is consistent (i.e., converges in probability to its true value) uniformly in the interval $[L, U]$. We then show that the estimator of the beta coefficients and that of the transformation jointly converge to a tight Gaussian process, and that the estimator of the beta coefficients attains the semiparametric efficiency bound. The latter implies that this estimator is (asymptotically) as efficient as possible under the assumptions of the model. We show with simulations and real data that the results from this approach and those from the CPM on the original data are very similar when only a small fraction of data are outside the bounds.

2. Method

2.1. Cumulative Probability Models

Let Y be the outcome of interest and Z be a vector of p covariates. The semiparametric linear transformation model is

$$Y = H(\beta^T Z + \epsilon), \tag{1}$$

where H is a transformation function assumed to be non-decreasing but unknown otherwise, β is a vector of coefficients, and ϵ is independent of Z and is assumed to follow a continuous distribution with cumulative distribution function $G(\cdot)$. An alternative expression of model (1) is

$$A(Y) = \beta^T Z + \epsilon, \tag{2}$$

where $A = H^{-1}$ is the inverse of H. For mathematical clarity, we assume H is left continuous and define $A(y) = \sup\{z : H(z) \leq y\}$; then, A is non-decreasing and right-continuous.

Model (1) is equivalent to the cumulative probability model (CPM) presented in Liu et al. [11]:

$$G^{-1}\{\mathrm{P}(Y \le y \mid Z)\} = A(y) - \beta^T Z, \text{ for any } y, \tag{3}$$

where $G^{-1}(\cdot)$ serves as a link function. One example of the distribution for ϵ is the standard normal distribution. In this case, the CPM becomes a normal linear model after a transformation, which includes log-linear models and linear models with a Box–Cox transformation as special cases. The CPM becomes a Cox proportional hazards model when ϵ follows the extreme value distribution, i.e. $G(x) = 1 - \exp(-e^x)$, or a proportional odds model when ϵ follows the logistic distribution, i.e. $G(x) = e^x/(1+e^x)$.

Suppose the data are i.i.d. and denoted as $(Y_i, Z_i), i = 1, \ldots, n$. Liu et al. [11] proposed to model the transformation A nonparametrically. The corresponding nonparametric (NP) likelihood is

$$\prod_{i=1}^{n}\Big[G\Big\{A(Y_i) - \beta^T Z_i\Big\} - G\Big\{A(Y_i^-) - \beta^T Z_i\Big\}\Big],$$

where $A(y^-) = \lim_{t \uparrow y} A(t)$. Since A can be any non-decreasing function, this likelihood will be maximized when the increments of $A(\cdot)$ are concentrated at the observed Y_i; if some increments of $A(\cdot)$ are not at the observed Y_i, its corresponding probability mass at non-observed values can always be reallocated to some observed values to increase the likelihood. Thus, we can maximize this likelihood by only considering step functions $A(\cdot)$ that have a jump at every observed Y_i. This leads to an expression of the likelihood that is the same as the likelihood of the CPM when the outcome variable is treated as if it were ordered categorically with the observed distinct values as the ordered categories. As a result, nonparametric maximum likelihood estimates (NPMLEs) can be obtained by fitting an ordinal regression model to the continuous outcome. Liu et al. [11] showed in simulations that CPMs perform well under a wide range of scenarios. However, it is difficult to prove the asymptotic properties for this approach. Since some Y_i can be extremely large or small and the observations at the tails are often sparse, there is high variability in the estimate of A at the tails. Moreover, the unboundedness of the transformation at the tails makes it difficult to control the compactness of the estimator of A, thus making most of asymptotic theory no longer applicable. In this paper, we prove asymptotic properties for CPMs when they are applied to continuous outcomes with slight modification. We describe the modification in Section 2.2 and show the asymptotic results in Section 2.3.

2.2. Cumulative Probability Models on Modified Data

In view of the challenges above, we hereby describe an approach in which the outcomes are modified at the two ends before a CPM is fit. We will then describe the asymptotic properties of this approach in Section 2.3 and show with simulations that the results from this approach and those of the CPM on the original data are similar when a small fraction of data are modified.

More specifically, we predetermine a lower bound L and an upper bound U, and consider all observations with $Y_i \le L$ as a single ordered category, which we conveniently denote as L, and those with $Y_i \ge U$ as a single ordered category, denoted as U. The bounds L and U should satisfy $\mathrm{P}(L < Y < U) > 0$, $\mathrm{P}(Y \le L) > 0$, and $\mathrm{P}(Y \ge U) > 0$. The new outcome variable, denoted as Y_i', follows a mixture distribution. When $Y_i' \in (L, U)$, the distribution is continuous with the same cumulative distribution function as that for Y_i; that is, $\mathrm{P}(Y_i' \le y \mid Z_i) = \mathrm{P}(Y_i \le y \mid Z_i) = G\{A(y) - \beta^T Z_i\}$ for $y \in (L, U)$. When $Y_i' = L$ or $Y_i' = U$, the distribution is discrete, with $\mathrm{P}(Y_i' = L \mid Z_i) = G\{A(L) - \beta^T Z_i\}$ and $\mathrm{P}(Y_i' = U \mid Z_i) = 1 - G\{A(U^-) - \beta^T Z_i\}$. Then, the nonparametric likelihood for the modified data is

$$\prod_{i=1}^{n}\Big(\Big[G\{A(Y_i)-\beta^T Z_i\}-G\{A(Y_i^-)-\beta^T Z_i\}\Big]^{I(Y_i\in(L,U))}\times$$
$$G\{A(L)-\beta^T Z_i\}^{I(Y_i\le L)}\times\Big[1-G\{A(U^-)-\beta^T Z_i\}\Big]^{I(Y_i\ge U)}\Big), \quad (4)$$

where $I(S)$ is the indicator function for event S with value 1 if S occurs and 0 otherwise.

Since $A(\cdot)$ can be any non-decreasing function over the interval $[L,U)$, the likelihood (4) will be maximized when the increments of $A(\cdot)$ are concentrated at the observed Y_i'. Hence, it suffices to consider only step functions with a jump at each distinct value of $Y_i' \in [L,U]$.

2.3. Asymptotic Results

From now on, we assume the outcome is continuous. Without loss of generality, we assume that in our models (1)–(3), the support of Y contains 0, the vector Z contains an intercept and has p dimensions, and $A(0)=0$. Furthermore, the two bounds satisfy $L<0$, $U>0$, $\mathrm{P}(L<Y<U)>0$, $\mathrm{P}(Y\le L)>0$, and $\mathrm{P}(Y\ge U)>0$. To establish the asymptotic properties described below, we further assume that

1. $G(x)$ is thrice-continuously differentiable, $G'(x)>0$ for any x, $G''(x)\mathrm{sign}(x)<0$ for $|x|\ge M$, where $M>0$ is a constant, and
$$\liminf_{x\to\infty} G'(x)/\{1-G(x)\}>0,\ \liminf_{x\to-\infty} G'(x)/G(x)>0.$$
2. The covariance matrix of Z is non-singular. In addition, Z and β are bounded so that $\beta^T Z\in[-m,m]$ almost certainly for some large constant m.
3. $A(y)$ is continuously differentiable in $(-\infty,\infty)$.

Condition 1 imposes restrictions on $G(x)$ at both tails; it holds for many residual distributions, including the standard normal distribution, the extreme value distribution and the logistic distribution. Conditions 2 and 3 are minimal assumptions for establishing asymptotic properties for linear transformation models.

Let $(\widehat{\beta},\widehat{A})$ denote the nonparametric maximum likelihood estimate of (β,A) that maximizes the likelihood (4) on the modified data. Then, $\widehat{A}$ is a step function with a jump at each of the distinct Y_i' in the modified data. To establish the asymptotic properties of $(\widehat{\beta},\widehat{A})$, we consider $\widehat{A}$ as a function over the closed interval $[L,U]$ by defining $\widehat{A}(U)=\widehat{A}(U^-)$. We have the following consistency theorem.

Theorem 1. *Under conditions 1–3, with probability one,*

$$\sup_{y\in[L,U]}|\widehat{A}(y)-A(y)|+\|\widehat{\beta}-\beta\|\to 0.$$

The proof of Theorem 1 is in Appendix A. Core steps of the proof include showing that $\widehat{A}$ is bounded in $[L,U]$ with probability one. Then, since $\widehat{A}(\cdot)$ is bounded and increasing in $[L,U]$, via the Helly selection theorem [20], for any subsequence, there exists a further subsequence that converges to a non-decreasing, right-continuous function at its continuity points. Thus, without confusion, we assume that $\widehat{A}\to A^*$ weakly in $[L,U]$ and $\widehat{\beta}\to\beta^*$. We then show that with probability one, $A^*(y)=A(y)$ for $y\in[L,U]$ and $\beta^*=\beta$. With this result, the consistency is established. Furthermore, since A is continuously differentiable, we conclude that $\widehat{A}(y)$ converges to $A(y)$ uniformly in $[L,U]$ with probability one.

We next describe the asymptotic distribution of $(\widehat{\beta},\widehat{A})$. The asymptotic distribution of $\widehat{A}$ will be expressed as that of a random functional in a metric space. We first define some notation. Let $BV[L,U]$ be the set of all functions defined over $[L,U]$ for which the total variation is at most one. Let $lin(BV[L,U])$ be the set of all linear functionals over $BV[L,U]$; that is, every element f in $lin(BV[L,U])$ is a linear function $f: BV[L,U]\to R$. For any $f\in lin(BV[L,U])$, its norm is defined as $\|f\|=\sup_{h\in BV[L,U]} f[h]$. A metric over

$lin(BV[L,U])$ can then be derived subsequently. Given any non-decreasing function A over $[L,U]$, a corresponding linear functional in $lin(BV[L,U])$, also denoted as A, can be defined such that for any $h \in BV[L,U]$,

$$A[h] = \int_L^U h(x)dA(x).$$

Similarly, for an nonparametric maximum likelihood estimate $\widehat{A}$, its corresponding linear functional in $lin(BV[L,U])$ is $\widehat{A}$, such that for any $h \in BV[L,U]$,

$$\widehat{A}[h] = \int_L^U h(x)d\widehat{A}(x).$$

The functional $\widehat{A}$ is a random element in the metric space $lin(BV[L,U])$. For any $y \in (L,U)$, there exists an $h \in BV[L,U]$, such that $\widehat{A}(y) = \widehat{A}[h]$. For example, suppose the estimated jump sizes at the distinct outcome values of a dataset, $\{a_1,\ldots,a_J\}$, are $\{\hat{s}_1,\ldots,\hat{s}_J\}$. Then, at $y_0 > 0$, $\widehat{A}(y_0) = \sum_{0<a_j\le y_0}\hat{s}_j = \widehat{A}[h_0]$, where $h_0(y) = I(0 < y \le y_0)$; and, similarly, at $y_0 < 0$, $\widehat{A}(y_0) = -\sum_{y_0<a_j<0}\hat{s}_j = \widehat{A}[h_0]$, where $h_0(y) = I(y_0 < y < 0)$.

Theorem 2. *Under conditions 1 – 3, $n^{1/2}(\widehat{\beta}-\beta, \widehat{A}-A)$ converges weakly to a tight Gaussian process in $R^p \times lin(BV[L,U])$. Furthermore, the asymptotic variance of $n^{1/2}(\widehat{\beta}-\beta)$ attains the semiparametric efficiency bound.*

The proof of Theorem 2 is in Appendix B and makes use of weak convergence theory for empirical processes and semiparametric efficiency theory. Its proof relies on verifying all the technical conditions in the Master Z-Theorem in [21]. In particular, it entails verification of the invertibility of the information operator for (β, A).

Because the information operator for (β, A) is invertible, the arguments given in [22] imply that the asymptotic variance-covariance matrix of $(\widehat{\beta}, \widehat{A}[h_1],\ldots,\widehat{A}[h_m])$ for any $h_1,\ldots,h_m \in BV[L,U]$ can be consistently estimated based on the information matrix for $\widehat{\beta}$ and the jump sizes of $\widehat{A}$. Specifically, suppose the estimated jump sizes at the distinct outcome values of a dataset, $\{a_1,\ldots,a_J\}$, are $\{\hat{s}_1,\ldots,\hat{s}_J\}$. Let $\widehat{I}_n$ be the estimated information matrix for both $\widehat{\beta}$ and $\{\hat{s}_1,\ldots,\hat{s}_J\}$. Then, the variance-covariance matrix for $(\widehat{\beta}, \widehat{A}[h_1],\ldots,\widehat{A}[h_m])$ is estimated as $V^T\widehat{I}_n^{-1}V$, where

$$V = \begin{pmatrix} I_{p\times p} & 0 \\ 0 & H \end{pmatrix}$$

and H is a $J \times m$ matrix with elements $\{h_k(a_j)\}_{1\le j\le J, 1\le k\le m}$.

3. Simulation Study

3.1. Simulation Set-Up

CPMs have been extensively simulated elsewhere to justify their use and have been largely seen to have good behavior [11]. Here we perform a more limited set of simulations to illustrate three major points which are particularly relevant to our study. First, the estimation of $A(y)$ using CPMs can be biased at extreme values of y. Even though $\widehat{A}(y)$ may have point-wise consistency for any y, $\widehat{A}(y)$ may not be uniformly consistent over all $y \in (-\infty,\infty)$. Second, in the modified approach, $\widehat{A}(y)$ is uniformly consistent over $y \in [L,U]$. Third, except for the estimation of extreme quantiles and $A(y)$ at extreme levels, the results are largely similar between CPMs fit to the original data and the modified data.

We roughly followed the simulation settings of Liu et al. [11]. Let $X_1 \sim$ Bernoulli(0.5), $X_2 \sim N(0,1)$, and $Y = \exp(\beta_1 X_1 + \beta_2 X_2 + \epsilon)$, where $\beta_1 = 1$, $\beta_2 = -0.5$, and $\epsilon \sim N(0,1)$. In this set-up, the correct transformation function is $A(y) = \log(y)$. We generated datasets $\{(X_1, X_2, Y)\}$ with sample sizes $n = 100$, 1000, and 5000. We fit CPMs that have the correctly

specified link function (probit) and model form (linear). The performance of misspecified models was extensively studied via simulations [11]. In CPMs, the transformation A and the parameters (β_1, β_2) are semi-parametrically estimated. We evaluated how well the transformation was estimated by comparing $\widehat{A}(y)$ with the correct transformation, $A(y) = \log(y)$, for various values of y.

We fit CPMs to the original data and CPMs to the modified data with (L, U) set to (e^{-4}, e^{4}), (e^{-2}, e^{2}), and $(e^{-1/2}, e^{1/2})$; these values correspond to approximately 0.2%, 13%, and 71% of Y being modified, respectively. All simulations had 1000 replications.

3.2. Simulation Results

Figure 1 shows the average estimate of $A(y)$ across 1000 simulation replicates compared with the true transformation, $\log(y)$. The left, center, and right panels are results based on sample sizes of 100, 1000, and 5000, respectively. With the original data, for all sample sizes, estimates are unbiased when y is around the center of its distribution (i.e., where the bulk of the probability mass lies), approximately in the range $[e^{-2}, e^{3}]$ when $n = 100$, in $[e^{-3}, e^{4}]$ when $n = 1000$, and in a wider range when $n = 5000$. However, at extreme values of y, we see biased estimation. This illustrates that, for a fixed y, one can find a sample size large enough that the estimation of $A(y)$ is unbiased, but that there will always be a more extreme value of y for which $\widehat{A}(y)$ may be biased. This motivates the need to categorize values outside a predetermined range (L, U) to achieve the uniform consistency of $\widehat{A}(y)$ for $y \in [L, U]$.

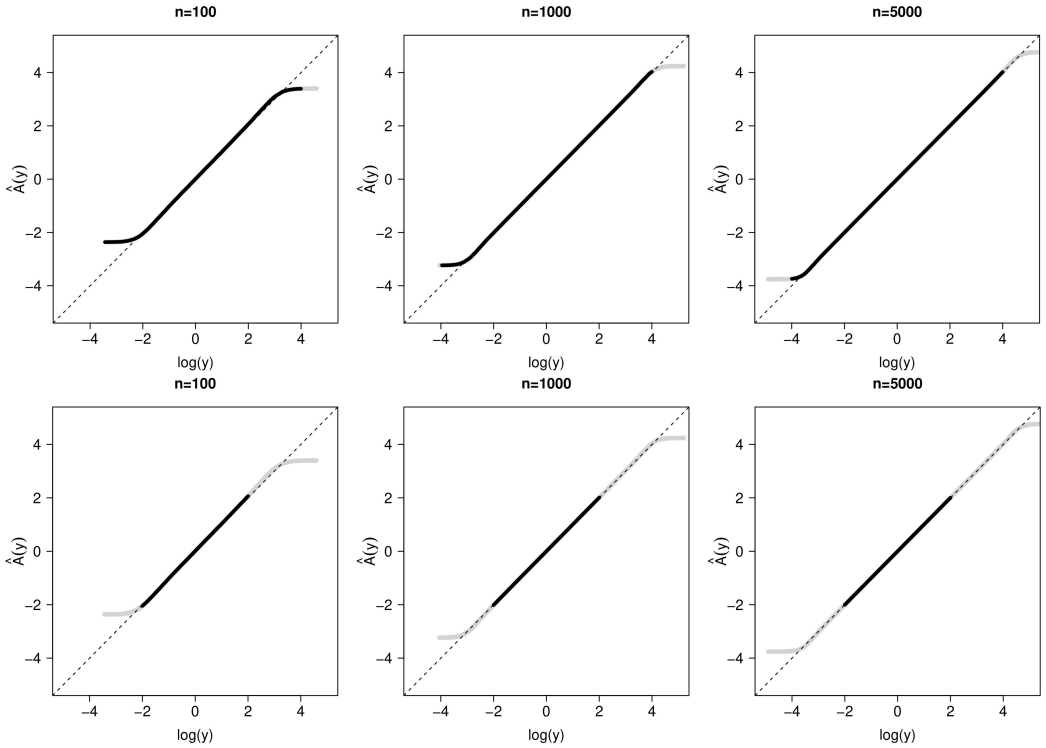

Figure 1. *Cont.*

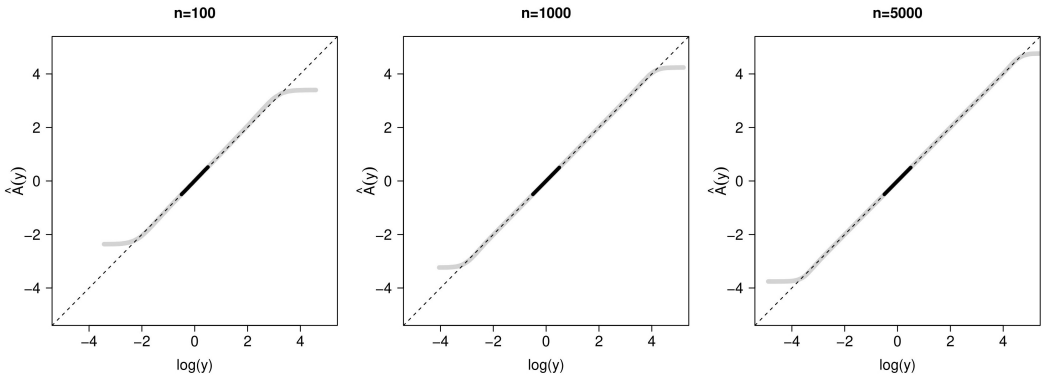

Figure 1. Average estimate of $A(y)$ after fitting properly specified CPMs compared with the true transformation, $\log(y)$. Gray curve: original data; black curve: modified data. Dashed lines are the diagonal. Top row: $(L, U) = (e^{-4}, e^{4})$; middle row: $(L, U) = (e^{-2}, e^{2})$; bottom row: $(L, U) = (e^{-1/2}, e^{1/2})$. Left to right: $n = 100, 1000, 5000$. Based on 1000 replications.

Figure 2 compares estimates of β_1 for the various sample sizes using the original data and using the modified data. As the sample size becomes larger, $\hat{\beta}_1$ becomes less biased in all approaches. At $n = 5000$, $\hat{\beta}_1$ is approximately unbiased even with a large proportion of the data having been categorized. Not surprisingly, with increasing proportions of categorized data, $\hat{\beta}_1$ becomes slightly more variable (Table 1) and slightly less correlated with that estimated from the original data. The results for β_2 have similar patterns (Supplementary Material Figure S1).

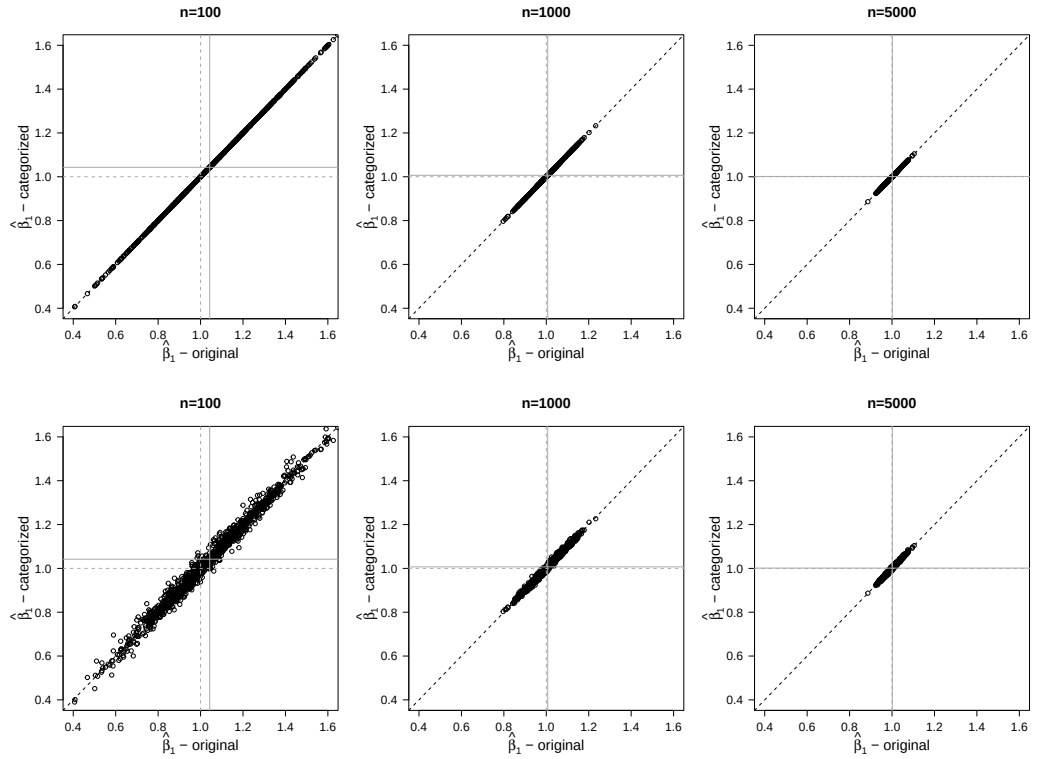

Figure 2. *Cont.*

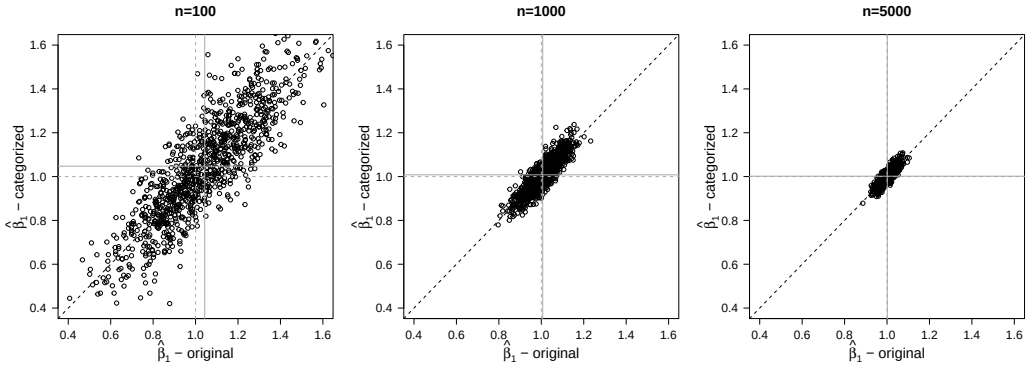

Figure 2. Estimates of β_1 using data categorized outside (L, U) compared with those using the original data and to the truth, $\beta_1 = 1$. Gray lines are mean estimates and dashed gray lines are the truth. Top row: $(L, U) = (e^{-4}, e^4)$; middle row: $(L, U) = (e^{-2}, e^2)$; bottom row: $(L, U) = (e^{-1/2}, e^{1/2})$. Left to right: $n = 100, 1000, 5000$. Based on 1000 replications.

Table 1 shows further results for five estimands: β_1, β_2, $A(e^{0.5})$, and the conditional median and mean of Y given $X_1 = 0$ and $X_2 = 0$. For each estimand, we compute the bias of the corresponding estimate, its standard deviation across replicates, the mean of estimated standard errors, and the mean squared error. For the estimands β_1, β_2, and $A(e^{0.5})$, estimation using the original data appears to be consistent, and the behavior of our estimators with the modified data is as predicted by the asymptotic theory. When $n = 100$, there appears to be only a modest amount of bias, even with 71% categorized; when $n = 1000$ (Table 1) and 5000 (shown in Supplementary Material Table S1), the bias is quite small. Although in Figure 1 we saw that estimates of $A(y)$ for extreme values of y were biased, we see no evidence that this impacts the estimation of β_1 and β_2. The average standard errors are very similar to the empirical results (i.e., the standard deviation of parameter estimates across replicates), suggesting that we are correctly estimating standard errors. These results hold regardless of the proportion categorized in our simulations. With increasing proportions being categorized, as expected, both absolute bias and standard deviation increase, and, as a result, the mean squared error increases. However, all these measures become smaller as the sample size increases.

Table 1. Simulation results for estimates from CPMs on original data and on data categorized outside (L, U); $n = 100, 1000$; based on 1000 replications.

n	Estimand		Original Data	Data Categorized Outside (L, U)		
				(e^{-4}, e^4)	(e^{-2}, e^2)	$(e^{-1/2}, e^{1/2})$
100	β_1	bias	0.043	0.043	0.042	0.048
		SD	0.228	0.228	0.229	0.260
		mean SE	0.217	0.217	0.219	0.251
		MSE	0.054	0.054	0.054	0.070
	β_2	bias	–0.022	–0.021	–0.020	–0.022
		SD	0.119	0.119	0.120	0.143
		mean SE	0.110	0.110	0.111	0.133
		MSE	0.015	0.015	0.015	0.021
	$A(e^{0.5})$	bias	0.019	0.019	0.019	0.020
		SD	0.177	0.177	0.177	0.183
		mean SE	0.174	0.174	0.175	0.182
		MSE	0.032	0.032	0.032	0.034
	median$(Y \mid X_1 = 0, X_2 = 0)$	bias	0.022	0.022	0.023	0.021
		SD	0.172	0.172	0.172	0.176

Table 1. *Cont.*

n	Estimand		Original Data	Data Categorized Outside (L, U) (e^{-4}, e^{4})	(e^{-2}, e^{2})	$(e^{-1/2}, e^{1/2})$
		MSE	0.030	0.030	0.030	0.031
	$E(Y \mid X_1 = 0, X_2 = 0)$	bias	–0.007	-	-	-
		SD	0.266	-	-	-
		mean SE	0.262	-	-	-
		MSE	0.071	-	-	-
1000	β_1	bias	0.007	0.007	0.007	0.008
		SD	0.068	0.068	0.068	0.076
		mean SE	0.067	0.067	0.068	0.077
		MSE	0.005	0.005	0.005	0.006
	β_2	bias	–0.001	–0.001	–0.001	–0.001
		SD	0.033	0.033	0.034	0.040
		mean SE	0.034	0.034	0.034	0.041
		MSE	0.001	0.001	0.001	0.002
	$A(e^{0.5})$	bias	0.003	0.003	0.003	0.003
		SD	0.055	0.055	0.055	0.056
		mean SE	0.054	0.054	0.054	0.057
		MSE	0.003	0.003	0.003	0.003
	median$(Y \mid X_1 = 0, X_2 = 0)$	bias	0.003	0.003	0.002	0.002
		SD	0.054	0.054	0.054	0.056
		MSE	0.003	0.003	0.003	0.003
	$E(Y \mid X_1 = 0, X_2 = 0)$	bias	–0.003	-	-	-
		SD	0.081	-	-	-
		mean SE	0.083	-	-	-
		MSE	0.007	-	-	-

SD, standard deviation of replicates; mean SE, average estimated standard error across replicates; MSE, mean squared error.

We cannot compute the standard error for the conditional median. Categorization also prohibits the sound estimation of the conditional mean; one could instead estimate the trimmed conditional mean, e.g., $E(Y \mid X_1 = 0, X_2 = 0, L \leq Y \leq U)$, which may substantially differ from $E(Y \mid X_1 = 0, X_2 = 0)$. The bias of $\widehat{A}(y)$ for extreme values of y had little impact on the estimation of $E(Y \mid X_1 = 0, X_2 = 0)$, which is computed using $\widehat{A}(y)$ over the entire range of observed y.

4. Example Data Analysis

CD4:CD8 ratio is a biomarker for measuring the strength of the immune system. A normal CD4:CD8 ratio is between 1 and 4, while people with HIV tend to have much lower values, and a low CD4:CD8 ratio is highly predictive of poor outcomes including non-communicable diseases and mortality. When people with HIV are put on antiretroviral therapy, their CD4:CD8 ratio tends to increase, albeit often slowly and quite variably. Castilho et al. [23] studied factors associated with the CD4:CD8 ratio among 2024 people with HIV who started antiretroviral therapy and maintained viral suppression for at least 12 months. They considered various factors including age, sex, race, the probable route of transmission, hepatitis C co-infection, hepatitis B co-infection, and the year of antiretroviral therapy initiation. Here we re-analyze their data using CPMs. We will focus on the associations of the CD4:CD8 ratio with age and sex, treating the other factors as covariates. The CD4:CD8 ratio tends to be right-skewed (Figure 3a), but there is no standard transformation for analyzing it. In various studies, it has been left untransformed [23], log-transformed [24], dichotomized (CD4:CD8 > 1 vs. ≤ 1) [25], put into ordered categories roughly based on quantiles [26], square-root transformed [27], and fifth-root transformed [28]. In contrast, CPMs do not require the specification of the transformation.

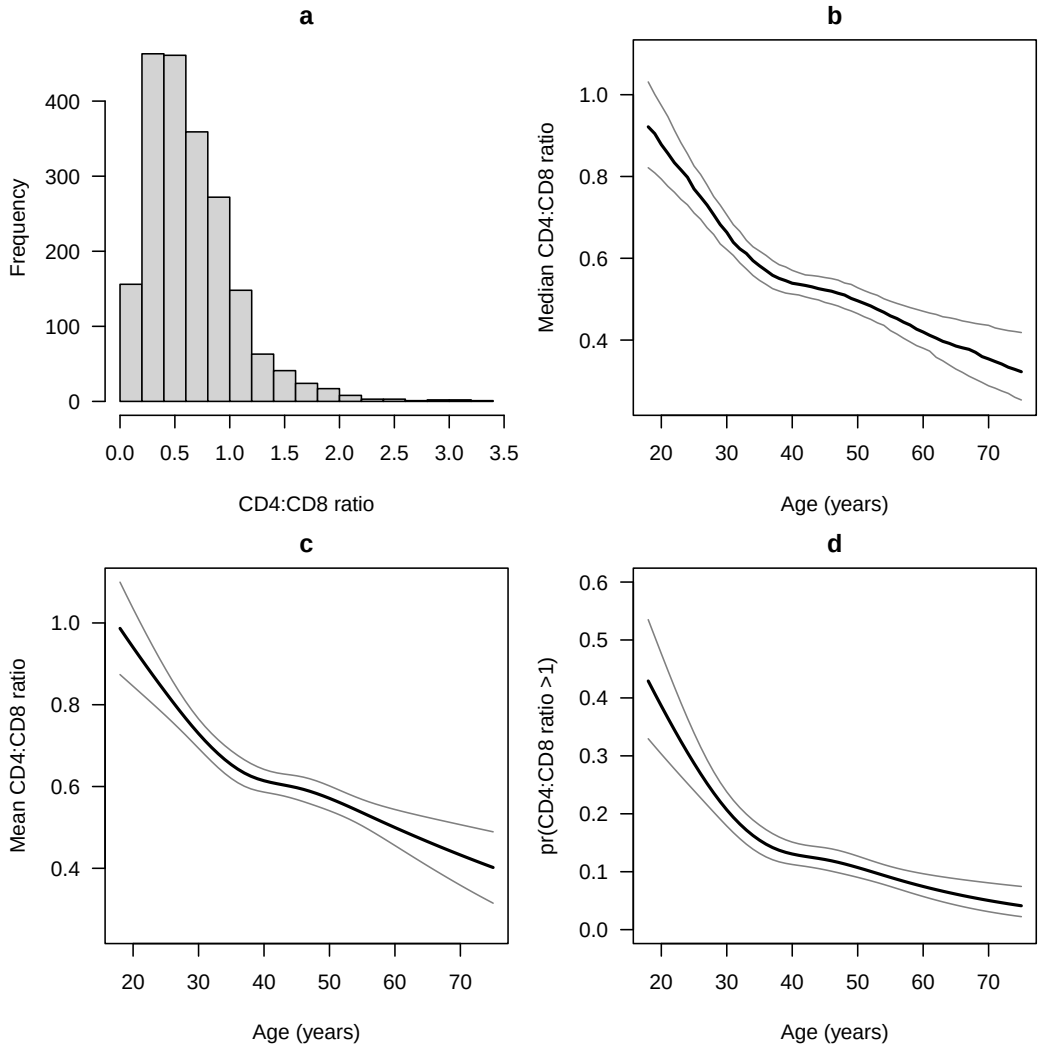

Figure 3. (**a**) Histogram of CD4:CD8 ratio in our dataset. (**b**–**d**) Estimated outcome measures and 95% confidence intervals as functions of age, holding other covariates constant at their medians/modes. (**b**) Median CD4:CD8 ratio; (**c**) mean CD4:CD8 ratio; (**d**) probability that CD4:CD8 > 1.

We fit three CPMs: Model 1 using the original data, Model 2 categorizing all CD4:CD8 ratios below $L = 0.1$ and above $U = 2.0$, and Model 3 categorizing below $L = 0.2$ and above $U = 1.5$. In a similar group of patients in a prior study [29], these values of L and U were approximately the 1.5th and 99.5th percentiles, respectively, for Model 2, and the 7th and 95th percentiles for Model 3. In our dataset, there were 19 (0.9%) CD4:CD8 ratios below 0.1 and 21 (1%) above 2.0, and 156 (7.7%) below 0.2 and 74 (3.7%) above 1.5. In our models, age was modeled using restricted cubic splines [30] with four knots at the 0.05, 0.35, 0.65, and 0.95 quantiles. All models were fit using a logit link function; quantile–quantile plots of probability-scale residuals [11] from the models suggested a good model fit (Supplementary Materials Figure S2).

All three models produced nearly identical results. Female sex had regression coefficients 0.6002, 0.6000, and 0.5994 in Models 1, 2, and 3, respectively (likelihood ratio $p < 0.0001$ in all models), suggesting that the odds of having a higher CD4:CD8 ratio, after controlling for all other variables in the model, were about $e^{0.6} = 1.82$ times higher for females than for males (95% Wald confidence interval 1.44–2.31). The median CD4:CD8 ratio holding all other covariates fixed at their medians/modes was estimated to be 0.67 (0.60–0.74) for females compared with 0.53 (0.51–0.56) for males; all models had the same estimates to two decimal places. The mean CD4:CD8 ratio holding all other covariates constant was estimated to be 0.73 (0.67–0.79) for females and 0.61 (0.58–0.63) for males

from Model 1. The mean estimates from Models 2 and 3 were slightly different (e.g., 0.72 for females); however, the mean should not be reported after categorization because the estimates arbitrarily assigned the categorized values to be L and U.

Older age was strongly associated with a lower CD4:CD8 ratio ($p < 0.0001$ in all models), and the association was non-linear ($p = 0.0080$, 0.0081, 0.0086, respectively). Figure 3b–d show the estimated median and mean CD4:CD8 ratio and the probability that CD4:CD8 > 1 as functions of age, all derived from the CPMs and holding other covariates fixed at their medians/modes. The median CD4:CD8 ratio and P(CD4:CD8 > 1) were not discernibly different between the three models. The mean as a function of age is only shown as derived from Model 1.

5. Discussion

We have now established the asymptotic properties for CPMs applied to data categorized at the tails. CPMs are flexible semiparametric regression models for continuous outcomes because the outcome transformation is nonparametrically estimated. We proved uniform consistency of the estimated coefficients $\widehat{\beta}$ and the estimated transformation function $\widehat{A}$ over the interval $[L, U]$, and showed that their joint asymptotic distribution is a tight Gaussian process. We demonstrated that these estimators perform well with simulations and illustrated their use in practice with a real data example.

Establishing uniform consistency requires a bounded range of the transformation function A, which is achieved by categorizing the outcome variable at both ends. Even if an outcome variable has a bounded support, the transformed values may not be bounded, and categorization will still be needed to establish uniform consistency. The proof of uniform consistency for $\widehat{\beta}$ also requires a bounded range of A even though β and A are separate components of the model.

Although the asymptotic properties for a similar nonparametric maximum likelihood approach in survival analysis have been established [8], the proofs here for CPMs with continuous data are different because we consider the nonparametric maximum likelihood estimate for the transformation in CPMs rather than the cumulative hazards function as in survival analysis. In addition, the transformation is estimated in the proofs directionally and separately for the two tails, which also differs from prior work.

For data without natural lower and upper bounds, the choice of L and U might be challenging in practice. In our CD4:CD8 ratio analysis, we were able to select values of L and U that corresponded with small and large CD4:CD8 percentiles in a prior study, therefore likely ensuring that a small fraction of the data would be modified in our analysis. In general, it is desirable to choose bounds so that only a small fraction of the data are categorized, although it should be reiterated that these bounds should be chosen prior to analysis. Both our simulations and our data example suggest that the results are robust to the specific choices of L and U as long as they do not lead to a high proportion of the data being categorized. For example, in our simulations, the results were nearly identical when categorization varied between 0.2% and 13%; in the data example, results were also nearly identical when categorization varied between 1.9% and 11.4%. Therefore, if one chooses to specify L and U, we suggest to select them so that approximately 5% or fewer of the observations would be modified at each end.

Our simulations and data example actually also suggest that without categorization, the estimators also perform well, which may support the use of CPMs with the original data in practice. CPMs applied to the original data do not require specifying L and U, and they permit the calculation of conditional means. However, its asymptotic theory has not been established; hence, there might be some risk to analyses using CPMs on the original data.

Continuous data that are skewed or subject to detection limits are common in applied research. Because of their ability to non-parametrically estimate a proper transformation, their robust rank-based nature, and their desirable properties proven and illustrated in this manuscript, CPMs are often an excellent choice for analyzing these types of data.

Extensions of CPMs to more complicated settings, e.g., clustered and longitudinal data, multivariate outcome data, or data with multiple detection limits, are warranted and are areas of ongoing research.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/math11244896/s1. Additional results from simulations and data example are in the Supplementary Materials. The code for simulations and data analysis is available at https://biostat.app.vumc.org/ArchivedAnalyses (accessed on 1 December 2023).

Author Contributions: Conceptualization, C.L. and B.E.S.; Methodology, D.Z.; Validation, Y.T.; Writing—original draft, C.L., D.Z. and B.E.S. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported in part by United States National Institutes of Health grants R01AI093234, P30AI110527, and K23AI20875.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: We thank Jessica Castilho and other Vanderbilt Comprehensive Care Clinic investigators for the use of the CD4:CD8 data.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proof of Theorem 1

Core steps of the proof: Let (β_0, A_0) be the true value of (β, A) and $(\widehat{\beta}, \widehat{A})$ be the NPMLE from the CPM applied to modified data. We will first prove that (I) $\widehat{A}$ is bounded in $[L, U]$ with probability one. Since $\widehat{A}(\cdot)$ is bounded and increasing in $[L, U]$, via the Helly selection theorem [20], for any subsequence, there exists a weakly convergent subsequence. Thus, without confusion, we assume that $\widehat{A} \to A^*$ weakly in $[L, U]$ and $\widehat{\beta} \to \beta^*$. We will then prove that (II) with probability one, $A^*(y) = A_0(y)$ for $y \in [L, U]$ and $\beta^* = \beta_0$. With this result, the consistency is established. Furthermore, since A_0 is continuously differentiable, we conclude that $\widehat{A}(y)$ converges to $A_0(y)$ uniformly in $[L, U]$ with probability one.

Proof. Proof of (I): Given a dataset of i.i.d. observations $\{(Y_i, Z_i)\}$, the nonparametric log-likelihood for CPM fitted to the modified data is

$$l_n(\beta, A) = \mathbf{P}_n\{I(Y \le L)\log G(A(L) - \beta^T Z) + I(Y \ge U)\log(1 - G(A(U-) - \beta^T Z)) + I(L < Y < U)\log(G(A(Y) - \beta^T Z) - G(A(Y-) - \beta^T Z))\Big\}.$$

Here, $\mathbf{P}_n$ denotes the empirical measure, i.e., $\mathbf{P}_n g(Y, Z) = n^{-1}\sum_{i=1}^n g(Y_i, Z_i)$, for any measurable function $g(Y, Z)$. Let $\widehat{A}\{Y_i\} \equiv \widehat{A}(Y_i) - \widehat{A}(Y_i-)$ be the jump size of $\widehat{A}$ at Y_i. Let $\widehat{A}(U) \equiv \widehat{A}(U-)$.

We first show that $\limsup \widehat{A}(U) < \infty$ a.s. If $\widehat{A}(Y_i-) \le M + m$ for all $Y_i > 0$, then $\widehat{A}(U) \le M + m$. Below, we assume there is a $Y_i > 0$ such that $\widehat{A}(Y_i-) > M + m$. Clearly, $\widehat{A}\{Y_i\}$ should be strictly positive, since, otherwise, $l_n(\beta, A) = -\infty$. Because of this, we differentiate $l_n(\beta, A)$ with respect to $A\{Y_i\}$ and then set it to zero to obtain the following equation:

$$\mathbf{P}_n\left\{I(Y \ge U)\frac{G'(\widehat{A}(U) - \widehat{\beta}^T Z)}{1 - G(\widehat{A}(U) - \widehat{\beta}^T Z)}\right\} - \mathbf{P}_n\left\{I(Y_i < Y < U)\frac{G'(\widehat{A}(Y) - \widehat{\beta}^T Z) - G'(\widehat{A}(Y-) - \widehat{\beta}^T Z)}{G(\widehat{A}(Y) - \widehat{\beta}^T Z) - G(\widehat{A}(Y-) - \widehat{\beta}^T Z)}\right\} = \frac{1}{n}\frac{G'(\widehat{A}(Y_i) - \widehat{\beta}^T Z_i)}{G(\widehat{A}(Y_i) - \widehat{\beta}^T Z_i) - G(\widehat{A}(Y_i-) - \widehat{\beta}^T Z_i)}. \quad \text{(A1)}$$

For any $Y > Y_i$, via condition 2,

$$\widehat{A}(Y-) - \widehat{\beta}^T Z \geq M + m - \widehat{\beta}^T Z \geq M.$$

According to condition 1, $G'(x)$ is decreasing when $x \geq M$. The left-hand side of (A1) is

$$\geq \mathbf{P}_n \left\{ I(Y > U) \frac{G'(\widehat{A}(U) - \widehat{\beta}^T Z)}{1 - G(\widehat{A}(U) - \widehat{\beta}^T Z)} \right\}.$$

For the right-hand side, we use the mean-value theorem on the denominator and then the decreasing property of $G'(x)$ when $x \geq M$ to obtain that

$$\frac{1}{n} \frac{G'(\widehat{A}(Y_i) - \widehat{\beta}^T Z_i)}{G(\widehat{A}(Y_i) - \widehat{\beta}^T Z_i) - G(\widehat{A}(Y_i-) - \widehat{\beta}^T Z_i)} = \frac{1}{n} \frac{G'(\widehat{A}(Y_i) - \widehat{\beta}^T Z_i)}{G'(\xi_i)\widehat{A}\{Y_i\}} \leq \frac{1}{n\widehat{A}\{Y_i\}},$$

where ξ_i is some value such that $\widehat{A}(Y_i-) - \widehat{\beta}^T Z_i \leq \xi_i \leq \widehat{A}(Y_i) - \widehat{\beta}^T Z_i$. Therefore, we have

$$\widehat{A}\{Y_i\} \leq \frac{1}{n} \left[\mathbf{P}_n \left\{ I(Y > U) \frac{G'(\widehat{A}(U) - \widehat{\beta}^T Z)}{1 - G(\widehat{A}(U) - \widehat{\beta}^T Z)} \right\} \right]^{-1},$$

and this holds for any Y_i between 0 and U and satisfying $\widehat{A}(Y_i-) > M + m$.

Let i_0 be the maximal index i for which $Y_i > 0$ and $\widehat{A}(Y_{i_0}-) \leq M + m$. We sum over all Y_i between 0 and U to obtain that

$$\begin{aligned} \widehat{A}(U) &= \widehat{A}(Y_{i_0}-) + \sum_{Y_i > 0, \widehat{A}(Y_i-) > M+m} \widehat{A}\{Y_i\} \\ &\leq M + m + \left[n^{-1} \sum_{i=1}^{n} I(0 < Y_i \leq U) \right] \left[\mathbf{P}_n \left\{ I(Y > U) \frac{G'(\widehat{A}(U) - \widehat{\beta}^T Z)}{1 - G(\widehat{A}(U) - \widehat{\beta}^T Z)} \right\} \right]^{-1}. \end{aligned}$$

We now show that $\widehat{A}(U)$ cannot diverge to ∞. Otherwise, suppose that $\widehat{A}(U) \to \infty$ for some subsequence. From the second half of condition 1, when n is large enough in the subsequence, for any Z,

$$\frac{G'(\widehat{A}(U) - \widehat{\beta}^T Z)}{1 - G(\widehat{A}(U) - \widehat{\beta}^T Z)} > \frac{1}{2} \liminf_{x \to \infty} \frac{G'(x)}{1 - G(x)} \equiv c_0 > 0,$$

and, therefore,

$$\widehat{A}(U) \leq M + m + \frac{n^{-1} \sum_{i=1}^{n} I(0 < Y_i \leq U)}{c_0 \mathbf{P}_n \{ I(Y > U) \}},$$

in which the last term converges to a constant. We thus have a contradiction. Hence, $\limsup \widehat{A}(U) < \infty$ with probability 1.

We can reverse the order of Y_i (change Y_i to $-Y_i$ so that the NPMLE is equivalent to maximizing the likelihood function, but instead of $A(y)$, we consider $-A(y)$). The same arguments as above apply to conclude that $\limsup -\widehat{A}(L) < \infty$ with probability 1, or equivalently, $\liminf \widehat{A}(L) > -\infty$ with probability 1. □

Proof. Proof of (II): We first show that $n\widehat{A}\{Y_i\}$ is bounded for all $Y_i \in [L, U]$. From the proof above, we know $\widehat{A}\{Y_i\} = O(n^{-1})$ uniformly in i for which $Y_i \in [L, U]$ satisfying $|\widehat{A}(Y_i-)| > M + m$. We prove that this is true for any Y_i. To do that, we define

$$\begin{aligned} H_n(y) &= \mathbf{P}_n\left\{ I(Y > U)\frac{G'(\widehat{A}(U) - \widehat{\beta}^T Z)}{1 - G(\widehat{A}(U) - \widehat{\beta}^T Z)} \right\} \\ &\quad -\mathbf{P}_n\left\{ I(y < Y \le U)\frac{G'(\widehat{A}(Y) - \widehat{\beta}^T Z) - G'(\widehat{A}(Y-) - \widehat{\beta}^T Z)}{G(\widehat{A}(Y) - \widehat{\beta}^T Z) - G(\widehat{A}(Y-) - \widehat{\beta}^T Z)} \right\}. \end{aligned}$$

First, we note that $H_n(y)$ has a total bounded variation in $[0, U]$. In fact, for any $0 < t < s < U$,

$$\begin{aligned} & |H_n(t) - H_n(s)| \\ = & \left|\mathbf{P}_n\left\{ I(t < Y \le s)\frac{G'(\widehat{A}(Y) - \widehat{\beta}^T Z) - G'(\widehat{A}(Y-) - \widehat{\beta}^T Z)}{G(\widehat{A}(Y) - \widehat{\beta}^T Z) - G(\widehat{A}(Y-) - \widehat{\beta}^T Z)} \right\}\right| \\ \le & c_1 \mathbf{P}_n\{I(t < Y \le s)\}, \end{aligned}$$

where $c_1 = \sup_{x\in[-m,c_0+m]} |G''(x)| / \inf_{x\in[-m,c_0+m]} |G'(x)|$. By choosing a subsequence, we assume that $H_n(y)$ converges weakly to $H^*(y)$. From the above inequality and taking limits, it is clear that

$$|H^*(t) - H^*(s)| \le c_1 P\ (t < Y \le s),$$

so $H^*(y)$ is Lipschitz-continuous in $y \in [0, U]$. The latter property ensures that $H_n(y)$ uniformly converges to $H^*(t)$ for $t \in [0, U]$.

According to Equation (A1), we know

$$|H_n(Y_i)| = \frac{1}{n}\frac{G'(\widehat{A}(Y_i) - \widehat{\beta}^T Z_i)}{G(\widehat{A}(Y_i) - \widehat{\beta}^T Z_i) - G(\widehat{A}(Y_i-) - \widehat{\beta}^T Z_i)} \ge \frac{c_2}{n\widehat{A}\{Y_i\}}, \tag{A2}$$

where $c_2 = \inf_{x\in[-m,c_0+m]G'(x)} / \sup_{x\in[-m,c_0+m]} G'(x)$. Thus,

$$\widehat{A}\{Y_i\} \ge \frac{c_2}{n}\frac{1}{|H_n(Y_i)| + \epsilon}$$

for any positive constant ϵ. This gives

$$\widehat{A}(U) \ge c_2 \mathbf{P}_n\left[\frac{I(Y \in [0, U])}{|H_n(Y)| + \epsilon}\right]. \tag{A3}$$

Since $H_n(Y)$ has a bounded total variation, $\{|H_n(Y)| + \epsilon\}^{-1}$ belongs to a Glivenko–Cantelli class bounded by $1/\epsilon$ and it converges in $L_2(P)$-norm to $\{|H^*(Y)| + \epsilon\}^{-1}$. As a result, the right-hand side of (A3) converges to $c_2 E[I(Y \in [0, U])(|H^*(Y)| + \epsilon)^{-1}]$, so we obtain that

$$c_0 \ge c_2 \int_0^U \frac{f_Y(y)}{|H^*(y)| + \epsilon} dy,$$

where $f_Y(y)$ is the marginal density of Y. Let ϵ decrease to zero; then, from the monotone convergence theorem, we conclude that

$$\int_0^U \frac{f_Y(y)}{|H^*(y)|} dy \le \frac{c_0}{c_2}. \tag{A4}$$

We use (A4) to show that $\min_{y\in[0,\tau]} |H^*(y)| > 0$. Otherwise, since $H^*(y)$ is continuous, there exists some $y_0 \in [0, \tau]$ such that $H^*(y_0) = 0$. However, since $H^*(y)$ is Lipschitz-continuous at y_0, the left-hand side of (A4) is at least larger than $\int_{y_0}^{y_0+\delta} \{c_1|y - y_0|\}^{-1} dy$ if

$y_0 < U$ or $\int_{y_0-\delta}^{y_0}\{c_1|y-y_0|\}^{-1}dy$ if $y_0 > 0$ for some small constant δ. The latter integrals are infinity. We obtain the contradiction. Hence, we conclude that $H^*(y)$ is uniformly bounded away from zero when $y \in [0, U]$. Thus, when n is large enough, $|\widehat{H}_n(Y_i)|$ is larger than a positive constant c_3 uniformly for all $Y_i > 0$. From (A1), we thus obtain that

$$c_3 \leq \frac{c_4}{n\widehat{A}\{Y_i\}},$$

where $c_4 = \sup_{x\in[-m,c_0+m]G'(x)}/\inf_{x\in[-m,c_0+m]} G'(x)$. In other words, $n\widehat{A}\{Y_i\} \leq c_4/c_3$. Using symmetric arguments, we can show that $n\widehat{A}\{Y_i\}$ is bounded by a constant for all $Y_i < 0$.

Finally, to establish the consistency in Theorem 1, since $\widehat{A}\{Y_i\}$ is of order n^{-1}, from Equation (A1), we obtain that

$$\widehat{A}\{Y_i\} = n^{-1}\left[\mathbf{P}_n\left\{I(Y > U)\frac{G'(\widehat{A}(U) - \widehat{\beta}^T Z)}{1 - G(\widehat{A}(U) - \widehat{\beta}^T Z)}\right\} - \mathbf{P}_n\left\{I(Y_i < Y \leq U)\frac{G''(\widehat{A}(Y) - \widehat{\beta}^T Z)}{G'(\widehat{A}(Y) - \widehat{\beta}^T Z)}\right\}\right]^{-1} + O(n^{-2}).$$

Following this expression, we define another step function, denoted by $\widetilde{A}(y)$, whose jump size at Y_i satisfies that

$$\widetilde{A}\{Y_i\} = n^{-1}\left[\mathbf{P}_n\left\{I(Y > U)\frac{G'(A_0(U) - \beta_0^T Z)}{1 - G(A_0(U) - \beta_0^T Z)}\right\} - \mathbf{P}_n\left\{I(Y_i < Y \leq U)\frac{G''(A_0(Y) - \beta_0^T Z)}{G'(A_0(Y) - \beta_0^T Z)}\right\}\right]^{-1}$$

so

$$\widetilde{A}(y) = n^{-1}\sum_{i=1}^{n} I(Y_i \leq y)\left[\mathbf{P}_n\left\{I(Y > U)\frac{G'(A_0(U) - \beta_0^T Z)}{1 - G(A_0(U) - \beta_0^T Z)}\right\} - \mathbf{P}_n\left\{I(Y_i < Y \leq U)\frac{G''(A_0(Y) - \beta_0^T Z)}{G'(A_0(Y) - \beta_0^T Z)}\right\}\right]^{-1}.$$

Via the strong law of large numbers and monotonicity of $\widehat{A}$, it is straightforward to show that $\widehat{A}(y)$ converges to

$$E\left\{I(Y \leq y)\left[\widetilde{\mathbf{P}}\left\{I(\widetilde{Y} > U)\frac{G'(A_0(U) - \beta_0^T \widetilde{Z})}{1 - G(A_0(U) - \beta_0^T \widetilde{Z})}\right\} - \widetilde{\mathbf{P}}\left\{I(Y < \widetilde{Y} \leq U)\frac{G''(A_0(\widetilde{Y}) - \beta_0^T \widetilde{Z})}{G'(A_0(\widetilde{Y}) - \beta_0^T \widetilde{Z})}\right\}\right]^{-1}\right\}$$

uniformly in $y \in [L, U]$. The limit can be verified to be the same as $A_0(y)$. Furthermore, we notice

$$\frac{\widehat{A}\{Y_i\}}{\widetilde{A}\{Y_i\}} = \frac{\mathbf{P}_n\left\{I(Y > U)\frac{G'(A_0(U)-\beta_0^T Z)}{1-G(A_0(U)-\beta_0^T Z)}\right\} - \mathbf{P}_n\left\{I(Y_i < Y \leq U)\frac{G''(A_0(Y)-\beta_0^T Z)}{G'(A_0(Y)-\beta_0^T Z)}\right\}}{\left[\mathbf{P}_n\left\{I(Y > U)\frac{G'(\widehat{A}(U)-\widehat{\beta}^T Z)}{1-G(\widehat{A}(U)-\widehat{\beta}^T Z)}\right\} - \mathbf{P}_n\left\{I(Y_i < Y \leq U)\frac{G''(\widehat{A}(Y)-\widehat{\beta}^T Z)}{G'(\widehat{A}(Y)-\widehat{\beta}^T Z)}\right\}\right] + O(n^{-1})}. \quad \text{(A5)}$$

Since $\widehat{A}(y)$ is bounded and increasing and $\widehat{\beta}^T Z$, $\widehat{A}(Y) - \widehat{\beta}^T Z$ belongs to a VC-hull of Donsker class. Via the preservation property under the monotone transformation,

$G^{(k)}(\widehat{A}(U) - \widehat{\beta}^T Z)$, $k = 0, 1, 2$, also belongs to a Donsker class. Therefore, the right-hand side of (A5) converges uniformly in Y_i to

$$g(Y_i) = \frac{\mathbf{P}\left\{ I(Y > U) \frac{G'(A_0(U) - \beta_0^T Z)}{1 - G(A_0(U) - \beta_0^T Z)} \right\} - \mathbf{P}\left\{ I(Y_i < Y \le U) \frac{G''(A_0(Y) - \beta_0^T Z)}{G'(A_0(Y) - \beta_0^T Z)} \right\}}{\mathbf{P}\left\{ I(Y > U) \frac{G'(A^*(U) - \beta^{*T} Z)}{1 - G(A^*(U) - \beta^{*T} Z)} \right\} - \mathbf{P}\left\{ I(Y_i < Y \le U) \frac{G''(A^*(Y) - \beta^{*T} Z)}{G'(A^*(Y) - \beta^{*T} Z)} \right\}}.$$

As a result, $A^*(y) = \int_0^y g(t) dA_0(t)$, or, equivalently, $dA^*(y)/dA_0(y) = g(y)$. Define

$$\widetilde{l}_n(\beta, A) = \mathbf{P}_n \Big\{ I(Y \le L) \log G(A(L) - \beta^T Z) + I(Y > U) \log(1 - G(A(U) - \beta^T Z))$$

$$+ I(L < Y \le U) \log(G(A(Y) - \beta^T Z) A\{Y\} \Big\}.$$

Since $\widetilde{A}\{Y_i\} = O(n^{-1})$ and $\widehat{A}\{Y_i\} = O(n^{-1})$,

$$l_n(\widehat{\beta}, \widehat{A}) = \widetilde{l}_n(\widehat{\beta}, \widehat{A}) + O(n^{-1}), \;\; l_n(\beta_0, \widetilde{A}) = \widetilde{l}_n(\beta_0, \widetilde{A}) + O(n^{-1}).$$

Since $l_n(\widehat{\beta}, \widehat{A}) \ge l_n(\beta_0, \widetilde{A}_0)$, we have

$$\widetilde{l}_n(\widehat{\beta}, \widehat{A}) \ge \widetilde{l}_n(\beta_0, \widetilde{A}) + O(n^{-1}).$$

That is,

$$\mathbf{P}_n \left\{ I(Y \le L) \log \frac{G(\widehat{A}(L) - \widehat{\beta}^T Z)}{G(\widetilde{A}(L) - \beta_0^T Z)} + I(Y > U) \log \frac{1 - G(\widehat{A}(U) - \widehat{\beta}^T Z)}{1 - G(\widetilde{A}(U) - \beta_0^T Z)} \right\}$$

$$+ \mathbf{P}_n \left\{ I(L < Y \le U) \log \frac{G(\widehat{A}(Y) - \widehat{\beta}^T Z)}{G(\widetilde{A}(Y) - \beta_0^T Z)} \right\} + n^{-1} \sum_{i=1}^{n} I(L < Y_i \le U) \frac{\widehat{A}\{Y_i\}}{\widetilde{A}\{Y_i\}} \ge O(n^{-1}).$$

We take limits on both sides. Using the Glivenko–Cantelli theorem with the first three two terms in the left-hand side and noting $\left| \widehat{A}\{Y_i\}/\widetilde{A}\{Y_i\} - g(Y_i) \right|$ converges to zero uniformly, we obtain that

$$\mathbf{P} \left\{ I(Y \le L) \log \frac{G(A^*(L) - \beta^{*T} Z)}{G(A_0(L) - \beta_0^T Z)} + I(Y > U) \log \frac{1 - G(A^*(U) - \beta^{*T} Z)}{1 - G(A_0(U) - \beta_0^T Z)} \right\}$$

$$+ \mathbf{P} \left\{ I(L < Y \le U) \log \frac{G(A^*(Y) - \beta^{*T} Z)}{G(A_0(Y) - \beta_0^T Z)} \right\} + \mathbf{P} \left\{ (L < Y \le U) \frac{dA^*(Y)}{dA_0(Y)} \right\} \ge 0.$$

The left-hand side is the negative Kullback–Leibler information for the density with parameter (β^*, A^*). Thus, the density function with parameter (β^*, A^*) should be the same as the true density. Immediately, we obtain that

$$G(A^*(Y) + \beta^{*T} Z) = G(A_0(Y) + \beta_0^T Z)$$

with probability one. From condition 2, we conclude that $\beta^* = \beta_0$ and $A^*(y) = A_0(y)$ for $y \in [L, U]$. □

Appendix B. Proof of Theorem 2

Proof. Let $BV[L, U]$ be the set of the functions over $[L, U]$ with $\|h\|_{TV} \le 1$, where $\|\cdot\|_{TV}$ denotes the total variation in $[L, U]$. For any $\nu \in R^p$ with $\|\nu\| \le 1$ and any $h \in BV[L, U]$,

we define the score function $\mathbf{\Psi}_n(\beta, A)[\nu, h]$ along the submodel for β with tangent direction ν and for A with the tangent function $\int_0^{\cdot} h(t)dA(t)$:

$$\begin{aligned}
& \mathbf{\Psi}_n(\beta, A)[\nu, h] \\
= & \lim_{\epsilon \to 0} \left(\frac{l_n(\beta + \epsilon\nu, A + \epsilon \int_0^{\cdot} h(t)dA(t)) - l_n(\beta, A)}{\epsilon} \right) \\
= & \mathbf{P}_n \Big\{ -F_1(Y, Z; \beta, A) Z^T \nu + F_2(Y, Z; \beta, A) Z^T \nu - F_3(Y, Z; \beta, A) Z^T \nu \Big\} \\
& + \mathbf{P}_n \Big\{ F_1(Y, Z; \beta, A) \int_0^L h(t)dA(t) - F_2(Y, Z; \beta, A) \int_0^U h(t)dA(t) \\
& \qquad + F_3(Y, Z; \beta, A) \int_0^Y h dA + F_4(Y, Z; \beta, A) h(Y) \Big\},
\end{aligned}$$

where

$$\begin{aligned}
F_1(Y, Z; \beta, A) &= \frac{I(Y \le L) G'(A(L) - \beta^T Z)}{G(A(L) - \beta^T Z)}, \\
F_2(Y, Z; \beta, A) &= \frac{I(Y \ge U) G'(A(U) - \beta^T Z)}{1 - G(A(U) - \beta^T Z)}, \\
F_3(Y, Z; \beta, A) &= \frac{I(L < Y < U) \big(G'(A(Y) - \beta^T Z) - G'(A(Y-) - \beta^T Z) \big)}{G(A(Y) - \beta^T Z) - G(A(Y-) - \beta^T Z)}, \\
F_4(Y, Z; \beta, A) &= \frac{G'(A(Y-) - \beta^T Z)}{G(A(Y) - \beta^T Z) - G(A(Y-) - \beta^T Z)} (A(Y) - A(Y-)).
\end{aligned}$$

Since $(\widehat{\beta}, \widehat{A})$ maximizes $l_n(\beta, A)$, we have, for any v and h,

$$\mathbf{\Psi}_n(\widehat{\beta}, \widehat{A})[\nu, h] = 0.$$

The rest of the proof contains the following main steps: we first show that $(\widehat{\beta}, \widehat{A})$ satisfies Equation (A6) (details below), then (A8), and finally (A10), from which the asymptotic distribution of $(\widehat{\beta}, \widehat{A})$ will be derived.

We know $\max_{L \le Y_i \le U} (\widehat{A}(Y_i) - \widehat{A}(Y_i-)) = O_p(n^{-1})$ from the proof in Appendix A. Thus, if we let

$$\widetilde{F}_3(Y, Z; \beta, A) = \frac{I(L < Y < U) G''(A(Y) - \beta^T Z)}{G'(A(Y) - \beta^T Z)},$$

then

$$F_3(Y, Z; \widehat{\beta}, \widehat{A}) = \widetilde{F}_3(Y, Z; \widehat{\beta}, \widehat{A}) + O_p(n^{-1})$$

and $F_4(Y, Z; \widehat{\beta}, \widehat{A}) = 1 + O_p(n^{-1})$ uniformly in (Y, Z). Consequently, we obtain that

$$\begin{aligned}
& \mathbf{P}_n \Big(-F_1(Y, Z; \widehat{\beta}, \widehat{A}) Z^T \nu + F_2(Y, Z; \widehat{\beta}, \widehat{A}) Z^T \nu - \widetilde{F}_3(Y, Z; \widehat{\beta}, \widehat{A}) Z^T \nu \Big) \\
& + \mathbf{P}_n \Big(F_1(Y, Z; \widehat{\beta}, \widehat{A}) \int_0^L h(t)d\widehat{A}(t) - F_2(Y, Z; \widehat{\beta}, \widehat{A}) \int_0^U h(t)d\widehat{A}(t) \\
& \qquad + \widetilde{F}_3(Y, Z; \widehat{\beta}, \widehat{A}) \int_0^Y h d\widehat{A} + h(Y) \Big) \\
= & \; O_p(n^{-1}),
\end{aligned}$$

and it holds uniformly in ν and h with $\|\nu\| \le 1$ and $\|h\|_{TV} \le 1$. Equivalently, we have that

$$\sqrt{n}(\mathbf{P}_n - \mathbf{P}) \begin{pmatrix} -F_1(Y, Z; \widehat{\beta}, \widehat{A}) Z^T \nu + F_2(Y, Z; \widehat{\beta}, \widehat{A}) Z^T \nu - \widetilde{F}_3(Y, Z; \widehat{\beta}, \widehat{A}) Z^T \nu \\ + F_1(Y, Z; \widehat{\beta}, \widehat{A}) \int_0^L h(t)d\widehat{A}(t) - F_2(Y, Z; \widehat{\beta}, \widehat{A}) \int_0^U h(t)d\widehat{A}(t) \\ + \widetilde{F}_3(Y, Z; \widehat{\beta}, \widehat{A}) \int_0^Y h d\widehat{A} + h(Y) \end{pmatrix}$$

$$= \sqrt{n}\mathbf{P}\begin{pmatrix} -F_1(Y,Z;\widehat{\beta},\widehat{A})Z^T\nu + F_2(Y,Z;\widehat{\beta},\widehat{A})Z^T\nu - \widetilde{F}_3(Y,Z;\widehat{\beta},\widehat{A})Z^T\nu \\ +F_1(Y,Z;\widehat{\beta},\widehat{A})\int_0^L h(t)d\widehat{A}(t) - F_2(Y,Z;\widehat{\beta},\widehat{A})\int_0^U h(t)d\widehat{A}(t) \\ +\widetilde{F}_3(Y,Z;\widehat{\beta},\widehat{A})\int_0^Y hd\widehat{A} + h(Y) \end{pmatrix} + O_p(n^{-1/2}). \quad \text{(A6)}$$

For the left-hand side of (A6), it is easy to see that for (β, A) in a neighborhood of (β_0, A_0) in the metric space $R^d \times BV[L,U]$, the classes of $F_1(Y,Z;\beta,A)$, $F_2(Y,Z;\beta,A)$ and $\widetilde{F}_3(Y,Z;\beta,A)$ are Lipschtiz classes of the P-Donsker classes $\{\beta^T Z\}$ and $\{A(Y) \in BV[L,U]\}$, so they are P-Donsker by preservation of the Donsker property. Additionally, the classes of $\left\{\int_0^Y hdA\right\}$, $\{Z^T\nu\}$, and $\{h(Y)\}$ are P-Donsker. As the result, since, via the consistency,

$$\begin{pmatrix} -F_1(Y,Z;\widehat{\beta},\widehat{A})Z^T\nu + F_2(Y,Z;\widehat{\beta},\widehat{A})Z^T\nu - \widetilde{F}_3(Y,Z;\widehat{\beta},\widehat{A})Z^T\nu \\ +F_1(Y,Z;\widehat{\beta},\widehat{A})\int_0^L h(t)d\widehat{A}(t) - F_2(Y,Z;\widehat{\beta},\widehat{A})\int_0^U h(t)d\widehat{A}(t) \\ +\widetilde{F}_3(Y,Z;\widehat{\beta},\widehat{A})\int_0^Y hd\widehat{A} + h(Y) \end{pmatrix}$$

converges in $L_2(P)$ to

$$\mathbf{S}(Y,Z)[\nu,h] \equiv \begin{pmatrix} -F_1(Y,Z;\beta_0,A_0)Z^T\nu + F_2(Y,Z;\beta_0,A_0)Z^T\nu - \widetilde{F}_3(Y,Z;\beta_0,A_0)Z^T\nu \\ +F_1(Y,Z;\beta_0,A_0)\int_0^L h(t)dA_0(t) - F_2(Y,Z;\beta_0,A_0)\int_0^U h(t)dA_0(t) \\ +\widetilde{F}_3(Y,Z;\beta_0,A_0)\int_0^Y hdA_0 + h(Y) \end{pmatrix},$$

Equation (A6) gives that

$$\begin{aligned} &\sqrt{n}(\mathbf{P}_n - \mathbf{P})\mathbf{S}(Y,Z)[\nu,h] \\ &= \sqrt{n}\mathbf{P}\begin{pmatrix} -F_1(Y,Z;\widehat{\beta},\widehat{A})Z^T\nu + F_2(Y,Z;\widehat{\beta},\widehat{A})Z^T\nu - \widetilde{F}_3(Y,Z;\widehat{\beta},\widehat{A})Z^T\nu \\ F_1(Y,Z;\widehat{\beta},\widehat{A})\int_0^L h(t)d\widehat{A}(t) - F_2(Y,Z;\widehat{\beta},\widehat{A})\int_0^U h(t)d\widehat{A}(t) \\ +\widetilde{F}_3(Y,Z;\widehat{\beta},\widehat{A})\int_0^Y hd\widehat{A} + h(Y) \end{pmatrix} + o_p(1). \quad \text{(A7)} \end{aligned}$$

On the other hand, we note that the first term in the right-hand side of (A7) is zero if replacing $(\widehat{\beta},\widehat{A})$ by (β_0, A_0). Thus, the right-hand side of (A7) is equal to

$$\sqrt{n}\mathbf{P}\begin{pmatrix} -F_1(Y,Z;\widehat{\beta},\widehat{A})Z^T\nu + F_2(Y,Z;\widehat{\beta},\widehat{A})Z^T\nu - \widetilde{F}_3(Y,Z;\widehat{\beta},\widehat{A})Z^T\nu \\ +F_1(Y,Z;\widehat{\beta},\widehat{A})\int_0^L h(t)d\widehat{A}(t) - F_2(Y,Z;\widehat{\beta},\widehat{A})\int_0^U h(t)d\widehat{A}(t) \\ +\widetilde{F}_3(Y,Z;\widehat{\beta},\widehat{A})\int_0^Y hd\widehat{A} \end{pmatrix}$$

$$-\sqrt{n}\mathbf{P}\begin{pmatrix} -F_1(Y,Z;\beta_0,A_0)Z^T\nu + F_2(Y,Z;\beta_0,A_0)Z^T\nu - \widetilde{F}_3(Y,Z;\beta_0,A_0)Z^T\nu \\ +F_1(Y,Z;\beta_0,A_0)\int_0^L h(t)dA_0(t) - F_2(Y,Z;\beta_0,A_0)\int_0^U h(t)dA_0(t) \\ +\widetilde{F}_3(Y,Z;\beta_0,A_0)\int_0^Y hdA_0 \end{pmatrix} + o_p(1).$$

We perform the linearization to the first two terms in the above expression. After some algebra, we obtain that this expression is equivalent to

$$\sqrt{n}(\mathbf{S}_{11}^T\nu + \mathbf{S}_{12}[h])^T(\widehat{\beta} - \beta_0) + \sqrt{n}\int(\mathbf{S}_{21}^T\nu + \mathbf{S}_{22}[h])d(\widehat{A} - A_0)(y)$$

$$+o_p(\sqrt{n}\|\widehat{\beta} - \beta_0\| + \sqrt{n}\|\widehat{A} - A_0\|_{TV}) + o_p(1),$$

where the operators $\mathbf{S}_{11} : R^d \to R^d$, $\mathbf{S}_{12} : BV[L,U] \to R^d$, $\mathbf{S}_{21}^T : R^d \to BV[L,U]$, and $\mathbf{S}_{22} : BV[L,U] \to BV[L,U]$ are defined as

$$
\begin{aligned}
\mathbf{S}_{11}v &= E\Big[\frac{d}{dt}\frac{G'(t)}{G(t)}\Big|_{t=A_0(L)-\beta_0^TZ} I(Y\le L)ZZ^Tv\Big] \\
&\quad -E\Big[\frac{d}{dt}\frac{G'(t)}{1-G(t)}\Big|_{t=A_0(U)-\beta_0^TZ} I(Y\ge U)ZZ^Tv\Big] \\
&\quad +E\Big[\frac{d}{dt}\frac{G''(t)}{G'(t)}\Big|_{t=A_0(Y)-\beta_0^TZ} I(L<Y<U)ZZ^Tv\Big], \\
\mathbf{S}_{12}[h] &= -E\Big[\frac{d}{dt}\frac{G'(t)}{G(t)}\Big|_{t=A_0(L)-\beta_0^TZ} I(Y\le L)Z\int_0^L h dA_0\Big] \\
&\quad +E\Big[\frac{d}{dt}\frac{G'(t)}{1-G(t)}\Big|_{t=A_0(U)-\beta_0^TZ} I(Y\ge U)Z\int_0^U h dA_0\Big] \\
&\quad -E\Big[\frac{d}{dt}\frac{G''(t)}{G'(t)}\Big|_{t=A_0(Y)-\beta_0^TZ} I(L<Y<U)Z\int_0^Y h dA_0\Big], \\
(\mathbf{S}_{21}^Tv)(y) &= -E\Big[\frac{d}{dt}\frac{G'(t)}{G(t)}\Big|_{t=A_0(L)-\beta_0^TZ} I(Y\le L)Z^TvI(Y>y)\Big] \\
&\quad +E\Big[\frac{d}{dt}\frac{G'(t)}{1-G(t)}\Big|_{t=A_0(U)-\beta_0^TZ} I(Y\ge U)Z^TvI(Y>y)\Big] \\
&\quad -E\Big[\frac{d}{dt}\frac{G''(t)}{G'(t)}\Big|_{t=A_0(Y)-\beta_0^TZ} I(L<Y<U)Z^TvI(Y>y)\Big], \\
\mathbf{S}_{22}[h](y) &= E\Big[\frac{d}{dt}\frac{G'(t)}{G(t)}\Big|_{t=A_0(L)-\beta_0^TZ} I(Y\le L)I(Y>y)\int_0^L h dA_0\Big] \\
&\quad -E\Big[\frac{d}{dt}\frac{G'(t)}{1-G(t)}\Big|_{t=A_0(U)-\beta_0^TZ} I(Y\ge U)I(Y>y)\int_0^U h dA_0\Big] \\
&\quad +E\Big[\frac{d}{dt}\frac{G''(t)}{G'(t)}\Big|_{t=A_0(Y)-\beta_0^TZ} I(L<Y<U)I(Y>y)\int_0^Y h dA_0\Big] \\
&\quad +E\{F_1(Y,Z;\beta_0,A_0)I(L\ge y)-F_2(Y,Z;\beta_0,A_0)I(U>y) \\
&\qquad +\widetilde{F}_3(Y,Z;\beta_0,A_0)I(Y\le y)\Big\}h(y).
\end{aligned}
$$

Combining the above results, we obtain from (A7) that

$$
\begin{aligned}
&\sqrt{n}(\mathbf{S}_{11}^T\nu+\mathbf{S}_{12}[h])^T(\widehat{\beta}-\beta_0)+\sqrt{n}\int(\mathbf{S}_{21}^T\nu+\mathbf{S}_{22}[h])d(\widehat{A}-A_0)(y) \\
&=\sqrt{n}(\mathbf{P}_n-\mathbf{P})\mathbf{S}(Y,Z)[\nu,h]+o_p(\sqrt{n}\|\widehat{\beta}-\beta_0\|+\sqrt{n}\|\widehat{A}-A_0\|_{TV})+o_p(1). \qquad \text{(A8)}
\end{aligned}
$$

Next, we show that the operator, $(\mathbf{S}_{11}^T\nu+\mathbf{S}_{12}[h],\mathbf{S}_{21}^T\nu+\mathbf{S}_{22}[h])$ that maps $(\nu,h)\in R^d\times BV[L,U]$ to $R^d\times BV[L,U]$, is invertible. This can be proven as follows: first, $\mathbf{S}_{11}^T\nu+\mathbf{S}_{12}[h]$ is finite-dimensional. Second, since the last term of $\mathbf{S}_{22}[h]$ is invertible and the other terms in $\mathbf{S}_{21}^T\nu+\mathbf{S}_{22}[h]$ map (ν,h) to a continuously differentiable function in $[L,U]$ which is a compact operator, $\mathbf{S}_{21}^T\nu+\mathbf{S}_{22}[h]$ is a Fredholm operator of the first kind. Therefore, to prove the invertibility, it suffices to show that $(\mathbf{S}_{11}^T\nu+\mathbf{S}_{12}[h],\mathbf{S}_{21}^T\nu+\mathbf{S}_{22}[h])$ is one-to-one. Suppose that $(\mathbf{S}_{11}^T\nu+\mathbf{S}_{12}[h],\mathbf{S}_{21}^T\nu+\mathbf{S}_{22}[h])=0$. Thus, we have

$$
(\mathbf{S}_{11}^T\nu+\mathbf{S}_{12}[h])^T\nu+\int(\mathbf{S}_{21}^T\nu+\mathbf{S}_{22}[h])dh(y)=0.
$$

From the previous derivation, we note that the left-hand side is in fact the negative Fisher information along the submodel $(\beta_0+\epsilon\nu,A_0+\int_0^{\cdot}h(t)dA_0(t))$. Thus, the score function along this submodel must almost certainly be zero. That is,

$$
-F_1(Y,Z;\beta_0,A_0)Z^T\nu+F_2(Y,Z;\beta_0,A_0)Z^T\nu-F_3(Y,Z;\beta_0,A_0)Z^T\nu
$$

$$+F_1(Y,Z;\beta_0,A_0)\int_0^L h(t)dA(t) - F_2(Y,Z;\beta_0,A_0)\int_0^U h(t)dA(t)$$
$$+\widetilde{F}_3(Y,Z;\beta_0,A_0)\int_0^Y hdA + h(Y) = 0$$

almost certainly. Consider any $Y = 0$; then, we have $Z^T\nu - h(0) = 0$, so $\nu = 0$ from condition 2. This further shows that $h(0) = 0$ and $h(y)$ satisfies

$$h(Y) + \widetilde{F}_3(Y,Z;\beta_0,A_0)\int_0^Y hdA = 0$$

for any $Y \in [L,U]$. This is a homogeneous integral equation and it is clear $h(y) = 0$ for any $y \in [L,U]$. We thus have established the invertibility of the operator $(\mathbf{S}_{11}^T\nu + \mathbf{S}_{12}[h], \mathbf{S}_{21}^T\nu + \mathbf{S}_{22}[h])$.

Therefore, from (A8), for any ν^* and h^*, by choosing (ν^-,h^-) as the inverse of the above operator applied to (ν^*,h^*), we obtain that

$$\begin{aligned}&\sqrt{n}\nu^{*T}(\widehat{\beta}-\beta_0) + \sqrt{n}\int h^*(y)d(\widehat{A}-A_0)(y)\\ &= \sqrt{n}(\mathbf{P}_n-\mathbf{P})\mathbf{S}(Y,Z)[\nu^-,h^-] + o_p(\sqrt{n}\|\widehat{\beta}-\beta_0\| + \sqrt{n}\|\widehat{A}-A_0\|_{TV}) + o_p(1),\end{aligned} \tag{A9}$$

and this holds uniformly for $\|\nu^*\| \le 1$ and $\|h^*\|_{TV} \le 1$. Using (A9), we obtain that

$$\sqrt{n}\|\widehat{\beta}-\beta_0\| + \sqrt{n}\|\widehat{A}-A_0\|_{TV} = O_p(1).$$

Thus,

$$\begin{aligned}&\sqrt{n}\nu^{*T}(\widehat{\beta}-\beta_0) + \sqrt{n}\int h^*(y)d(\widehat{A}-A_0)(y)\\ &= \sqrt{n}(\mathbf{P}_n-\mathbf{P})\mathbf{S}(Y,Z)[\nu^-,h^-] + o_p(1).\end{aligned} \tag{A10}$$

This implies that

$$\sqrt{n}(\widehat{\beta}-\beta_0, \widehat{A}-A_0),$$

as a random map on (ν^*,h^*), converges weakly to a mean-zero and tight Gaussian process. Furthermore, by letting $\nu^* = 1$ and $h^* = 0$, we conclude that $\sqrt{n}(\widehat{\beta}-\beta_0)$ has an influence function given by $\mathbf{S}(Y,Z)[\nu^-,h^-]$. Since the latter lies on the score space, it must be the efficient influence function. Hence, the asymptotic variance of $\sqrt{n}(\widehat{\beta}-\beta_0)$ achieves the semiparametric efficiency bound. □

References

1. Box, G.E.P.; Cox, D.R. An analysis of transformations (with Discussion). *J. R. Stat. Soc. Ser. B* **1964**, *26*, 211–252.
2. Doksum, K.A. An extension of partial likelihood methods for proportional hazard models to general transformation models. *Ann. Statist.* **1987**, *15*, 325–345. [CrossRef]
3. Cuzick, J. Rank regression. *Ann. Statist.* **1988**, *16*, 1369–1389. [CrossRef]
4. Pettitt, A.N. Inference for the linear model using a likelihood based on ranks. *J. R. Statist. Soc. Ser. B* **1982**, *44*, 234–243. [CrossRef]
5. Kalbfleisch, J.D.; Prentice, R.L. Marginal likelihoods based on Cox's regression and life model. *Biometrika* **1973**, *60*, 267–278. [CrossRef]
6. Cheng, S.C.; Wei, L.J.; Ying, Z. Analysis of transformation models with censored data. *Biometrika* **1995**, *82*, 835–845. [CrossRef]
7. Chen, K.; Jin, Z.; Ying, Z. Semiparametric analysis of transformation models with censored data. *Biometrika* **2002**, *89*, 659–668. [CrossRef]
8. Zeng, D.; Lin, D.Y. Maximum likelihood estimation in semiparametric regression models with censored data (with Discussion). *J. R. Statist. Soc. Ser. B* **2007**, *69*, 507–564. [CrossRef]
9. Manuguerra, M.; Heller, G.Z. Ordinal regression models for continuous scales. *Int. J. Biostat.* **2010**, *6*, 14. [CrossRef]
10. Hothorn, T.; Möst, L.; Bühlmann, P. Most likely transformations. *Scand. J. Stat.* **2018**, *45*, 110–134. [CrossRef]
11. Liu, Q.; Shepherd, B.E.; Li, C.; Harrell, F.E. Modeling continuous response variables using ordinal regression. *Stat. Med.* **2017**, *36*, 4316–4335. [CrossRef] [PubMed]
12. Spertus, J.A.; Jones, P.G.; Maron, D.J.; O'Brien, S.M.; Reynolds, H.R.; Rosenberg, Y.; Stone, G.W.; Harrell, F.E.; Boden, W.E.; Weintraub, W.S.; et al. Health-status outcomes with invasive or conservative care in coronary disease. *N. Engl. J. Med.* **2020**, *382*, 1408–1419. [CrossRef] [PubMed]

13. Pun, B.T.; Badenes, R.; La Calle, G.H.; Orun, O.M.; Chen, W.; Raman, R.; Simpson, B.-G.K.; Wilson-Linville, S.; Olmedillo, B.H.; de la Cueva, A.V.; et al. Prevalence and risk factors for delirium in critically ill patients with COVID-19 (COVID-D): A multicentre cohort study. *Lancet* **2021**, *9*, 239–250. [CrossRef] [PubMed]
14. Pasquali, S.K.; Thibault, D.; O'Brien, S.M.; Jacobs, J.P.; Gaynor, J.W.; Romano, J.C.; Gaies, M.; Hill, K.D.; Jacobs, M.L.; Shahian, D.M.; et al. National variation in congenital heart surgery outcomes. *Circulation* **2020**, *142*, 1351–1360. [CrossRef] [PubMed]
15. Williams, Z.J.; Failla, M.D.; Davis, S.L.; Heflin, B.H.; Okitondo, C.D.; Moore, D.J.; Cascio, C.J. Thermal perceptual thresholds are typical in autism spectrum disorder but strongly related to intra-individual response variability. *Sci. Rep.* **2019**, *9*, 12595. [CrossRef] [PubMed]
16. Hatch, L.D.; Scott, T.A.; Slaughter, J.C.; Xu, M.; Smith, A.H.; Stark, A.R.; Patrick, S.W.; Ely, E.W. Outcomes, resource use, and financial costs of unplanned extubations in preterm infants. *Pediatrics* **2020**, *145*, e20192819. [CrossRef]
17. Wang, J.-H.; Wong, R.C.B.; Liu, G.-S. Retinal transcriptome and cellular landscape in relation to the progression of diabetic retinopathy. *Investig. Ophthalmol. Vis. Sci.* **2022**, *63*, 26. [CrossRef]
18. Ioannidis, J.P.A.; Kim, B.Y.S.; Trounson, A. How to design preclinical studies in nanomedicine and cell therapy to maximize the prospects of clinical translation. *Nat. Biomed. Eng.* **2018**, *2*, 797–809. [CrossRef]
19. French, B.; Shotwell, M.S. Regression models for ordinal outcomes. *JAMA* **2022**, *328*, 772–773. [CrossRef]
20. Billingsley, P. *Probability and Measure*, 3rd ed.; Wiley: New York, NY, USA, 1995.
21. Van der Vaart, A.W.; Wellner, J.A. *Weak Convergence and Empirical Processes*; Springer: New York, NY, USA, 1996.
22. Murphy, S.A.; van der Vaart, A.W. On profile likelihood. *J. Am. Stat. Assoc.* **2000**, *95*, 449–465. [CrossRef]
23. Castilho, J.L.; Shepherd, B.E.; Koethe, J.R.; Turner, M.; Bebawy, S.; Logan, J.; Rogers, W.B.; Raffanti, S.; Sterling, T.R. CD4/CD8 ratio, age, and risk of serious non-communicable diseases in HIV-infected adults on antiretroviral therapy. *AIDS* **2016**, *30*, 899–908. [CrossRef]
24. Sauter, R.; Huang, R.; Ledergerber, B.; Battegay, M.; Bernasconi, E.; Cavassini, M.; Furrer, H.; Hoffman, M.; Rougemont, M.; Günthard, H.F.; et al. CD4/CD8 ratio and CD8 counts predict CD4 response in HIV-1-infected drug naive and in patients on cART. *Medicine* **2016**, *95*, e5094. [CrossRef]
25. Petoumenos, K.; Choi, J.Y.; Hoy, J.; Kiertiburanakul, S.; Ng, O.T.; Boyd, M.; Rajasuriar, R.; Law, M. CD4:CD8 ratio comparison between cohorts of HIV-positive Asians and Caucasians upon commencement of antiretroviral therapy. *Antivir. Ther.* **2017**, *22*, 659–668.
26. Serrano-Villar, S.; Sainz, T.; Lee, S.A.; Hunt, P.W.; Sinclair, E.; Shacklett, B.L.; Ferre, A.L.; Hayes, T.L.; Somsouk, M.; Hsue, P.Y.; et al. HIV-infected individuals with low CD4/CD8 ratio despite effective antiretroviral therapy exhibit altered T cell subsets, heightened CD8+ T cell activation, and increased risk of non-AIDS morbidity and mortality. *PLoS Pathog.* **2014**, *10*, e1004078. [CrossRef]
27. Silva, C.; Peder, L.; Silva, E.; Previdelli, I.; Pereira, O.; Teixeira, J.; Bertolini, D. Impact of HBV and HCV coinfection on CD4 cells among HIV-infected patients: A longitudinal retrospective study. *J. Infect. Dev. Ctries.* **2018**, *12*, 1009–1018. [CrossRef]
28. Gras, L.; May, M.; Ryder, L.P.; Trickey, A.; Helleberg, M.; Obel, N.; Thiebaut, R.; Guest, J.; Gill, J.; Crane, H.; et al. Determinants of restoration of CD4 and CD8 cell counts and their ratio in HIV-1-positive individuals with sustained virological suppression on antiretroviral therapy. *J. Acquir. Immune Defic. Syndr.* **2019**, *80*, 292–300. [CrossRef]
29. Serrano-Villar, S.; Perez-Elias, M.J.; Dronda, F.; Casado, J.L.; Moreno, A.; Royuela, A.; Perez-Molina, J.A.; Sainz, T.; Navas, E.; Hermida, J.M.; et al. Increased risk of serious non-AIDS-related events in HIV-infected subjects on antiretroviral therapy associated with a low CD4/CD8 ratio. *PLoS ONE* **2014**, *9*, e85798. [CrossRef]
30. Harrell, F.E., Jr. *Regression Modeling Strategies*, 2nd ed.; Springer: Cham, Switzerland; Berlin/Heidelberg, Germany; New York, NY, USA; Dordrecht, The Netherlands; London, UK, 2015.