Supplementary Material for "Simple Semiparametric Estimation of Ordered Response Models"

Ruixuan Liu[1] and Zhengfei Yu[2]

[1]Department of Decision Sciences and Managerial Economics, Chinese University of Hong Kong

[2]Faculty of Humanities and Social Sciences, University of Tsukuba

S1 Auxiliary Results

S1.1 Asymptotic Variance of Two-Stage Rank Estimator

This section states the asymptotic variance of the two-stage rank estimator (described in Section 2.4.1) and compares it to our two-stage isotonic estimator. It complements Remark 3.2. Recall that $X = (X_1, X_{-1}')'$ and denote $\beta_0 = (1, \beta_{0-}')'$ (the first element of β_0 is normalized to 1). Let $\hat{\beta}_{R-}$ and $\hat{\alpha}_R$ be the two-stage rank estimators for the coefficients β_{0-} and threshold α_0 of the ordered response model (1.1). In addition, let g_0 denote the density function of $X'\beta_0$, and F_0 and f_0 denote the distribution and density functions of the error term ε. The asymptotic variances of $\hat{\beta}_{R-}$ and $\hat{\alpha}_R$ are given in Theorem S1.1, whose proof is presented at the end of this section.

Condition S1. The density of X_1, conditional on X_{-1} and Y, has bounded derivatives up to the third order.

Condition S1, adapted from Sherman (1993) [p.137], is sufficient for Assumption 7(i) to (iv) of Cavanagh and Sherman (1998).

Theorem S1.1. *Under Conditions 1 to 9 and Condition S1, we have*

$$\begin{aligned} \sqrt{n}(\hat{\beta}_{R-} - \beta_{0-}) &\Rightarrow \mathbb{N}\left(0, H_{\beta_0,R}^{-1} J_{\beta_0,R} H_{\beta_0,R}^{-1}\right), \\ \sqrt{n}\left(\hat{\alpha}_R - \alpha_0\right) &\Rightarrow \mathbb{N}\left(0, V_{\alpha_0,R}^{-1} J_{\alpha_0,R} V_{\alpha_0,R}^{-1}\right), \end{aligned}$$

where

$$\begin{aligned}
H_{\beta_0,R} &= -\mathbb{E}\left[\{X_{-1}-\mathbb{E}[X_{-1}|X'\beta_0]\}^{\otimes 2}\, g_0(X'\beta_0)\left(f_0(X'\beta_0)+f_0(X'\beta_0+\alpha_0)\right)\right],\\
J_{\beta_0,R} &= \mathbb{E}\left[\{X_{-1}-\mathbb{E}[X_{-1}|X'\beta_0]\}^{\otimes 2}\, g_0(X'\beta_0)^2\left(Y+F_0(X'\beta_0)+F_0(X'\beta_0+\alpha_0)-3\right)^2\right],\\
V_{\alpha_0,R} &= -\mathbb{E}\left[f_0(X'\beta_0)g_0(X'\beta_0-\alpha_0)+f(X'\beta_0+\alpha_0)g_0(X'\beta_0+\alpha_0)\right],\\
J_{\alpha_0,R} &= \mathbb{E}\left[\left[\left(F_0(X'\beta_0)-\Delta_1\right)g_0(X'\beta_0-\alpha_0)+\left(1-\Delta_3-F_0(X'\beta_0+\alpha_0)\right)g_0(X'\beta_0+\alpha_0)\right]^2\right].
\end{aligned}$$

Despite that the asymptotic variance formulae of the isotonic two-stage estimator (Theorem 3.1) and the two-stage rank estimator (Theorem S1.1) have a few comparable terms, there is no definite ranking between them. We conjecture that the isotonic two-stage estimator may have smaller asymptotic variances in some circumstances as it is likelihood-based, and also uses the information of the distribution function $F(;,\beta)$ and moment conditions.

We present a numerical example to illustrate the relationship between the asymptotic variances of the isotonic and rank estimators. Let β_{0-} contain two elements $\beta_{0-}=(\beta_{02},\beta_{03})'$ and $X\sim\mathbb{N}(0,I_3)$ so that $X_{-1}\sim\mathbb{N}(0,I_2)$ and $K=3$. Define the linear index $U=X'\beta_0$. It follows that $U\sim N(0,\sigma_U^2)$, where $\sigma_U^2=\beta_0'\beta_0$. We observe that the conditional distribution of covariates X given the linear index $U=X'\beta_0$ is

$$X_{-1}|U\sim\mathbb{N}(\beta_{0-}U/\sigma_U^2, I_{(K-1)\times(K-1)}-\beta_{0-}\beta_{0-}'/\sigma_U^2).$$

Therefore,

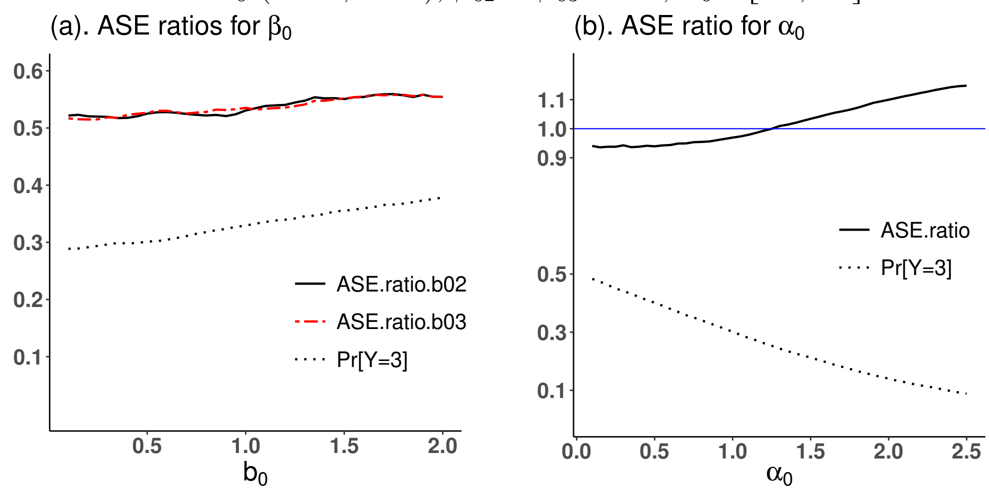
$$X_{-1}-\mathbb{E}[X_{-1}|U]=X_{-1}-\beta_{0-}U/\sigma_U^2,\quad\text{and}\quad \mathbb{E}\left[(X_{-1}-\mathbb{E}[X_{-1}|U])^{\otimes 2}|U\right]=I_2-\beta_{0-}\beta_{0-}'/\sigma_U^2.$$

The Hessian matrices H_{β_0} (in Theorem 3.1) and $H_{\beta_0,R}$ (in Theorem S1.1) simplify to

$$\begin{aligned}
H_{\beta_0} &= \mathbb{E}[f_0(U)]\times[I_2-\beta_{0-}\beta_{0-}'/\sigma_U^2],\\
H_{\beta_0,R} &= \mathbb{E}\left[\left(f_0(U)+f_0(U+\alpha_0)\right)g_0(U)\right]\times[I_2-\beta_{0-}\beta_{0-}'/\sigma_U^2].
\end{aligned}$$

For this example, Figure S1.1 plots the ratio of asymptotic standard errors (ASE) between the isotonic and the rank estimators for β_{02}, β_{03}, and α_0, when F_0 and f_0 are the normal CDF and PDF with the mean equal to 0 and the variance equal to 2. Panel (a) depicts the ratio ASE($\hat\beta_2$)/ASE($\hat\beta_{2,R}$) (black solid line) and ASE($\hat\beta_3$)/ASE($\hat\beta_{3,R}$) (red dash-dotted line) as the true value $\beta_{02}=\beta_{03}=b_0$ is running over $[0.1,2]$. Panel (b) plots the ratio ASE($\hat\alpha$)/ASE($\hat\alpha_R$) (black solid line) as the true value α_0 is running over $[0.1,2.5]$ and $\beta_{02}=\beta_{03}=0.5$. Both figures also include a curve representing $\Pr[Y=3]$ (black dotted line) to show that values of β_0 and α_0 are varying within a reasonable range.[1]

Figure S1: Asymptotic standard error (ASE) ratios between the isotonic and the rank estimators, the error term $\varepsilon \sim N(0,2)$, the black dotted line denotes $\Pr[Y=3]$. Panel (a): ASE ratios for β_{02} (black, solid) and β_{03} (red, dashed), $\beta_{02} = \beta_{03} = b_0 \in [0.1, 2]$, $\alpha_0 = 1$. Panel (b): ASE ratio for α_0 (black, solid), $\beta_{02} = \beta_{03} = 0.5$, $\alpha_0 \in [0.1, 2.5]$.



Panel (a) shows that the ASEs of the isotonic estimator for the coefficients β_{01} and β_{02} are about one half of those of the rank estimator, with the ratio varying from 0.52 to 0.56. Panel (b), on the other hand, shows that the ASE for the threshold α_0 is comparable between these two estimators, as the ratio varies from 0.94 to 1.14. Overall, in this numerical example, the isotonic two-stage estimator tends to have a smaller asymptotic variance for estimating β_0 while the relative magnitude of two variances depends more on the case for the estimation of α_0. From a heuristic perspective, the first stage of the rank estimator (Han, 1987; Cavanagh and Sherman, 1998) only uses the monotonicity of $\mathbb{E}[Y|X]$ with respect to $X'\beta_0$. Its second stage, on the other hand, uses more structural information from the ordered response, because it compares conditional probabilities across different categories.

Proof of Theorem S1.1. A quick inspection confirms that all the regularity conditions from Cavanagh and Sherman (1998) and Chen (2002) are imposed. Therefore, our main task is to derive explicit forms for the rank scores and Hessian matrices. Consider the first-stage rank estimator $\hat{\beta}_{R-}$. Applying Theorems 2 and 3 of Cavanagh and Sherman (1998) with M being the identity function leads to

[1] Note that in this example, $\Pr[Y = 1] = 0.5$ regardless of the values of β_0 and α_0.

$$\sqrt{n}(\hat{\beta}_{R-} - \beta_{0-}) \quad \Rightarrow \quad \mathbb{N}\left(0, H^{-1}_{\beta_0,R} J_{\beta_0,R} H^{-1}_{\beta_0,R}\right),$$

where

$$
\begin{aligned}
H_{\beta_0,R} &= \mathbb{E}\left[\{X_{-1} - \mathbb{E}[X_{-1}|X'\beta_0]\}^{\otimes 2} g_0(X'\beta_0)\rho'(X'\beta_0)\right], && \text{(S.1)}\\
J_{\beta_0,R} &= \mathbb{E}\left[\{X_{-1} - \mathbb{E}[X|X'\beta_0]\}^{\otimes 2} g_0(X'\beta_0)^2 S(Y, X'\beta_0)^2\right], && \text{(S.2)}\\
S(y,t) &= Y - \mathbb{E}[Y|X'\beta_0 = t], && \text{(S.3)}\\
\rho(t) &= \mathbb{E}[Y|X'\beta_0 = t]. && \text{(S.4)}
\end{aligned}
$$

In the ordered response model (1.1),

$$Y = \mathbb{I}\{-X_i'\beta_0 + \varepsilon_i \leqslant 0\} + 2\mathbb{I}\{0 < -X_i'\beta_0 + \varepsilon_i \leqslant \alpha_0\} + 3\mathbb{I}\{-X_i'\beta_0 + \varepsilon_i \geqslant \alpha_0\},$$

and hence

$$
\begin{aligned}
S(Y, X'\beta_0) &= Y + F_0(X'\beta_0) + F_0(X'\beta_0 + \alpha_0) - 3,\\
\rho'(X'\beta_0) &= -f_0(X'\beta_0) - f_0(X'\beta_0 + \alpha_0).
\end{aligned}
$$

Then we consider the second stage rank estimator $\hat{\alpha}_R$. Theorem 1 of Chen (2002) gives

$$\sqrt{n}\left(\hat{\alpha}_R - \alpha_0\right) \quad \Rightarrow \quad \mathbb{N}\left(0, V^{-1}_{\alpha_0,R} J_{\alpha_0,R} V^{-1}_{\alpha_0,R}\right),$$

where

$$V_{\alpha_0,R} = \mathbb{E}\left[\frac{\partial^2 \tau(Y, X, \beta_0, \alpha_0)}{\partial \alpha^2}\right], J_{\alpha_0,R} = \mathbb{E}\left[\left(\frac{\partial \tau(Y, X, \beta_0, \alpha_0)}{\partial \alpha}\right)^2\right], \qquad \text{(S.5)}$$

and

$$
\begin{aligned}
\tau(y, x, \alpha, \beta) \equiv\ & \mathbb{E}\left[\left(\mathbb{I}\{y = 1\} - \mathbb{I}\{Y \leqslant 2\}\right)\mathbb{I}\{x'\beta - X'\beta \geqslant \alpha\}\right]\\
& + \mathbb{E}\left[\left(\mathbb{I}\{Y = 1\} - \mathbb{I}\{y \leqslant 2\}\right)\mathbb{I}\{X'\beta - x'\beta \geqslant \alpha\}\right]. \qquad \text{(S.6)}
\end{aligned}
$$

In the following, we provide an explicit expression for $\tau(y, x, \alpha, \beta)$ by computing the conditional expectation given X and then applying the law of iterated expectation. The first term of (S.6) becomes

$$\int_{(X-x)'\beta\leqslant-\alpha}(\mathbb{I}\{y=1\}-F_0(X'\beta_0+\alpha_0))\,dF_X(X),$$

and the second term of (S.6) reduces to

$$\int_{(X-x)'\beta\geqslant\alpha}(F_0(X'\beta_0)-\mathbb{I}\{y\leqslant 2\})\,dF_X(X).$$

We calculate the partial derivative with respect to α. Let $g_0(\cdot|r)$ be the conditional density function of $X'\beta_0$ given $X_{-1}=r$ and $g_0(\cdot)$ be the marginal density of $X'\beta_0$. Notice that

$$\begin{aligned}\tau(y,x,\alpha,\beta)=&\int^{x'\beta_0-\alpha+(x-X)'(\beta-\beta_0)}S_1(y,t)g_0(t|r)dtdF_{X_{-1}}(r)\\&+\int_{x'\beta_0+\alpha+(x-X)'(\beta-\beta_0)}S_2(y,t)g_0(t|r)dtdF_{X_{-1}}(r),\end{aligned}\tag{S.7}$$

where

$$S_1(y,t)\equiv\mathbb{E}\left[\mathbb{I}\{y=1\}-\mathbb{I}\{Y\leqslant 2\}|X'\beta_0=t\right]=\mathbb{I}\{y=1\}-F_0(t+\alpha_0),\tag{S.8}$$

$$S_2(y,t)\equiv\mathbb{E}\left[\mathbb{I}\{Y=1\}-\mathbb{I}\{y\leqslant 2\}|X'\beta_0=t\right]=F_0(t)-\mathbb{I}\{y\leqslant 2\}.\tag{S.9}$$

Therefore, for $\nu\to 0$,

$$\begin{aligned}&\tau(y,x,\alpha+\nu,\beta_0)-\tau(y,x,\alpha,\beta_0)\\&=-\int_{x'\beta_0-(\alpha+\nu)}^{x'\beta_0-\alpha}S_1(y,t)g_0(t)dt-\int_{x'\beta_0+\alpha}^{x'\beta_0+\alpha+\nu}S_2(y,t)g_0(t)dt\\&=-\nu S_1(y,x'\beta_0-\alpha)g_0(x'\beta_0-\alpha)-\nu S_2(y,x'\beta_0+\alpha)g_0(x'\beta_0+\alpha).\end{aligned}\tag{S.10}$$

Combining equations (S.8), (S.9), and (S.10) yields

$$\begin{aligned}\frac{\partial\tau(Y,X,\alpha,\beta_0)}{\partial\alpha}=&\left[F_0(X'\beta_0-\alpha+\alpha_0)-\Delta_1\right]g_0(X'\beta_0-\alpha)\\&+\left[1-\Delta_3-F_0(X'\beta_0+\alpha)\right]g_0(X'\beta_0+\alpha),\end{aligned}\tag{S.11}$$

and thus

$$\frac{\partial\tau(Y,X,\alpha_0,\beta_0)}{\partial\alpha}=\left[F_0(X'\beta_0)-\Delta_1\right]g_0(X'\beta_0-\alpha_0)+\left[1-\Delta_3-F_0(X'\beta_0+\alpha_0)\right]g_0(X'\beta_0+\alpha_0).\tag{S.12}$$

Furthermore, observe that

$$\mathbb{E}\left[S_1(Y, X'\beta_0 - \alpha_0)|X'\beta_0\right] = 0, \tag{S.13}$$
$$\mathbb{E}\left[S_2(Y, X'\beta_0 + \alpha_0)|X'\beta_0\right] = 0. \tag{S.14}$$

Using (S.11), (S.13), and (S.14), we have

$$V_{\alpha_0,R} = -\mathbb{E}\left[f_0(X'\beta_0)g_0(X'\beta_0 - \alpha_0) + f(X'\beta_0 + \alpha_0)g_0(X'\beta_0 + \alpha_0)\right], \tag{S.15}$$

which is negative and thus satisfies Assumption 5 of Chen (2002). □

Note that even for the binary choice data, there is no general ranking of the asymptotic variances between our Stage 1 estimator in Section 2.2 and the maximum rank estimator. (The estimators of Sherman (1993) and Cavanagh and Sherman (1998) coincide in this case.) They deviate from the efficient estimator (Klein and Spady, 1993) in different ways. Our Stage 1 estimator is not efficient because it uses a simple moment condition rather than the efficient score function, in order to avoid tuning parameters. The efficiency loss in the maximum rank estimator, on the other hand, lies in its ignorance of the information contained in the distribution function. Write the asymptotic variance of both estimators for β_{-1} in the sandwich form $H_d^{-1}\Sigma_d H_d^{-1}$, for $d \in \{npmle, rank\}$, we have (see e.g. Table 2 of Groeneboom and Hendrickx (2019))

$$\begin{aligned}
H_{npmle} &= \mathbb{E}\left[\{X_{-1} - \mathbb{E}[X_{-1}|X'\beta_0]\}^{\otimes 2} f_0(X'\beta_0)\right], \\
J_{npmle} &= \mathbb{E}\left[\{X_{-1} - \mathbb{E}[X_{-1}|X'\beta_0]\}^{\otimes 2} (F_0(X'\beta_0) - \Delta_1)^2\right], \\
H_{rank} &= \mathbb{E}\left[\{X_{-1} - \mathbb{E}[X_{-1}|X'\beta_0]\}^{\otimes 2} g_0(X'\beta_0) f_0(X'\beta_0)\right], \\
J_{rank} &= \mathbb{E}\left[\{X_{-1} - \mathbb{E}[X_{-1}|X'\beta_0]\}^{\otimes 2} g_0(X'\beta_0)^2 (F_0(X'\beta_0) - \Delta_1)^2\right].
\end{aligned}$$

The additional factor $g_0(X'\beta_0)$ in both H and J for the rank estimator prevents a definite ranking of two asymptotic variances. Due to the explicit form of the efficient score for the binary choice model, Groeneboom and Hendrickx (2018) show that one can smooth the NPMLE and build the estimating equation with the additional weight $f_0/(F_0(1 - F_0))$, which leads to the efficient estimation.

S1.2 Computation of NPMLE and Zero-crossing Points

This section provides computational details for the joint estimator proposed in Section 2.3. Recall that the NPMLE $\tilde{F}_n(\cdot;\alpha,\beta)$ can be computed by the iterative convex minorant algorithm in Groeneboom and Wellner (1992) and Groeneboom and Jongbloed (2014). Here we provide a more detailed description. The number of mass points, denoted by p, is smaller than $2n$, because for any i with $\Delta_{2i}=0$, either $X_i'\beta+\alpha$ (if $\Delta_{1i}=1$) or $X_i'\beta$ (if $\Delta_{3i}=1$) does not enter the log-likelihood function. Denote the remaining elements in the set $\{X_i'\beta, X_i'\beta+\alpha : i=1,2,...,n\}$ as $U_j^{(\alpha,\beta)}, j=1,2,...,p$. Partition the observations into the following four groups:

$$\begin{aligned}
I_1 &= \{1 \leqslant j \leqslant p : U_j^{(\alpha,\beta)} = X_i'\beta \ \text{ for some } \ i \ \text{ and } \ \Delta_{1i}=1\},\\
I_{2l} &= \{1 \leqslant j \leqslant p : U_j^{(\alpha,\beta)} = X_i'\beta \ \text{ for some } \ i \ \text{ and } \ \Delta_{2i}=1\},\\
I_{2r} &= \{1 \leqslant j \leqslant p : U_j^{(\alpha,\beta)} = X_i'\beta+\alpha \ \text{ for some } \ i \ \text{ and } \ \Delta_{2i}=1\},\\
I_3 &= \{1 \leqslant j \leqslant p : U_j^{(\alpha,\beta)} = X_i'\beta+\alpha \ \text{ for some } \ i \ \text{ and } \ \Delta_{3i}=1\}.
\end{aligned}$$

Define k as a function that maps any index from I_{2l} to I_{2r} for a given observation i with $\Delta_{2i}=1$: $k(j)=m$ if $U_j^{(\alpha,\beta)}=X_i'\beta$ and $U_m^{(\alpha,\beta)}=X_i'\beta+\alpha$, for $\Delta_{2i}=1$. Let $v^{(t)} \equiv (v_1^{(t)},...,v_p^{(t)})'$ be the output from the t-th iteration, then $v^{(t+1)}$ is the left derivative of the cumulative sum diagram consisting of the following points:

$$P_0=(0,0), P_j=\left(\sum_{i=1}^{j} H_j(v^{(t)}), \sum_{i=1}^{j} v_i^{(t)} H_j(v^{(t)}) - G_j(v^{(t)})\right), j=1,...,p,$$

where

$$G_j(v)=\begin{cases} -v_j^{-1} \quad \text{if} \quad j \in I_1,\\ (v_{k(j)}-v_j)^{-1} \quad \text{if} \quad j \in I_{2l},\\ -(v_j - v_{k^{-1}(j)})^{-1} \quad \text{if} \quad j \in I_{2r},\\ (1-v_j)^{-1} \quad \text{if} \quad j \in I_3,\end{cases}$$

and

$$H_j(v) = \begin{cases} v_j^{-2} \quad \text{if} \quad j \in I_1, \\ (v_{k(j)} - v_j)^{-2} \quad \text{if} \quad j \in I_{2l}, \\ (v_j - v_{k^{-1}(j)})^{-2} \quad \text{if} \quad j \in I_{2r}, \\ (1 - v_j)^{-2} \quad \text{if} \quad j \in I_3. \end{cases}$$

The initial value can be set as $v^{(0)} = (1/p, 2/p, ..., 1)'$, which assigns the same probability mass on each jump point.

Regarding the zero-crossing point, we adopt the modified Barzilai-Borwein method (Varadhan and Gilbert, 2009), which does not require the differentiability of the estimating equations. We illustrate with our joint estimator for $\tilde{\theta}_n \equiv (\tilde{\alpha}_n, \tilde{\beta}_n')'$ which solves $\Phi_n(\cdot)$ in equation (2.7). If the estimating equations were differentiable, then the Newton-Raphson method would iterate with

$$\theta_{k+1} = \theta_k - \dot{\Phi}_n^{-1}(\theta_k)\Phi_n(\theta_k), \text{ for } k = 1, 2, \ldots.$$

Instead, we proceed with the following iteration:

$$\theta_{k+1} = \theta_k - l_k \Phi_n(\theta_k), \text{ for } k = 1, 2, \ldots,$$

where l_k is known as the spectral step-length. One popular choice is

$$l_k = \frac{s_{k-1}' s_{k-1}}{s_{k-1}' y_{k-1}}, \tag{S.16}$$

where $s_{k-1} = \theta_k - \theta_{k-1}$ and $y_{k-1} = \Phi_n(\theta_k) - \Phi_n(\theta_{k-1})$. The initial step-length is set as $l_0 = \min\left\{1, \frac{1}{\|\Phi_n(\beta_0)\|}\right\}$. To achieve the global convergence, the spectral iterate scheme needs to be combined with a suitable line search technique:

$$\phi_n(\theta_{k+1}) \leqslant \max_{0 \leqslant j \leqslant M} \phi_n(\theta_{k-j}) + \eta_k - 10^{-4} l_k^2 \phi_n(\theta_k), \tag{S.17}$$

in which $\phi_n(\theta) = \Phi_n(\theta)'\Phi_n(\theta)$ and η_k is a positive decreasing sequence that $\sum_{k=0}^{\infty} \eta_k < \infty$. We refer interested readers to Varadhan and Gilbert (2009) for the theoretical background and additional Monte Carlo evidence.

S1.3 Models with Four or More Categories

Here we discuss the general ordered response model where the dependent variable can take more than three values such as in the empirical applications of Cameron and Heckman (1998) or Klein and Sherman (2002). Formally, the dependent variable is determined by

$$Y_i = \begin{cases} 1 \quad \text{if} \quad \varepsilon_i \leqslant X_i'\beta_0, \\ 2 \quad \text{if} \quad X_i'\beta_0 < \varepsilon_i \leqslant X_i'\beta_0 + \alpha_{0,1}, \\ 3 \quad \text{if} \quad X_i'\beta_0 + \alpha_{0,1} < \varepsilon_i \leqslant X_i'\beta_0 + \alpha_{0,2}, \\ \vdots \\ J+1 \quad \text{if} \quad \varepsilon_i > X_i'\beta_0 + \alpha_{0,J-1}, \end{cases} \tag{S.18}$$

for $i = 1, \ldots, n$ and $J \geqslant 3$, with a set of ordered thresholds $(\alpha_{0,1}, \ldots, \alpha_{0,J-1})$. In the same spirit of Lewbel (2002), our two-stage estimator in Section 2.2 directly applies to this setting. In the first stage, the binary data $(\mathbb{I}(Y_i > 1), X_i)_{i=1}^n$ is employed to obtain $\hat{\beta}_n$ and $\hat{F}_n(\cdot; \hat{\beta}_n)$ in the same manner as Stage 1 in Section 2.2. Then one utilizes the data $(\mathbb{I}(Y_i > j+1), X_i)_{i=1}^n$ in the second stage to estimate $\alpha_{0,j}$ through the following estimating equation of α_j for $j = 1, \ldots, J-1$,

$$\Psi_{j,n}\left(\alpha_j, \hat{\beta}_n, \hat{F}_n(\cdot; \hat{\beta}_n)\right) = \frac{1}{n}\sum_{i=1}^n \left[1 - \mathbb{I}(Y_i > j+1) - \hat{F}_n(X_i'\hat{\beta}_n + \alpha_j; \hat{\beta}_n)\right]. \tag{S.19}$$

The large sample properties of the two-stage estimator are presented in the following. We collect the finite dimensional parameter in $\theta_0^J = (\alpha_{0,1}, \alpha_{0,2}, \ldots, \alpha_{0,J-1}, \beta_{0-}')'$ and denote the two-stage semiparametric estimator by $\hat{\theta}_n^J = (\hat{\alpha}_{n,1}, \hat{\alpha}_{n,2}, \ldots, \hat{\alpha}_{n,J-1}, \hat{\beta}_{n-}')'$ with $J \geqslant 3$. Furthermore, we introduce the following notations:

$$\begin{aligned} \psi_{0,j}(Z_i) &= \left[F_0(U_i + \alpha_{0,j}) - \mathbb{I}(Y_i \leqslant j+1)\right], \\ V_{\alpha_{0,j}} &= \frac{\partial}{\partial \alpha}\mathbb{E}[F_0(X'\beta_0 + \alpha)]\big|_{\alpha=\alpha_{0,j}}, \\ \psi_{\alpha_{0,j}} &= V_{\alpha_{0,j}}^{-1}(\psi_{0,j} + \psi_{F_0} + V_{\beta_0}\psi_{\beta_0}), \end{aligned}$$

for $j = 1, 2, \ldots, J-1$. Corollary S1.1 presents the asymptotic normality of $\hat{\theta}_n^J$. Its proof follows from a straightforward modification of our Theorem 3.1 up to some notation changes, and thus is omitted.

Corollary S1.1. *Suppose Conditions 1 to 9 hold and $V_{\alpha_{0,j}} \neq 0$ for all j. Then we have*

$$\sqrt{n}\left(\hat{\theta}_n^J - \theta_0^J\right) \Rightarrow \mathbb{N}(0, \Sigma_0^J),$$

where

$$\Sigma_0^J = \mathbb{E}[(\psi_{\alpha_{0,1}}, \psi_{\alpha_{0,2}}, \ldots, \psi_{\alpha_{0,J-1}}\psi'_{\beta_0})'(\psi_{\alpha_{0,1}}, \psi_{\alpha_{0,2}}, \ldots, \psi_{\alpha_{0,J-1}}, \psi'_{\beta_0})].$$

Our joint estimator in Section 2.3 is also applicable. From a computational point of view, if there are more than three categories, only the interval corresponding to the chosen category and its adjacent ones are relevant for the computation of the NPMLE; the other intervals can be discarded; see Groeneboom (2014), [p.2093]. Therefore, the construction of the NPMLE is almost the same as the case with three categories. The consistency of the NPMLE for multiple categories is shown in Schick and Yu (2000). However, the rate of convergence or the asymptotic properties of its linear functionals remain unknown. We leave this challenging issue to the future research. Thereafter, we recommend practitioners use the methods in Klein and Sherman (2002)[2] or Coppejans (2007), if efficiency is the main concern. Since our empirical application in Section 4.2 involves three categories, we will focus on the setup specified by (1.1).

S1.4 Estimation of the Trend Function in the Honoré-Paula Model

Honoré and de Paula (2010) apply the ordered response model (with three categories) to identify and estimate an interdependent duration model of two players. Our simulation design and empirical application are based on this model. Let (T_1, T_2) be the time of switching from an initial activity to an alternative activity. The utility flow of the alternative activity for one player depends on whether the other player has switched or not, which causes an endogenous interaction effect. The equilibrium of two duration variables (T_1, T_2) are characterized by

$$\begin{aligned} T_1 &= \inf\left\{t_1 : \Lambda(t_1)\exp\left(X_1'\beta_0\right)\exp\left[\alpha^*\mathbb{I}\left\{T_2 \leqslant t_1\right\}\right] \geqslant \epsilon_1\right\}, \\ T_2 &= \inf\left\{t_2 : \Lambda(t_2)\exp\left(X_2'\beta_0\right)\exp\left[\alpha^*\mathbb{I}\left\{T_1 \leqslant t_2\right\}\right] \geqslant \epsilon_2\right\}, \end{aligned} \tag{S.20}$$

where the unknown scalar α^* captures the interaction effect, and the function $\Lambda(t)$ captures the deterministic trend. Each player j $(j = 1, 2)$ has covariate X_j and the initial random

[2] The K-S estimator is also semiparametrically efficient under an additional periodicity restriction on the covariates; see Section 3.3 of Coppejans (2007).

utility flow ϵ_j. This interdependent duration model induces an ordered response model that involves the parameters β_0 and α^*, but not the deterministic trend function $\Lambda(t)$. However, combining Theorem 3 of Honoré and de Paula (2010) and the key idea of Horowitz (1996) yields a nonparametric estimator for $\ln \Lambda(t)$. To elaborate on the proposal, let

$$h(t_1, t_2; x_1, x_2) \equiv \Pr\{T_1 \leqslant t_1, T_2 > t_2 | X_1 = x_1, X_2 = x_2\}, \text{ for } t_1 < t_2.$$

The proof of Theorem 3 in Honoré and de Paula (2010) implies that

$$\frac{\partial \ln \Lambda(t_1)}{\partial t_1} = \beta_{0k} \frac{\partial h / \partial t_1}{\partial h / \partial x_{1k}}. \tag{S.21}$$

Thus, one can adopt the estimator given by equation (2.4) in Horowitz (1996) and plug in a nonparametric kernel estimator for the partial derivative of $h(t_1, t_2; x_1, x_2)$. Given that the theoretical properties of such an estimator follow from Horowitz (1996), we will not expand on the issue.

S2 Additional Simulation Results

S2.1 The Effect of Trimming

We repeat the Monte Carlo exercises in Section 4.1 for the isotonic two-stage estimator and the NPMLE-based joint estimator but use the *truncated* estimating equations. The trimming scheme follows that of Groeneboom and Hendrickx (2018), which restricts observations to those with the estimated $\hat{F}_n(X_i'\beta)$ within the interval $[\tau, 1-\tau]$, where $\tau \in [0, 1/2)$ is the truncation parameter.

For the two-stage approach, the trimming revises the estimation procedure described in Section 2.2 as follows:

Stage 1(ii) (Trimmed). Given $\hat{F}_n(\cdot; \beta)$, the estimator $\hat{\beta}_{tr}$ for the regression coefficient is the zero-crossing point of the truncated estimating equation with respect to β:

$$\frac{1}{n} \sum_{i=1}^{n} X_{i,-1} \left[\Delta_{1i} - \hat{F}_n(X_i'\beta; \beta) \right] \mathbb{I}\{\tau \leqslant \hat{F}_n(X_i'\beta; \beta) \leqslant 1 - \tau\} = 0. \tag{S.22}$$

Stage 2 (Trimmed). Given $\hat{\beta}_{tr}$ and $\hat{F}_n(\cdot; \hat{\beta}_{tr})$, we estimate α_0 by $\hat{\alpha}_{tr}$, which is the zero-crossing point of the truncated estimating equation $\Psi_{n,tr}\left(\hat{\alpha}_{tr}, \hat{\beta}_{tr}, \hat{F}_n(\cdot;\hat{\beta}_{tr})\right) = 0$, where

$$\Psi_{n,tr}\left(\alpha, \hat{\beta}_{tr}, \hat{F}_n(\cdot;\hat{\beta}_{tr})\right) = \frac{1}{n}\sum_{i=1}^{n}\left[1-\Delta_{3i}-\hat{F}_n(X_i'\hat{\beta}_{tr}+\alpha;\hat{\beta}_{tr})\right]\mathbb{I}\{\tau \leqslant \hat{F}_n(X_i'\hat{\beta}_{tr}+\alpha;\hat{\beta}_{tr}) \leqslant 1-\tau\}. \tag{S.23}$$

For the joint approach, the trimmed version is described as follows. Given $\tilde{F}_n(\cdot;\alpha,\beta)$, the estimators $(\tilde{\alpha}_{tr}, \tilde{\beta}_{tr})$ are the zero-crossing points of the following estimating equations:

$$\Phi_{n,tr}(\tilde{\alpha}_{tr}, \tilde{\beta}_{tr}) = 0, \tag{S.24}$$

where

$$\Phi_{n,tr}(\alpha,\beta) \equiv \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} X_{i,-1}\left[\Delta_{1i}-\tilde{F}_n(X_i'\beta;\alpha,\beta)\right]\mathbb{I}\{\tau \leqslant \tilde{F}_n(X_i'\beta;\alpha,\beta) \leqslant 1-\tau\} \\ \frac{1}{n}\sum_{i=1}^{n}\left[1-\Delta_{3i}-\tilde{F}_n(X_i'\beta+\alpha;\alpha,\beta)\right]\mathbb{I}\{\tau \leqslant \tilde{F}_n(X_i'\beta+\alpha;\alpha,\beta) \leqslant 1-\tau\} \end{bmatrix}.$$

Their finite sample performances are summarized in Tables S1 and S2, along with the two-stage and the joint estimators without any trimming. To evaluate the effect of trimming, we set the truncation parameter $\tau = 0.01$, larger than the one used by Groeneboom and Hendrickx (2018) ($\tau = 0.001$) . According to Tables S1 and S2, even with this relatively large value of τ, the bias and RMSE for the estimators with trimming are similar to those without trimming. The effect of trimming is particularly negligible for the joint estimator of both β_{0-} and α_0, and for the two-stage estimator of β_{0-}. When it comes to the two-stage estimator of α_0, the trimming reduces its bias to some extent, but that effect diminishes with the increase of sample size. Overall, trimming is not critical for implementing our two-stage and joint estimators.

Table S1: Performance of estimators with and without trimming in the estimating equations, normal errors.

		n = 250		n = 500		n = 750		n = 1000	
Methods		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
Two-stage	β_{02}	-.0483	.1834	-.0419	.1360	-.0271	.1094	-.0211	.0938
	β_{03}	-.0505	.1822	-.0409	.1333	-.0331	.1064	-.0281	.0893
	α^*	-.0532	.1526	-.0408	.1113	-.0320	.0933	-.0263	.0804
Two-stage	β_{02}	-.0483	.1834	-.0418	.1361	-.0270	.1095	-.0208	.0939
(truncated,	β_{03}	-.0505	.1822	-.408	.1333	-.0330	.1065	-.0279	.0897
$\tau = 0.01$)	α^*	-.0368	.1603	-.0318	.1135	-.0267	.0943	-.0213	.0810
Joint	β_{02}	-.0089	.1802	-.0161	.1320	-.0064	.1054	-.0036	.0888
	β_{03}	-.0119	.1803	-.0149	.1312	-.0129	.0989	-.008	.0828
	α^*	-.0275	.1337	-.0213	.0951	-.0153	.0777	-.0117	.0652
Joint	β_{02}	-.0086	.1803	-.0159	.1320	-.0059	.1059	-.0034	.0886
(truncated,	β_{03}	-.0116	.1807	-.0148	.1316	-.0128	.0995	-.0107	.0834
$\tau = 0.01$)	α^*	-.0274	.1338	-.0215	.0948	-.0155	.0777	-.0117	.0653

Table S2: Performance of estimators with and without trimming in the estimating equations, exponential errors.

		n = 250		n = 500		n = 750		n = 1000	
Methods		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
Two-stage	β_{02}	-.0505	.1745	-.0400	.1249	-.0337	.0965	-.0259	.0915
	β_{03}	-.0572	.1714	-.0409	.1203	-.0319	.0996	-.0272	.0891
	α^*	-.0417	.1411	-.0353	.1009	-.0290	.0802	-.0264	.0745
Two-stage	β_{02}	-.0505	.1745	-.0399	.1252	-.0333	.0966	-.0255	.0919
(truncated,	β_{03}	-.0572	.1714	-.0408	.1203	-.0315	.0996	-.0269	.0893
$\tau = 0.01$)	α^*	-.0253	.1474	-.0266	.1014	-.0221	.0809	-.0209	.0748
Joint	β_{02}	-.0144	.1744	-.0158	.1193	-.0165	.0896	-.0117	.0839
	β_{03}	-.0228	.1634	-.0174	.1143	-.0144	.0927	-.0132	.0814
	α^*	-.0193	.1253	-.0164	.0850	-.0136	.0652	-.0138	.0590
Joint	β_{02}	-.0114	.1745	-.0154	.1198	-.0161	.0895	-.0112	.0842
(truncated,	β_{03}	-.0229	.1633	-.0168	.1150	-.0142	.0928	-.0131	.0810
$\tau = 0.01$)	α^*	-.0194	.1255	-.0163	.0853	-.0133	.0655	-.0135	.0588

S2.2 Confidence Intervals

Tables S3 and S4 report the empirical coverage rate (CR) and the mean length (ML) of the 95% confidence intervals (CIs) based on the nonparametric bootstrap for the proposed semiparametric methods and several alternatives described in Section 3 of the main paper. The simulation design and the computation of estimators follow Section 4.1. The trimming proportion p is set to 0.2 in the K-S estimator. The number of bootstrap replications is 200. The sample sizes are 250, 500, 750 and 1000. The number of simulations is 1000.

We make the following observations regarding the confidence intervals. First, the CIs of both two-stage and joint estimators have good coverage rates for the coefficients β_{0-} and the threshold parameter α^* with moderate sample sizes. Between them, the coverage rate of the joint estimator is closer to the nominal rate. Second, the CIs of two-stage and joint estimators are substantially shorter than the rank and K-S estimators. Third, the coverage rates of the bootstrap-based CIs of the rank and K-S estimator are also quite precise. On the other hand, the SMS exhibits over-coverage in all scenarios.

Table S3: Coverage proportion (CR) and mean length (ML) of 95% bootstrap-based confidence intervals, normal errors.

		$n = 250$		$n = 500$		$n = 750$		$n = 1000$	
Methods		CR	ML	CR	ML	CR	ML	CR	ML
Two-stage	β_{02}	.889	.683	.900	.480	.927	.393	.936	.338
	β_{03}	.901	.656	.912	.465	.923	.378	.939	.329
	α^*	.876	.555	.897	.393	.914	.320	.923	.277
Joint	β_{02}	.916	.739	.925	.519	.933	.425	.947	.365
	β_{03}	.917	.714	.924	.506	.949	.412	.954	.355
	α^*	.897	.523	.929	.382	.943	.311	.945	.267
Rank	β_{02}	.926	1.327	.940	1.025	.938	.929	.938	.883
	β_{03}	.955	1.338	.945	1.056	.950	.957	.921	.907
	α^*	.884	.797	.912	.577	.921	.500	.903	.457
K-S	β_{02}	.901	1.376	.932	1.048	.946	.866	.940	.757
	β_{03}	.909	1.426	.943	1.073	.961	.871	.952	.768
	α^*	.897	1.324	.922	.961	.917	.782	.921	.702
SMS	β_{02}	.988	2.506	.996	.662	.997	.454	.995	.385
	β_{03}	.990	2.505	.993	.678	.990	.464	.995	.397
	α^*	.995	2.381	.989	.694	.991	.500	.995	.427

Table S4: Coverage proportion (CR) and mean length (ML) of 95% bootstrap-based confidence intervals, exponential errors.

		$n = 250$		$n = 500$		$n = 750$		$n = 1000$	
Methods		CR	ML	CR	ML	CR	ML	CR	ML
Two-stage	β_{02}	.895	.649	.916	.454	.936	.365	.922	.317
	β_{03}	.902	.622	.918	.435	.940	.354	.928	.305
	α^*	.896	.531	.920	.371	.935	.300	.933	.259
Joint	β_{02}	.920	.690	.935	.485	.959	.390	.945	.338
	β_{03}	.926	.662	.943	.466	.950	.380	.939	.384
	α^*	.918	.496	.952	.360	.955	.289	.948	.247
Rank	β_{02}	.945	1.278	.938	1.000	.938	.914	.930	.853
	β_{03}	.955	1.281	.956	1.025	.935	.945	.919	.886
	α^*	.911	.759	.929	.553	.931	.480	.911	.441
K-S	β_{02}	.921	1.354	.947	.970	.958	.795	.933	.704
	β_{03}	.947	1.370	.957	.998	.967	.817	.959	.720
	α^*	.933	1.740	.911	.866	.929	.709	.904	.637
SMS	β_{02}	.998	1.174	.994	.514	.991	.422	.997	.384
	β_{03}	.997	1.136	.996	.521	.996	.431	.994	.395
	α^*	.992	1.127	.989	.444	.993	.353	.988	.314

S2.3 Unbounded Covariates and Errors

This section repeats the simulation exercises in Section 4.1 for covariates X_j and error terms ϵ_j ($j = 1, 2$) with unbounded support. To be sepcific, X_{j1} is a standard normal variable; X_{j2} is a $\chi^2(1)$ variable standardized to mean zero and variance one; the remaining components (X_{j3}, X_{j4}, X_{j5}) are multivariate standard normal with the pairwise correlation coefficient between X_{jk_1} and X_{jk_2} equal to $0.5^{|k_1-k_2|}$. The error terms (ϵ_1, ϵ_2) once again take two types: (I). Normal errors: $\log(\epsilon_1)$ and $\log(\epsilon_2)$ have the standard normal distribution, and (II). Exponential errors: $\log(\epsilon_1)$ and $\log(\epsilon_2)$ have the unit exponential distribution. The following Figures S2 and S3 show that the performance of estimators in the case of unbounded covariates and errors is similar to those in the bounded cases shown in Figures 3 and 4. Observations and discussions in Section 4.1 apply here.

Figure S2a: Finite sample performances of estimators for $(\beta_{02}, \beta_{03}, \alpha^*)$, normal errors: two-stage (black, dashed, $\bullet$), joint (red, solid, $\blacksquare$), rank (blue, long-dashed, $\blacktriangle$), K-S (green, dotted, $\times$ for β_{02}, β_{03}; $\times$ and $+$ for α^* depending on p), SMS (brown, dot-dashed, $\blacklozenge$), ordered probit (violet, two-dashed, $\boxdot$), ordered logit (cyan, very long-dashed, $\circ$).

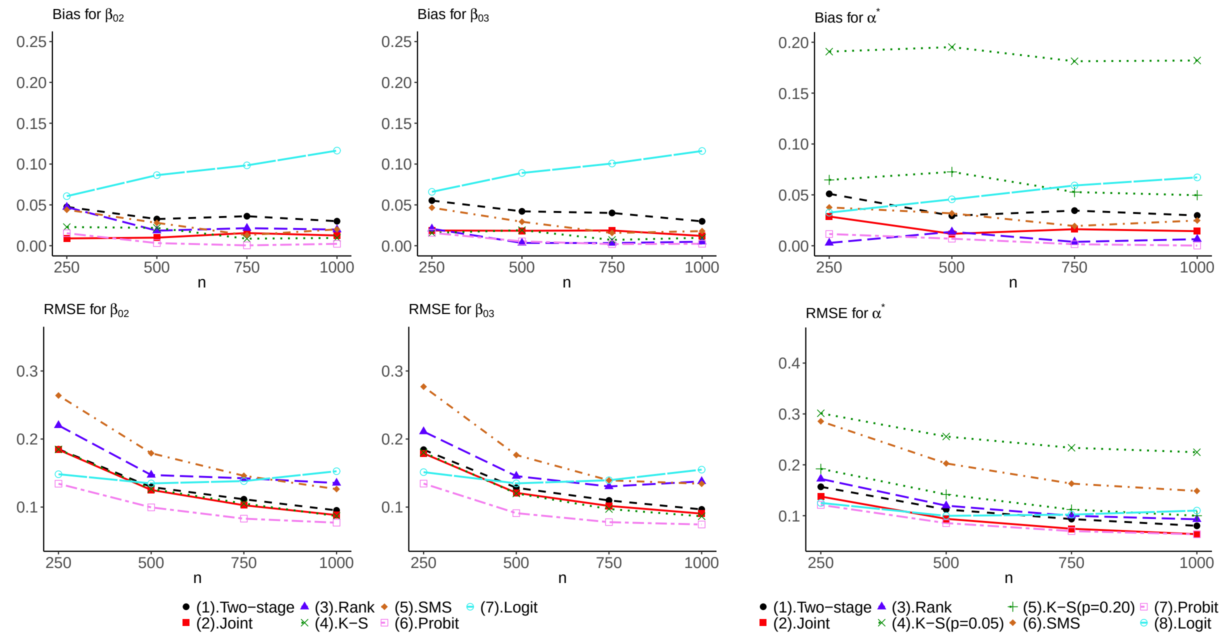


Figure S2b: Finite sample performances of estimators for $(\beta_{02}, \beta_{03}, \alpha^*)$, exponential errors.

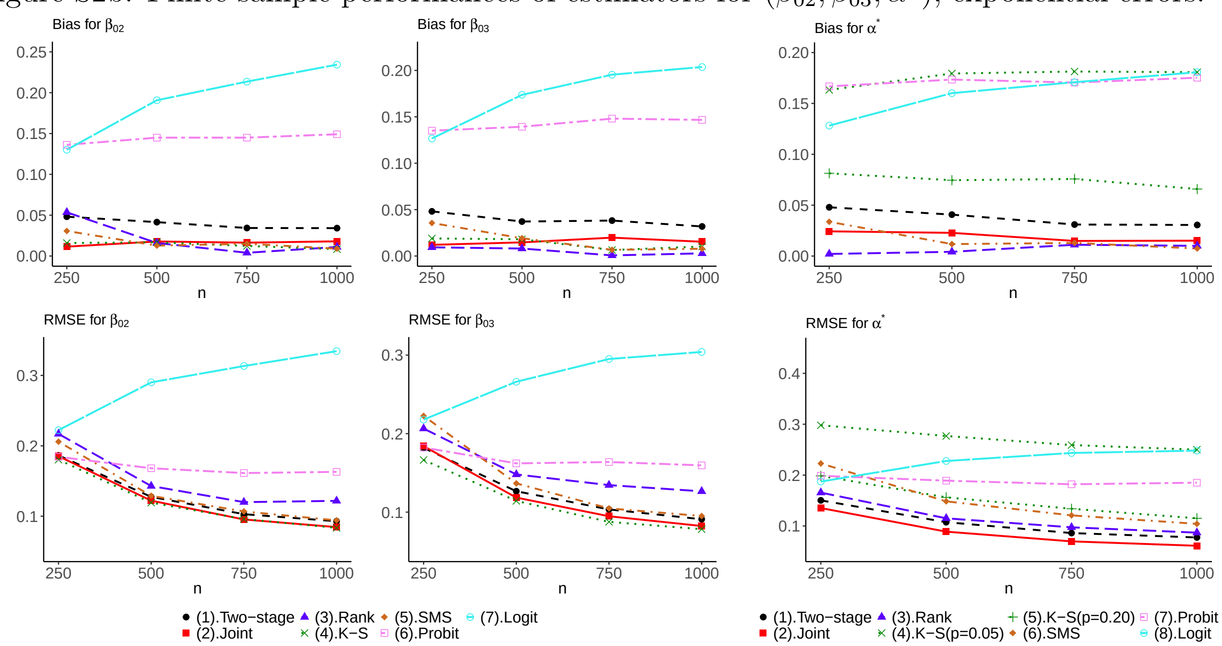


Figure S3a: Pointwsie estimators for the function $H(w)$ at $w = -2, -1$ and 1, normal errors: two-stage (black, dashed, $\bullet$), joint (red, solid, $\blacksquare$), rank (blue, long-dashed, $\blacktriangle$), K-S (green, dotted, $\times$), SMS (brown, dot-dashed, $\blacklozenge$), ordered probit (violet, two-dashed, $\square$), ordered logit (cyan, very long-dashed, $\circ$).

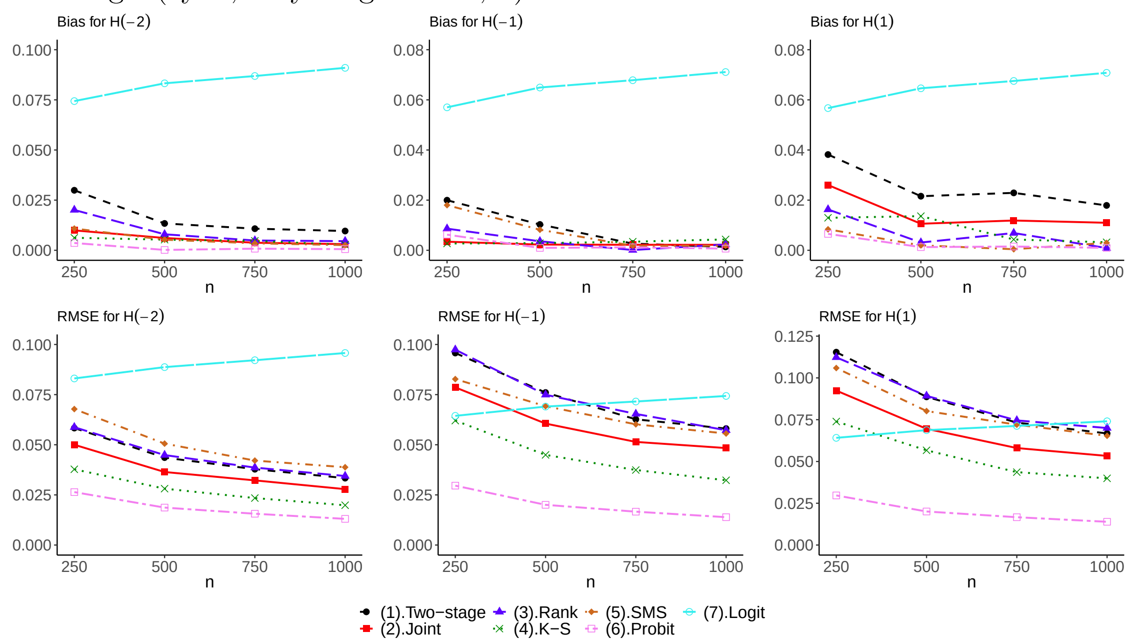
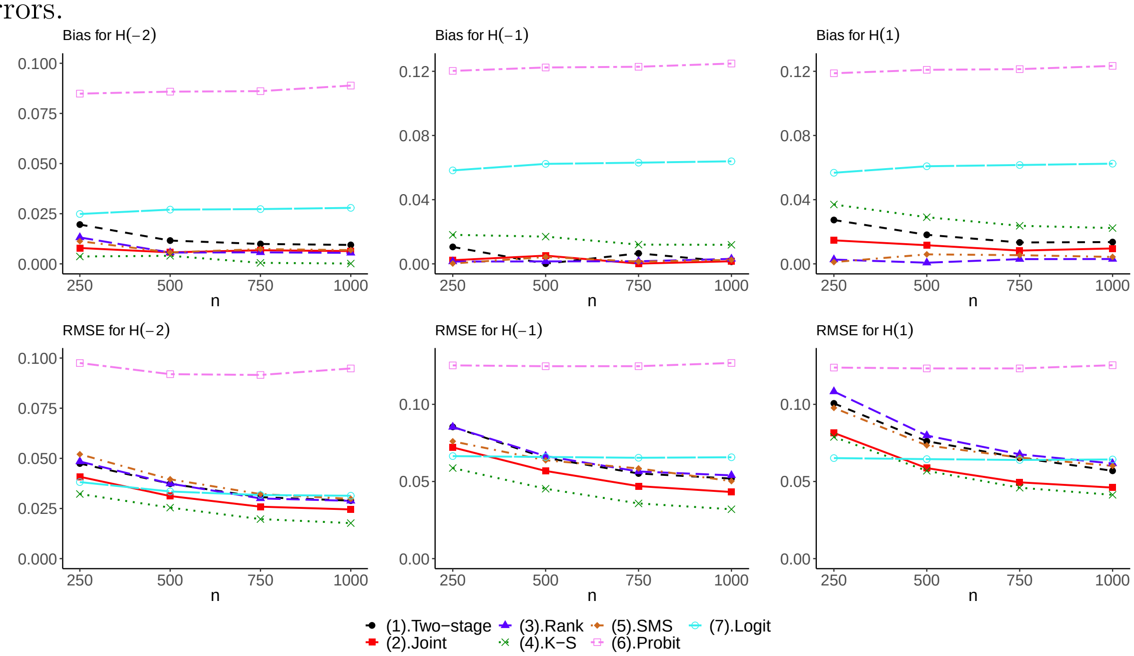


Figure S3b: Pointwsie estimators for the function $H(w)$ at $w = -2, -1$ and 1, exponential errors.



S3 Technical Proofs Related to Two-stage Estimation

This section presents and proves the technical lemmas and other results that are used in the proof of Theorem 3.1. We also prove Theorem 3.2.

We first restate some necessary definitions and Theorem 2.4.1 in Van Der Vaart and Wellner (1996) that will be used repeatedly in the sequel. Let $\mathcal{F}$ be the class of functions and $L_2(Q)$ be the L_2-norm defined by a probability measure Q. For any probability measure Q, let $N(\varepsilon, \mathcal{F}, L_2(Q))$ be the minimal number of balls of radius ε needed to cover the class $\mathcal{F}$. The entropy integral $J(\delta, \mathcal{F})$ is defined as

$$J(\delta, \mathcal{F}) \equiv \sup_Q \int_0^\delta \sqrt{1 + \log N(\varepsilon, \mathcal{F}, L_2(Q))} d\varepsilon.$$

An envelope function of a functional class $\mathcal{F}$ is a function F such that $|f(x)| \leqslant F(x)$ for all x and $f \in \mathcal{F}$.

Lemma S1 (Theorem 2.14.1 in Van Der Vaart and Wellner (1996))**.** *Let P_0 be the distribution of the underlying observation and let $\mathcal{F}$ be a P_0-measurable class with an envelope function F. We have*

$$\mathbb{E} \sup_{f \in \mathcal{F}} |\mathbb{G}_n f| \lesssim J(1, \mathcal{F}) \parallel F \parallel_{P_0,2} . \tag{S.25}$$

We need to apply the following well-known entropy bounds concerning monotone functions and functions of bounded variation repeatedly. The bounds actually hold for the entropy integral uniformly over the underlying probability measure, which will be used in Section S4 as well. We refer readers to Theorem 2.7.5 on [p.159] of Van Der Vaart and Wellner (1996) or Lemma 3.8 on [p.36] of Van de Geer (2000) for the proofs.

Lemma S2 (Entropy Bounds)**.** *Let $\mathcal{A}_C$ be the class of monotone functions with values in $[0, C]$, then for all $\delta > 0$,*

$$J(\delta, \mathcal{A}_C) \lesssim \sqrt{\delta}. \tag{S.26}$$

Let $\mathcal{B}_C$ be the class of functions of bounded variation with values in $[c, C]$, then for all $\delta > 0$,

$$J(\delta, \mathcal{B}_C) \lesssim \sqrt{\delta}. \tag{S.27}$$

Now we obtain the entropy bounds for the key functional class in our context and prove the asymptotic characterizations for several terms appearing in our proof of Theorem 3.1.

Lemma S3. *The functional class $\mathcal{G}$ defined by*

$$\mathcal{G} \equiv \left\{(x, \delta_3) \mapsto (1 - \delta_3 - F(x'\beta + \alpha)) : (\alpha, \beta_-) \in \Theta, F(\cdot) \in \mathcal{A}\right\} \tag{S.28}$$

has bounded entropy integral. Therefore, we have the following Glivenko-Cantelli property:

$$(\mathbb{P}_n - P)\left[\hat{F}_n\left(X'\hat{\beta}_n + \hat{\alpha}_n; \hat{\beta}_n\right)\right] = o_p(1).$$

Moreover, we obtain the stochastic equicontinuity as

$$\mathbb{G}_n\left[\hat{F}_n\left(X'\hat{\beta}_n + \hat{\alpha}_n; \hat{\beta}_n\right) - F_0\left(X'\beta_0 + \alpha_0\right)\right] = o_p(1). \tag{S.29}$$

Proof. We first verify that the uniform entropy integral $J(1, \mathcal{G})$ is bounded. Because the isotonic estimator $\hat{F}_n(t, \beta)$ is a monotonically increasing function for any given β, $\mathcal{G}$ is the class of composite functions involving a monotonically increasing link/ridge function and a linear index $x'\beta + \alpha$ with parameters (α, β) belonging to a compact Euclidean space. By Lemma 2.3 in Balabdaoui, Groeneboom, and Hendrickx (2019), we get $\log N(\varepsilon, \mathcal{G}) \lesssim 1/\varepsilon$, so the uniform entropy integral $J(1, \mathcal{G})$ is indeed bounded. Therefore, the functional class $\mathcal{G}$ is P-Donsker, which directly implies the stated Glivenko-Cantelli property.

Regarding the stochastic equicontinuity, let $\theta' \equiv (\alpha, \beta_-')$ and consider the following class:

$$\mathcal{G}_\epsilon \equiv \left\{x \mapsto (F(x'\beta + \alpha) - F_0(x'\beta_0 + \alpha_0)) : \theta \in \Theta, F(\cdot) \in \mathcal{A}, |\theta - \theta_0| \vee \parallel F - F_0 \parallel_\infty \leqslant \epsilon\right\},$$

for some small positive ϵ. Again $\mathcal{G}_\epsilon$ has bounded entropy integral similarly as $\mathcal{G}$. Moreover, $\hat{F}_n\left(X'\hat{\beta}_n + \hat{\alpha}_n; \hat{\beta}_n\right) - F_0\left(X'\beta_0 + \alpha_0\right)$ belongs to $\mathcal{G}_\epsilon$ with probability tending to 1, because

$$\begin{aligned}
&\parallel \hat{F}_n\left(X'\hat{\beta}_n + \hat{\alpha}_n; \hat{\beta}_n\right) - F_0\left(X'\beta_0 + \alpha_0\right) \parallel_\infty \\
\leqslant &\parallel \hat{F}_n\left(X'\hat{\beta}_n + \hat{\alpha}_n; \hat{\beta}_n\right) - F_0\left(X'\hat{\beta}_n + \hat{\alpha}_n; \hat{\beta}_n\right) \parallel_\infty + \parallel F_0\left(X'\hat{\beta}_n + \hat{\alpha}_n; \hat{\beta}_n\right) - F_0\left(X'\beta_0 + \alpha_0\right) \parallel_\infty \\
\to_p &\ 0.
\end{aligned}$$

The convergence of the first term on the right hand side of the inequality follows from the uniform consistency of the isotonic estimator as in (S.40), whereas the convergence of the second term is due to the smoothness of $F_0(u; \beta)$ (w.r.t. both u and β) and the consistency of $\hat{\alpha}_n$ and $\hat{\beta}_n$. Thereafter, the desired stochastic equicontinuity follows from applying (S.25) to the class $\mathcal{G}_\epsilon$. □

Proof of the existence of $\hat{\alpha}_n$. We show the existence of a unique zero-crossing point of Ψ_n (defined in Stage 2 of Section 2.2) with probability approaching to 1. Because α is a scalar, the zero-crossing point of $\Psi_n(\alpha)$ can be equivalently defined as $\hat{\alpha}_n$ such that for any α:

$$(\hat{\alpha}_n - \alpha)\Psi_n(\alpha) \geqslant 0, \tag{S.30}$$

see Lemma 4.1 of Groeneboom and Hendrickx (2018). If the zero-crossing point does not exist, then for all α_1 there exists some α_2 such that

$$(\alpha_1 - \alpha_2)\Psi_n(\alpha_2) \leqslant -c < 0, \tag{S.31}$$

for some finite positive constant c. Such a constant term c exists because the isotonic estimate $\hat{F}_n(u; \hat{\beta}_n)$ is a piece-wise constant function with finitely many jumps for any n, so is $\Psi_n(\alpha)$ for all α. In particular, we have

$$(\alpha_0 - \alpha_2)\Psi_n(\alpha_2) \leqslant -c. \tag{S.32}$$

By the corresponding Glivenko-Cantelli property, we get

$$(\alpha_0 - \alpha_2)\Psi(\alpha_2) \leqslant -c/2, \tag{S.33}$$

with probability tending to 1. However, this contradicts the fact that α_0 is the unique zero-crossing point of $\Psi(\alpha)$, since $\Psi(\alpha)$ is monotone and continuous with respect to α, given the monotonicity and absolute continuity of F_0. Thus, the zero-crossing point $\hat{\alpha}_n$ exists with probability tending to 1. □

An immediate consequence of Lemma S3 is Lemma S4, which shows the negligibility of the terms I_{2n}^b and I_{3n}^c in our proof of Theorem 3.1. The claims directly follow from (S.29).

Lemma S4. *Suppose Conditions 1 to 9 hold. We characterize the following smaller order terms:*

$$\sqrt{n} I_{2n}^b = o_p(1), \quad \textit{and} \quad \sqrt{n} I_{3n}^c = o_p(1). \tag{S.34}$$

Now we prove several preparatory lemmas related to the linear representation of I_{2n}^a. Recall that $U = X'\beta_0$.

Lemma S5. *Suppose Conditions 1 to 9 hold. The following representation holds:*

$$I_{2n}^a = -\int \phi_{\alpha_0}(u)(\hat{F}_n(u;\beta_0) - \delta_1) dP(u, \delta_1), \tag{S.35}$$

where

$$\phi_{\alpha_0}(u) = g_0(u - \alpha_0)/g_0(u). \tag{S.36}$$

Proof. The result follows from a similar argument used in Lemma 4.1 of Groeneboom, Jongbloed, and Witte (2010):

$$\begin{aligned} I_{2n}^a &= -\int \left(\hat{F}_n(u+\alpha_0;\beta_0) - F_0(u+\alpha_0)\right) g_0(u) du \qquad \text{(S.37)} \\ &= -\int \left(\hat{F}_n(u;\beta_0) - F_0(u)\right) g_0(u-\alpha_0) du \\ &= -\int \phi_{\alpha_0}(u) \left(\hat{F}_n(u;\beta_0) - F_0(u)\right) dG_0(u) \\ &= -\int \phi_{\alpha_0}(u)(\hat{F}_n(u;\beta_0) - \delta_1) dP(u, \delta_1), \end{aligned}$$

where the last line uses the fact that $\delta_1 dP = F_0(u)g_0(u)$, since the probability density function of the binary choice data (U, Δ_1) is

$$p(u, \delta_1) = F_0(u)^{\delta_1}(1 - F_0(u))^{1-\delta_1} g_0(u). \tag{S.38}$$

□

We consider the piece-wise constant version of ϕ_{α_0} which is constant on the same intervals where the isotonic estimator $\hat{F}_n(\cdot;\beta)$ remains constant. Denote those intervals by $J_i = [\tau_i, \tau_{i+1})$. We define

$$\bar{\phi}_{\alpha_0}(u) = \phi_{\alpha_0}(\hat{A}_n(u;\beta)), \tag{S.39}$$

where

$$\hat{A}_n(u;\beta) = \begin{cases} \tau_i, & \text{if} \quad \forall t \in J_i : F_0(t) > \hat{F}_n(\tau_i;\beta), \\ s, & \text{if} \quad \exists s \in J_i : F_0(s) = \hat{F}_n(s;\beta), \\ \tau_{i+1}, & \text{if} \quad \forall t \in J_i : F_0(t) < \hat{F}_n(\tau_i;\beta), \end{cases}$$

for $u \in J_i$.

The following convergence results of the isotonic estimator are available from the first equation on [p.79] of Groeneboom and Wellner (1992), Lemma 3.1 of Groeneboom and Hendrickx (2018), Proposition 2 of Balabdaoui, Groeneboom, and Hendrickx (2019), and

Lemma 5.9 of Groeneboom and Wellner (1992).

Lemma S6. *Suppose Conditions 1 to 9 hold, then we have*

$$P\left(\lim_{n\to\infty}\sup_{\beta\in\mathcal{B},u}\left|\hat{F}_n(u;\beta)-F_0(u;\beta)\right|=0\right)=1, \tag{S.40}$$

and

$$\sup_{\beta\in\mathcal{B}}\left(\int\left|\hat{F}_n(u;\beta)-F_0(u;\beta)\right|^2 dG(u)\right)^{1/2}=O_p\left(\log n\times n^{-1/3}\right),$$

$$\sup_{\beta\in\mathcal{B}}\left(\int\left|\hat{A}_n(u;\beta)-u\right| dG(u)\right)^{1/2}=O_p\left(\log n\times n^{-1/3}\right).$$

Note that the statement in Lemma 3.1(ii) of Groeneboom and Hendrickx (2018) applied trimming on the distribution function in order to be comparable with the efficient estimators they proposed. Here we do not need any trimming, in the same spirit of Proposition 2 of Balabdaoui, Groeneboom, and Hendrickx (2019). The isotonic estimator is consistent in terms of the Hellinger distance without any trimming (Van de Geer, 1993). This implies pointwise consistency if the true error distribution function is absolutely continuous. Since both the isotonic estimator and the true distribution are monotone, pointwise consistency implies uniform consistency; see Example 3.3 (a) in Van de Geer (1993) for a nice exposition. To show the convergence rate, one can first obtain the cubic root rate (modulo the logarithm factor) in terms of the Hellinger distance, which means:

$$\sup_{\beta}\int\left(\sqrt{\hat{F}_n(u;\beta)}-\sqrt{F_0(u;\beta)}\right)^2 dG(u)=O_p(\log^2 n\times n^{-2/3}).$$

Then we translate it into the L_2 norm as in Lemma 3.1(i) of Groeneboom and Hendrickx (2018) or Proposition 2 of Balabdaoui, Groeneboom, and Hendrickx (2019) using the fact that $(\hat{F}_n-F_0)^2\leqslant 4\left(\sqrt{\hat{F}_n}-\sqrt{F_0}\right)^2$.

The following lemma is adapted from Lemma A.4 of Groeneboom, Jongbloed, and Witte (2010), which connects the rate of convergence of the piece-wise approximation $\bar{\phi}_{\alpha_0}$ to the convergence rate of the isotonic estimator.

Lemma S7. *Suppose Conditions 1 to 9 hold. For any u,β, we have*

$$|\bar{\phi}_{\alpha_0}(u)-\phi_{\alpha_0}(u)|\leqslant L|\hat{F}_n(\cdot;\beta)-F_0(\cdot)|, \tag{S.41}$$

for a finite positive constant L.

Given the above lemmas, we get the following characterization of I_{2n}^a.

Lemma S8. *Suppose Conditions 1 to 9 hold, then we have the following linear representation:*

$$\sqrt{n}I_{2n}^a = \sqrt{n}P\left[F_0 - \hat{F}_n(\cdot, \beta_0)\right] = \mathbb{G}_n\psi_{F_0} + o_p(1), \tag{S.42}$$

where ψ_{F_0} is defined in Theorem 3.1.

Proof. Given the characterization of the isotonic estimator $\hat{F}_n(u;\beta_0)$ and the piece-wise constant nature of $\bar{\phi}_{\alpha_0}$, we get

$$\int \bar{\phi}_{\alpha_0}[\hat{F}_n(u;\beta_0) - \delta_1]d\mathbb{P}_n = 0, \tag{S.43}$$

by equality (8.15) in Groeneboom and Jongbloed (2014). Therefore, starting with the representation of I_{2n}^a in Lemma S5 we get

$$I_{2n}^a = \int \bar{\phi}_{\alpha_0}(\hat{F}_n(u;\beta_0) - \delta_1)d(\mathbb{P}_n - P) \tag{S.44}$$

$$+ \int [\bar{\phi}_{\alpha_0} - \phi_{\alpha_0}](\hat{F}_n(u;\beta_0) - \delta_1)dP(u,\delta_1). \tag{S.45}$$

In the next lemma, we show that

$$\int \bar{\phi}_{\alpha_0}(\hat{F}_n(u;\beta_0) - \delta_1)d(\mathbb{P}_n - P) = \int \phi_{\alpha_0}(F_0(u) - \delta_1)d(\mathbb{P}_n - P) + o_p(n^{-1/2}), \tag{S.46}$$

and

$$\int [\bar{\phi}_{\alpha_0} - \phi_{\alpha_0}](\hat{F}_n(u;\beta_0) - \delta_1)dP(u,\delta_1) = o_p(n^{-1/2}), \tag{S.47}$$

which lead to the desired conclusion. □

Lemma S9. *Suppose Conditions 1 to 9 hold, then the following hold:*

$$R_n \equiv \int \bar{\phi}_{\alpha_0}(\hat{F}_n(u;\beta_0) - F_0(u))d(\mathbb{P}_n - P) = o_p(n^{-1/2}),$$

and

$$S_n \equiv \int [\bar{\phi}_{\alpha_0} - \phi_{\alpha_0}](\hat{F}_n(u;\beta_0) - \delta_1)dP(u,\delta_1) = o_p(n^{-1/2}). \tag{S.48}$$

Proof. We first handle the term S_n as follows.

$$\begin{aligned} S_n &= \int [\bar{\phi}_{\alpha_0} - \phi_{\alpha_0}](\hat{F}_n(u;\beta_0) - F_0(u))dG(u) \\ &\lesssim \| \hat{F}_n(u;\beta_0) - F_0(u) \|_2^2 = O_p(n^{-2/3} \times \log^2 n), \end{aligned} \tag{S.49}$$

where the second line uses Lemmas S7 and S6.

Referring to the term R_n, we introduce some notations adapted from the proof of Lemma 7 in Groeneboom, Jongbloed, and Witte (2010). Define

$$\xi_B(u) = \bar{\phi}_{\alpha_0}(u)B(u), \tag{S.50}$$

where $B \in \mathcal{B}_M$, the class of functions of bounded variation and with the supremum norm M. Let

$$\mathcal{G}_C \equiv \{\xi_B(u) : B \in \mathcal{B}_M\}. \tag{S.51}$$

By Lemma S6, for any small $\gamma > 0$ we can find a finite constant term C such that for all n sufficiently large:

$$\Pr\{\Upsilon_{n,C}\} \equiv \Pr\{\sup_{\beta} \| \hat{F}_n(u;\beta) - F_0(u;\beta) \|_2 \leqslant Cn^{-1/3} \log n\} \geqslant 1 - \gamma/2.$$

Now for an vanishing sequence ν_n, we have

$$\begin{aligned} \Pr\{|n^{1/2}R_n| > \nu_n\} &= \Pr\{|n^{1/2}R_n| > \nu_n \cap \Upsilon_{n,C}\} + \Pr\{|n^{1/2}R_n| > \nu_n \cap \Upsilon_{n,C}^c\} \\ &\leqslant \nu_n^{-1}\mathbb{E}\left[|n^{1/2}R_n| 1\{\Upsilon_{n,C}\}\right] + \gamma/2, \end{aligned}$$

for any small γ. Again by Lemma S6, we have

$$\begin{aligned} \mathbb{E}\left[|n^{1/2}R_n| 1\{\Upsilon_{n,C}\}\right] &\leqslant \mathbb{E} \sup_{B \in \mathcal{B}_C} \left| n^{1/2-1/3} \log n \int \bar{\phi}_{\alpha_0}(u)B(u)d(\mathbb{P}_n - P) \right| \\ &\leqslant n^{-1/3} \log n \mathbb{E} \sup_{\xi \in \mathcal{G}_C} \left| \int \xi(u) d\mathbb{G}_n(u) \right|. \end{aligned}$$

The rest of our proof uses Theorem 2.14.1 in Van Der Vaart and Wellner (1996) to bound the expectation in the last display. Following the construction in Groeneboom, Jongbloed, and Witte (2010), the entropy integral of $\mathcal{G}_C$ is bounded above by a finite constant, i.e., $J(1, \mathcal{G}_C) < \infty$. The L_2-norm of the envelope function is also bounded. Then applying

(S.25) yields

$$\mathbb{E}|R_n| \leqslant n^{-5/6}\log n \times \mathbb{E}\sup_{\xi\in\mathcal{G}_C}\left|\int \xi(u)d\mathbb{G}_n(u)\right| \lesssim n^{-5/6}\log n, \tag{S.52}$$

which immediately leads to $R_n = o_p(n^{-1/2})$.

□

Lemma S10. *Suppose Conditions 1 to 9 hold. We characterize the following smaller order term:*

$$\sqrt{n}I_{3n}^b = o_p(1). \tag{S.53}$$

Proof. Recall that

$$I_{3n}^b = P\left[\hat{F}_n(X'\hat{\beta}_n + \hat{\alpha}_n;\hat{\beta}_n) - F_0(X'\hat{\beta}_n + \hat{\alpha}_n;\hat{\beta}_n)\right] - P\left[\hat{F}_n(X'\beta_0 + \alpha_0;\beta_0) - F_0(X'\beta_0 + \alpha_0)\right].$$

Following the arguments in Lemma S8, we get

$$I_{3n}^b = \int\left[\left(\frac{g_0(u-\hat{\alpha}_n;\hat{\beta}_n)}{g_0(u;\hat{\beta}_n)}[F_0(u;\hat{\beta}_n)-\delta_1]\right) - \left(\frac{g_0(u-\alpha_0)}{g_0(u)}[F_0(u)-\delta_1]\right)\right]d(\mathbb{P}_n - P) \tag{S.54}$$
$$+ o_p(n^{-1/2}).$$

The smoothness assumption in Condition 8 implies that the function in the bracket of (S.54) belongs to a P-Donsker class by Example 19.7 in Van Der Vaart (1998). The convergence of $\hat{\alpha}_n$ and $\hat{\beta}_n$ leads to the desired conclusion that $\sqrt{n}I_{3n}^b = o_p(1)$. □

We then prove Theorem 3.2. Consider the general exchangeable bootstrap weights $M_n = (M_{n1},\ldots,M_{nn})'$ as in Section 3.6 of Van Der Vaart and Wellner (1996). Commonly used exchangeable bootstrap schemes include: (i) nonparametric bootstrap in which the weights M_n follow the multinomial distribution Multi $(n,(n^{-1},\ldots,n^{-1}))$; (ii) Bayesian bootstrap in which the bootstrap weights $M_{ni} = \omega_i/\sum_{i=1}^n \omega_i$ for $i = 1,\ldots,n$ and ω_i has the unit exponential distribution (Rubin, 1981); and (iii) Delete-h jackknives in which the bootstrap weights are generated from permuting the deterministic weights $w_{ni} = n/(n-h)$ for $i = 1,\ldots,n-h$ and $M_{nj} = w_{nR_n(j)}$ where R_n is a random permutation uniformly over $\{1,\ldots,n\}$ (Wu, 1990). We use the following notations: $\mathbb{P}_n^* f = n^{-1}\sum_{i=1}^n M_{ni} f(Z_i)$, and $\mathbb{G}_n^* f = n^{-1/2}\sum_{i=1}^n (M_{ni}-1) f(Z_i)$, where $Z = (Y, X)$. To take into account the joint randomness from the observed data and the bootstrap weights, we consider the underlying product probability space $(\mathcal{Z}^\infty \times \mathcal{M}, \mathcal{A}^\infty \times \Omega, P_{ZM})$. Furthermore, the bootstrap weights are independent of the sample observations, i.e., $P_{ZM} = P_Z \times P_M$.

Lemma S11 (Lemma 3.6.7 in Van Der Vaart and Wellner (1996))**.** *Let $Z_{n1}, \ldots, Z_{nn}$ be arbitrary stochastic processes and $(M_{n1}, \ldots, M_{nn})'$ be any exchangeable random vector independent of $Z_{n1}, \ldots, Z_{nn}$. For any $n_0 > 0$ and $n > n_0$, we have*

$$\begin{aligned}\mathbb{E}_{ZM}\left(\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} M_{ni}Z_{ni}\right\|\right) \leqslant\; & n_0\mathbb{E}_Z\left(\|Z_{n1}\|\right)\left(\frac{\mathbb{E}_M\left(\max_{1\leqslant i\leqslant n}|M_{ni}|\right)}{\sqrt{n}}\right) \\ & + \left(\int_0^\infty \sqrt{P_M\left(M_{n1} \geqslant u\right)}du\right)\mathbb{E}_Z\left(\left|\max_{n_0<i\leqslant n}\left\|\frac{1}{\sqrt{n}}\sum_{j=n_0+1}^{i} Z_{nj}\right\|\right|\right).\end{aligned}$$

When it comes to proving the stochastic equicontinuity related to the bootstrap version, applying the multiplier inequality in Lemma S11 to $\mathcal{G}_\epsilon$, we get

$$\mathbb{E}_{ZM} \,|||\, \mathbb{G}_n^* \,|||\, \lesssim \mathbb{E}_Z|G_\epsilon|\frac{1}{\sqrt{n}}\mathbb{E}_M\left|\max_i M_{ni}\right| + \mathbb{E}_Z\left|\max_{n_0\leqslant k\leqslant n} \| \,\mathbb{G}_k\, \|\right|, \tag{S.55}$$

where G_ϵ is the corresponding envelope function. The first term is of smaller order since

$$\frac{1}{\sqrt{n}}\mathbb{E}_M\left|\max_i M_{ni}\right| = o_p(1),$$

under our assumptions on the bootstrap weights. Meanwhile, the Levy inequality (Proposition A.1.2 of Van Der Vaart and Wellner (1996)) implies:

$$\Pr\{\max_{k\leqslant n} \| \,\mathbb{G}_k\, \| > \lambda\} \leqslant 2\Pr\{\| \,\mathbb{G}_n\, \| > \lambda\},$$

which makes the second term negligible. In analogous to Lemma S6, the bootstrapped isotonic estimator and its jump locations satisfy

$$\begin{aligned}\sup_{\beta\in\mathcal{B}} \| \,\hat{F}_n^*(u;\beta) - F_0(u;\beta)\, \|_2 &= O_p\left(\log n \times n^{-1/3}\right), \\ \sup_{\beta\in\mathcal{B}} \| \,\hat{A}_n^*(u;\beta) - u\, \|_2 &= O_p\left(\log n \times n^{-1/3}\right),\end{aligned}$$

in P_Z-probability.

Proof of Theorem 3.2. The bootstrap validity of $\hat{\beta}_{n-}^*$ has been shown in Groeneboom and Hendrickx (2017)[p3465, equation (4.19)]. Here we focus on $\hat{\alpha}_n^*$. To prove its conditional weak convergence, we start with the bootstrap estimating equation and decompose it into

$$\frac{1}{n}\sum_{i=1}^{n} M_{ni}\left[1-\hat{F}_n^*(X_i'\hat{\beta}_n^*+\hat{\alpha}_n^*;\hat{\beta}_n^*)-\Delta_{3i}\right]=I_{1n}^*+I_{2n}^*+I_{3n}^*, \tag{S.56}$$

where

$$I_{1n}^* = \frac{1}{n}\sum_{i=1}^{n} M_{ni}\left[1-F_0(X_i'\beta_0+\alpha_0)-\Delta_{3i}\right], \tag{S.57}$$

$$I_{2n}^* = \frac{1}{n}\sum_{i=1}^{n} M_{ni}\left[F_0(X_i'\beta_0+\alpha_0)-\hat{F}_n^*(X_i'\beta_0+\alpha_0;\beta_0)\right], \tag{S.58}$$

$$I_{3n}^* = \frac{1}{n}\sum_{i=1}^{n} M_{ni}\left[\hat{F}_n^*(X_i'\beta_0+\alpha_0;\beta_0)-\hat{F}_n^*(X_i'\hat{\beta}_n^*+\hat{\alpha}_n^*;\hat{\beta}_n^*)\right]. \tag{S.59}$$

The general scheme is analogous to our proof of Theorem 3.1. First of all, note that $I_{1n}^* = O_{p_M}(n^{-1/2})$ in P_Z-probability. Referring to I_{2n}^*, we get $I_{2n}^* = I_{2n}^{*a}+I_{2n}^{*b}$ where

$$I_{2n}^{*a} = P\left[F_0(U+\alpha_0)-\hat{F}_n^*(U+\alpha_0;\beta_0)\right] \quad \text{and} \quad I_{2n}^{*b} = (\mathbb{P}_n^*-P)\left[F_0(U+\alpha_0)-\hat{F}_n^*(U+\alpha_0;\beta_0)\right]. \tag{S.60}$$

We shall utilize P-Donsker property (Van Der Vaart and Wellner, 1996) to show $I_{2n}^{*b} = o_{p_M}(n^{-1/2})$ as in Lemma S4. We also have the following linear representation as in Lemma S8:

$$\sqrt{n}I_{2n}^{*a} = \sqrt{n}P\left[F_0-\hat{F}_n^*(\cdot,\beta_0)\right] = \sqrt{n}\left[\mathbb{P}_n^*-P\right]\psi_{F_0}+o_{p_M}(1), \tag{S.61}$$

in P_Z-probability.

When it comes to I_{3n}^*, we decompose it into three terms: $I_{3n}^* = I_{3n}^{*a}+I_{3n}^{*b}+I_{3n}^{*c}$, where

$$I_{3n}^{*a} = P\left[F_0(X'\hat{\beta}_n^*+\hat{\alpha}_n^*;\hat{\beta}_n^*)-F_0(X'\beta_0+\alpha_0)\right], \tag{S.62}$$

$$I_{3n}^{*b} = P\left[\hat{F}_n^*(X'\hat{\beta}_n^*+\hat{\alpha}_n^*;\hat{\beta}_n^*)-\hat{F}_n^*(X'\beta_0+\alpha_0;\beta_0)-F_0(X'\hat{\beta}_n^*+\hat{\alpha}_n^*;\hat{\beta}_n^*)+F_0(X'\beta_0+\alpha_0)\right],$$

$$I_{3n}^{*c} = (\mathbb{P}_n^*-P)\left[\hat{F}_n^*(X'\hat{\beta}_n^*+\hat{\alpha}_n^*;\hat{\beta}_n^*)-\hat{F}_n^*(X'\beta_0+\alpha_0;\beta_0)\right].$$

Similar to Lemma S10 and Lemma S4, we have $I_{3n}^{*b} = o_p(n^{-1/2})$ and $I_{3n}^{*c} = o_p(n^{-1/2})$ by the P-Donsker property of the corresponding functional classes. We also have the following expansion:

$$I_{3n}^{*a} = V_{\alpha_0}(\hat{\alpha}_n^*-\alpha_0)+V_{\beta_0}(\hat{\beta}_{n-}^*-\beta_{0-})+o_p(n^{-1/2}+\hat{\alpha}_n^*-\alpha_0+|\hat{\beta}_n^*-\beta_0|). \tag{S.63}$$

By taking the difference of the linear representations for $\hat{\alpha}_n^*$ and $\hat{\alpha}_n$, we get

$$\sqrt{n}\left(\hat{\alpha}_n^* - \hat{\alpha}_n\right) = V_{\alpha_0}^{-1}\mathbb{G}_n^*\left[\psi_0 + \psi_{F_0} + \psi_{\beta_0}\right] + o_p(1). \tag{S.64}$$

The desired result follows from Theorem 3.6.13 in Van Der Vaart and Wellner (1996). □

S4 Technical Proofs Related to Joint Estimation

This section presents and proves the technical lemmas and other results that are used in the proof of Theorem 3.3. We also present the proof of Theorem 3.4.

Related to the P-Glivenko-Cantelli or P-Donsker property, it is more convenient to work with the bracketing entropy bounds. For that purpose, we collect the necessary definitions from Van Der Vaart and Wellner (1996) as follows. The bracketing number $N_{[]}\left(\epsilon, \mathcal{F}, \|\cdot\|_2\right)$ for subclass $\mathcal{F}$ is defined to be the minimum of m such that $\exists\ f_1^L, f_1^U, \ldots, f_m^L, f_m^U$ for $\forall f \in \mathcal{F}$, $f_j^L \leqslant f \leqslant f_j^U$ for some j, and $\left\|f_j^U - f_j^L\right\|_2 \leqslant \epsilon$. Denote $H_{[]}\left(\epsilon, \mathcal{F}, \|\cdot\|_2\right) \equiv \log N_{[]}\left(\epsilon, \mathcal{F}, \|\cdot\|_2\right)$.

We need a few more notations to establish the consistency and rate of convergence. Let q_0 be the true density function and q be a member of the class of densities, denoted by $\mathcal{P}$. We define

$$\bar{q} \equiv \frac{q + q_0}{2},\ \bar{\mathcal{P}} \equiv \{\bar{q} : q \in \mathcal{P}\},\ \text{and } \bar{\mathcal{P}}^{1/2} \equiv \{\bar{q}^{1/2} : \bar{q} \in \bar{\mathcal{P}}\}. \tag{S.65}$$

In addition, we denote a ball (measured according to the Hellinger distance) around the true density q_0, intersected with $\bar{\mathcal{P}}^{1/2}$, by

$$\bar{\mathcal{P}}^{1/2}(\delta) \equiv \{\bar{q}^{1/2} \in \bar{\mathcal{P}}^{1/2} : \mathbf{h}(\bar{q}, q_0) \leqslant \delta\},$$

for some small positive δ. We refer to $H_{[]}(u, \bar{\mathcal{P}}^{1/2}(\delta))$ as the local entropy with bracketing and its corresponding entropy integral is given by

$$J_{[]}(\delta, \bar{\mathcal{P}}^{1/2}(\delta)) \equiv \int_{\delta^2/2^{13}}^{\delta} H_{[]}(u, \bar{\mathcal{P}}^{1/2}(\delta))du \vee \delta. \tag{S.66}$$

Similar to Geskus and Groeneboom (1997), we consider the following density functions:

$$q_F(u, \delta_1, \delta_2, \delta_3; \alpha, \beta) = \delta_1 F(u; \alpha, \beta) + \delta_2\{F(u+\alpha; \alpha, \beta) - F(u; \alpha, \beta)\} + \delta_3\{1 - F(u+\alpha; \alpha, \beta)\}. \tag{S.67}$$

Recall the notations in the proof of Theorem 3.3: $q_{0,\theta}$ denotes the density q_F for given $\theta = (\alpha, \beta_-')'$ and $F = F_0$, and $\tilde{q}_{n,\theta}$ denotes the maximum likelihood estimator for given θ.

The Hellinger distance $\mathbf{h}$ between the two densities is given by

$$
\begin{aligned}
2\mathbf{h}^2(\tilde{q}_{n,\theta}, q_{0,\theta}) = & \int_{\delta_1=1} \left(\tilde{F}_n^{1/2}(u;\theta) - F_0^{1/2}(u;\theta)\right)^2 dQ \\
& + \int_{\delta_2=1} \left((\tilde{F}_n(u+\alpha;\theta) - \tilde{F}_n(u;\theta))^{1/2} - (F_0(u+\alpha;\theta) - F_0(u;\theta))^{1/2}\right)^2 dQ \\
& + \int_{\delta_3=1} \left((1-\tilde{F}_n(u+\alpha;\theta))^{1/2} - (1-F_0(u+\alpha;\theta))^{1/2}\right)^2 dQ. \qquad \text{(S.68)}
\end{aligned}
$$

We record Theorem 7.4 in Van de Geer (2000) which establishes a general result for the convergence rate of the maximum likelihood estimator. We are going to apply it to our context. In Lemma S12, q_0 denotes the true density and $\tilde{q}_n$ denotes the maximum likelihood estimator.

Lemma S12 (Theorem 7.4 in Van de Geer (2000))**.** *Take $\Psi(\delta) \geqslant J_{[]}(\delta, \bar{\mathcal{P}}^{1/2}(\delta))$ (the entropy integral function defined in (S.66)) in such a way that $\Psi(\delta)/\delta^2$ is a non-increasing function of δ. Then for a universal constant c, and for $\sqrt{n}\delta_n^2 \geqslant c\Psi(\delta_n)$, we have for all $\delta \geqslant \delta_n$*

$$
\Pr\left\{\mathbf{h}(\tilde{q}_n, q_0) \geqslant \delta\right\} \leqslant c \exp\left(-n\delta^2/c^2\right).
$$

Lemma S12 implies that the maximum likelihood estimator $\tilde{q}_n$ converges at the rate $O_p(\delta_n)$. The next lemma delivers the rate of convergence for the NPMLE in terms of L_2 norm uniformly over the finite dimensional parameter.

Lemma S13. *Regarding the convergence by L_2-norm, we have*

$$
\sup_{\theta} \parallel \tilde{F}_n(\alpha + x'\beta;\theta) - F_0(\alpha + x'\beta;\theta) \parallel_2 = O_p\left(\log^2 n \times n^{-1/3}\right). \qquad \text{(S.69)}
$$

Proof. In order to obtain the rate of convergence, we need to first bound the entropy number for the likelihood function:

$$
F^{\Delta_{1i}}(X_i'\beta;\theta) \times \left(F(X_i'\beta + \alpha;\theta) - F(X_i'\beta;\theta)\right)^{\Delta_{2i}} \times \left(1 - F(X_i'\beta + \alpha;\theta)\right)^{\Delta_{3i}}.
$$

The only complication comes from the term

$$
\mathcal{F}_D \equiv \left\{\sqrt{F(x'\beta + \alpha;\theta) - F(x'\beta;\theta)} : (\alpha, \beta, F)\right\}. \qquad \text{(S.70)}
$$

Its entropy number can be bounded in the following way; see Example 3.3(b) of Van de

Geer (1993). If $F(x'\beta+\alpha;\theta)-F(x'\beta;\theta)>\delta$ or $\bar{F}(x'\bar{\beta}+\bar{\alpha};\bar{\theta})-\bar{F}(x'\bar{\beta};\bar{\theta})>\delta$, then

$$\begin{aligned}&\left|\sqrt{F(x'\beta+\alpha;\theta)-F(x'\beta;\theta)}-\sqrt{\bar{F}(x'\bar{\beta}+\bar{\alpha};\bar{\theta})-\bar{F}(x'\bar{\beta};\bar{\theta})}\right|\\&<\frac{1}{\sqrt{\delta}}\left\{\left|F(x'\beta+\alpha;\theta)-\bar{F}(x'\bar{\beta}+\bar{\alpha};\bar{\theta})\right|+\left|F(x'\beta;\theta)-\bar{F}(x'\bar{\beta};\bar{\theta})\right|\right\}.\end{aligned}$$

If both $F(x'\beta+\alpha;\theta)-F(x'\beta;\theta)\leqslant\delta$ or $\bar{F}(x'\bar{\beta}+\bar{\alpha};\bar{\theta})-\bar{F}(x'\bar{\beta};\bar{\theta})\leqslant\delta$, obviously one has

$$\left|\sqrt{F(x'\beta+\alpha;\theta)-F(x'\beta;\theta)}-\sqrt{\bar{F}(x'\bar{\beta}+\bar{\alpha};\bar{\theta})-\bar{F}(x'\bar{\beta};\bar{\theta})}\right|\leqslant 2\sqrt{\delta}.$$

In sum, for any probability measure Q, we get

$$N(4\sqrt{\delta},\mathcal{F}_D,L_2(Q))\leqslant N(\delta,\mathcal{F}_0,L_2(Q)),$$

where $\mathcal{F}_0\equiv\{F(x'\beta+\alpha;\theta):(\alpha,\beta,F)\}$. Compared with the calculation in Van de Geer (1993), one needs to account for the presence of finite dimensional parameter, which incurs an additional $\log n$ factor. Therefore, one can apply Lemma S12 to get

$$\sup_{\theta}\mathbf{h}(\tilde{q}_{n,\theta},q_{0,\theta})=O_p(\log^2 n\times n^{-1/3}). \tag{S.71}$$

Also, note that

$$(\tilde{F}_n-F_0)^2\leqslant 4\left(\sqrt{\tilde{F}_n}-\sqrt{F_0}\right)^2\quad\text{and}\quad(\tilde{F}_n-F_0)^2\leqslant 4\left(\sqrt{1-\tilde{F}_n}-\sqrt{1-F_0}\right)^2,$$

we get that

$$\sup_{\theta}\parallel\tilde{F}_n(\cdot;\theta)-F_0(\cdot;\theta)\parallel_2=O_p(\log^2 n\times n^{-1/3}).$$

□

In the following, we provide some details for our proof of Theorem 3.3. Specifically, we show the existence of zero-crossing points and the stochastic equicontinuity of negligible terms related to our estimating equations.

Existence of the zero-crossing point of Φ_n. Recall that $\theta=(\alpha,\beta_-')'$. The first coordinate of β is normalized to be 1; i.e., the overall number of unknown parameters is equal to K. The uniform convergence of the estimating equation leads to

$$\Phi_n(\theta)=\dot{\Phi}_{\theta_0}(\theta-\theta_0)+r_n(\theta), \tag{S.72}$$

where $r_n(\theta) = o_p(1) + o(\theta - \theta_0)$. We define for $h > 0$

$$\Phi_{n,h}(\theta) \equiv \dot{\Phi}_{\theta_0}(\theta - \theta_0) + \tilde{r}_{n,h}(\theta), \tag{S.73}$$

with

$$\tilde{r}_{n,h}(\theta) = h^{-d} \int k_h(u_1 - \alpha) \cdots k_h(u_K - \beta_K) r_n(u_1, \cdots, u_K) du_1 \cdots du_K, \tag{S.74}$$

where $k(\cdot)$ is a standard kernel density function supported on $[-1, 1]$. Note that $\lim_{h \to 0} \tilde{r}_{n,h}(\theta) = r_n(\theta)$. We re-parameterize θ and θ_0 as

$$\gamma = \dot{\Phi}_{\theta_0}\theta, \text{ and } \gamma_0 = \dot{\Phi}_{\theta_0}\theta_0, \tag{S.75}$$

which gives

$$\Phi_{n,h}(\theta) = \gamma - \gamma_0 + \tilde{r}_{n,h}(\dot{\Phi}_{\theta_0}^{-1}\gamma). \tag{S.76}$$

Given the result in (S.72), the mapping

$$\gamma \mapsto \gamma_0 - \tilde{r}_n(\dot{\Phi}_{\theta_0}^{-1}\gamma) \tag{S.77}$$

maps, for each $\delta > 0$, the ball $B_\delta(\gamma_0) = \{\gamma : |\gamma - \gamma_0| \leqslant \delta\}$ into $B_{\delta/2}(\gamma_0) = \{\gamma : |\gamma - \gamma_0| \leqslant \delta/2\}$ with probability approaching to 1. Therefore by Brouwer's fixed point theorem (Groeneboom and Hendrickx, 2018), the mapping $\gamma \mapsto \gamma_0 - \tilde{r}_{n,h}(\dot{\Phi}_{\theta_0}^{-1}\gamma)$ has a fixed point which we denote by $\gamma_{n,h}$. Let $\theta_{n,h} \equiv \dot{\Phi}_{\theta_0}^{-1}\gamma_{n,h}$, then we have

$$\Phi_{n,h}(\theta_{n,h}) = 0. \tag{S.78}$$

By compactness of the parameter space, the sequence $(\theta_{n,1/k})_{k=1}^{\infty}$ must have a subsequence $(\theta_{n,1/k_l})$ with a limit point $\bar{\theta}_n$ as $l \to \infty$.

Finally, we prove that $\Phi_n(\theta)$ has a zero-crossing point at $\bar{\theta}_n$ by contradiction, following Groeneboom and Hendrickx (2018)([p.14] in their supplementary material). Suppose that the j-th component Φ_n^j of Φ_n does not have a crossing of zero at $\bar{\theta}_n$. Then there must be an open ball $B_\delta(\bar{\theta}_n) = \{\theta : |\theta - \bar{\theta}_n| < \delta\}$ of $\bar{\theta}_n$ such that Φ_n^j has a constant sign in $B_\delta(\bar{\theta}_n)$, say $\Phi_n^j(\theta) \geqslant c > 0$ for all $\theta \in B_\delta(\bar{\theta}_n)$ and some constant $c > 0$. Following the argument of Groeneboom and Hendrickx (2018), the j-th component of $\Phi_{n,h}^j$ of $\Phi_{n,h}$ satisfies

$$\Phi_{n,h}^j(\theta) \geqslant \frac{c}{2}, \tag{S.79}$$

for sufficiently small h and all $\theta \in B_\delta(\bar{\theta}_n)$, which contradicting (S.78), since $\theta_{n,h}$ for $h = 1/k_l$ belongs to $B_\delta(\bar{\theta}_n)$ for large k_l.

We also show that (α_0, β'_{0-}) is the unique zero-crossing point for the population level estimating equation. By the self-consistency of the NPMLE (Groeneboom and Wellner, 1992), we have $\mathbb{E}[\Delta_1 | X'\beta = u, X'\beta + \alpha = u + \alpha] = F(u; \alpha, \beta)$. It is clear that given the value of $X'\beta$, $X'\beta + \alpha$ does not provide additional information about Δ_1. Thereafter, we can apply Lemma 4.1 of Groeneboom and Hendrickx (2018) to show β_0 is the unique zero-crossing point of the first set of estimating equations associated with (Δ_1, X_{-1}). For the second estimating equation $\mathbb{E}[1 - \Delta_3 - F(X'\beta + \alpha; \alpha, \beta)]$, it is enough to fix $\beta = \beta_0$. Following the calculation in Lemma S21, it is easy to see that the estimating equation is monotone with respect to α for fixed β. Hence, α_0 is the unique zero-crossing point.

□

Lemma S14. *Under Conditions 1 to 10, we have*

$$(\mathbb{P}_n - P)\zeta(Z; \tilde{\alpha}_n, \tilde{\beta}_n, \tilde{F}_n(\cdot; \tilde{\alpha}_n, \tilde{\beta}_n)) = \mathbb{P}_n \zeta(Z; \alpha_0, \beta_0, F_0) + o_p(n^{-1/2}),$$

and

$$\begin{aligned} &P[\zeta(Z; \tilde{\alpha}_n, \tilde{\beta}_n, \tilde{F}_n(\cdot; \tilde{\alpha}_n, \tilde{\beta}_n)) - \zeta(Z; \alpha_0, \beta_0, \tilde{F}_n(\cdot; \alpha_0, \beta_0))] \\ &= P[\zeta(Z; \tilde{\alpha}_n, \tilde{\beta}_n, F_0(\cdot; \tilde{\alpha}_n, \tilde{\beta}_n)) - \zeta(Z; \alpha_0, \beta_0, F_0(\cdot))] + o_p(n^{-1/2}). \end{aligned}$$

Proof. The proof essentially follows from Lemma S3 by the stochastic equicontinuity of the related P-Donsker classes and the consistency of $\tilde{\alpha}_n$ and $\tilde{\beta}_n$. The only change applies to the functional class that the NPMLE $\tilde{F}_n$ belongs to, given that the NPMLE is a sub-distribution or a defective distribution. However, the entropy bound for the monotone functions in Lemma S2 does not depend on the range of the function, as long as it is finite. Hence, the desired results follow. □

The remaining proofs characterize the asymptotic property of the linear functional for the NPMLE given the data and the true unknown parameter θ_0. We denote the empirical probability measure of the ordered response data by Q_n and its population version by Q_{F_0}, where the distribution is set to be the true unknown F_0. We mentioned in the main text that the NPMLE could be a defective distribution (or a sub-distribution) in finite samples; i.e., $\tilde{F}_n(u) < 1$ for any u in the support. However, this plays a minor role regarding the large sample properties because the defectiveness does not occur with probability 1 as $n \to \infty$.

Lemma S15 (Proposition 1 in Geskus and Groeneboom (1997)). *We have*

$$\lim_{n\to\infty} \Pr\{\tilde{F}_n \text{ is defective}\} = 0.$$

We record the following lemma from Corollary 1 in Geskus and Groeneboom (1997), which characterizes the NPMLE $\tilde{F}_n$.

Lemma S16. *Any function σ that is constant at the same intervals as $\tilde{F}_n$ satisfies*

$$\int \sigma(u)\left[\frac{\delta_1}{\tilde{F}_n(u)} - \frac{1-\delta_1-\delta_2}{1-\tilde{F}_n(u+\alpha_0)} + \frac{\delta_2}{\tilde{F}_n(u+\alpha_0)-\tilde{F}_n(u)}\right] dQ_n(u,\delta_1,\delta_2) = 0. \tag{S.80}$$

We draw on Geskus and Groeneboom (1996, 1997, 1999), where the authors developed a systematic approach to characterize the linear functional of NPMLE for the interval censored data (case 2). For that purpose, we define $c_1(u) = \int_{C_L}^{u} \mathbb{E}[X_{-1}|v]g_0(v)dv$, $c_3(u) = G_0(u-\alpha_0)$, and $c(u) = (c_1'(u), c_3(u))'$. Consider the linear functional $\kappa(F_0) = \int c(v)dF_0(v)$ and its canonical (with zero mean) gradient

$$\tilde{\kappa}_F(u) = c(u) - \int c(v)dF(v). \tag{S.81}$$

A key component in determining the asymptotic property of $(\tilde{\alpha}_n, \tilde{\beta}_n)$ is $\kappa_{\tilde{F}_n(\cdot;\alpha_0,\beta_0)}$; i.e., the linear functional of the NPMLE when the finite dimensional parameter is set to be its true value. The influence function of the latter one crucially depends on whether there is a unique element ϕ_F satisfying

$$L^*\phi_F = \tilde{\kappa}_F, \tag{S.82}$$

given the differentiability of $\tilde{\kappa}_F$ in the sense of Van der Vaart (1991), where L^* denotes the adjoint operator of L defined in equation (S.83). We further denote its derivative by $\dot{\tilde{\kappa}}_F$.

To present the solution ϕ_F, we denote $u = x'\beta_0$ and the support of it as $[C_L, C_U]$; see Coppejans (2007). For any function a in the tangent set, the score operator for the nonparametric component is

$$L[a](u,\delta_1,\delta_2) = \frac{\delta_1\int_{C_L}^{u} a dF}{F(u)} + \frac{\delta_2\int_{u}^{u+\alpha_0} a dF}{F(u+\alpha_0)-F(u)} - \frac{(1-\delta_1-\delta_2)\int_{u+\alpha_0}^{C_U} a dF}{1-F(u+\alpha_0)}. \tag{S.83}$$

For any function $b(v,\delta_1,\delta_2)$, we also have the adjoint operator L^* specified as follows:

$$L^*[b](u) = \int_{u}^{C_U} b(v,1,0)g_0(v)dv + \int_{u-\alpha}^{u} b(v,0,1)g_0(v)dv + \int_{C_L}^{u-\alpha} b(v,0,0)g_0(v)dv. \tag{S.84}$$

Letting $\varsigma_F(u) \equiv \int_{C_L}^{u} a(v)dF(v)$ be the integrated score function, we have

$$L^*L[a](u) = \int_u^{C_U} \frac{\varsigma(v)}{F(v)} g_0(v)dv + \int_{u-\alpha_0}^{u} \frac{\varsigma(v+\alpha_0) - \varsigma(v)}{F(v+\alpha_0) - F(v)} g_0(v)dv + \int_{C_L}^{u-\alpha_0} \frac{\varsigma(v+\alpha_0)}{1 - F(v+\alpha_0)} g_0(v)dv. \tag{S.85}$$

Following equation (4) in Geskus and Groeneboom (1997) and Example 4.2 in Van de Geer (1995), the solution to (S.82) can be written as

$$\phi_F(u, \delta_1, \delta_2) = \delta_1 \frac{\varsigma_F(u)}{F(u)} + \delta_2 \frac{\varsigma_F(u+\alpha_0) - \varsigma_F(u)}{F(u+\alpha_0) - F(u)} - (1 - \delta_1 - \delta_2) \frac{\varsigma_F(u+\alpha_0)}{1 - F(u+\alpha_0)}, \tag{S.86}$$

where

$$\varsigma_F(u) = \begin{cases} -F(u)\left[(1 - F(u))\omega(u) + (1 - F(u+\alpha_0))\omega(u+\alpha_0)\right], & \text{for } C_L \leqslant u \leqslant \alpha_0 \\ (1 - F(u))\left[F(u)\omega(u) + F(u-\alpha_0)\omega(u-\alpha_0)\right], & \text{for } \alpha_0 \leqslant u \leqslant C_U, \end{cases}$$

and $\omega(u) \equiv \frac{\dot{c}(u)}{g_0(u)}$.

We consider the piece-wise approximation of $\varsigma_{\tilde{F}_n}$, which is constant on the same intervals where the NPMLE $\tilde{F}_n(\cdot; \alpha, \beta)$ remains constant. Denote those intervals by $J_i = [\tau_i, \tau_{i+1})$. We define

$$\bar{\varsigma}_{\tilde{F}_n}(u) = \varsigma_{\tilde{F}_n}(\tilde{A}_n(u; \alpha, \beta)), \tag{S.87}$$

where

$$\tilde{A}_n(u; \alpha, \beta) = \begin{cases} \tau_i, & \text{if } \forall t \in J_i : F_0(t) > \tilde{F}_n(\tau_i; \alpha, \beta), \\ s, & \text{if } \exists s \in J_i : F_0(s) = \tilde{F}_n(s; \alpha, \beta), \\ \tau_{i+1}, & \text{if } \forall t \in J_i : F_0(t) < \tilde{F}_n(\tau_i; \alpha, \beta), \end{cases}$$

for $u \in J_i$.

We define the function ξ_F as

$$\xi_F(u) = \frac{\varsigma_F(u)}{F(u)(1 - F(u))}. \tag{S.88}$$

We also consider the piece-wise constant version of $\xi_{\tilde{F}_n}$, denoted by $\bar{\xi}_{\tilde{F}_n}$, which is defined by $\bar{\xi}_{\tilde{F}_n}(u) = \bar{\varsigma}_{\tilde{F}_n} / [\tilde{F}_n(u)(1 - \tilde{F}_n(u))]$. In addition, let $\bar{\phi}_{\tilde{F}_n}$ denote the function $\phi_{\tilde{F}_n}$ defined in (S.86), but with $\varsigma_{\tilde{F}_n}$ replaced by $\bar{\varsigma}_{\tilde{F}_n}$.

Lemma S17. *Under Conditions 1 to 10, we have*

$$\sqrt{n}(\kappa(\tilde{F}_n) - \kappa(F_0)) = \sqrt{n} \int \phi_{F_0} d(Q_n - Q_{F_0}) + o_p(1). \tag{S.89}$$

Proof. The proof of this result requires several intermediate lemmas that we present afterwards. Here we describe the crux of the arguments in four steps.

Step 1. The first step rewrites the effect from estimating the distribution using NPMLE in terms of its linear functional:

$$\sqrt{n}(\kappa(\tilde{F}_n) - \kappa(F_0)) = \sqrt{n} \int \tilde{\kappa}_{F_0} d(\tilde{F}_n - F_0). \tag{S.90}$$

Step 2. The second step is similar to the proof of Lemma S5 where we apply integration by parts. Now we have

$$\int \tilde{\kappa}_{F_0} d(\tilde{F}_n - F_0) = -\int \phi_{\tilde{F}_n} dQ_{F_0}, \tag{S.91}$$

by Lemma 1 in Geskus and Groeneboom (1997).

Step 3. We consider the piece-wise approximation $\bar{\phi}_{\tilde{F}_n}$ defined below (S.88). By Lemma S16, one gets

$$\int \bar{\phi}_{\tilde{F}_n} dQ_n = 0.$$

Thus, we have

$$-\int \phi_{\tilde{F}_n} dQ_{F_0} = -\int \bar{\phi}_{\tilde{F}_n} d(Q_n - Q_{F_0}) + o_p(n^{-1/2}), \tag{S.92}$$

in which we use

$$\int \left(\bar{\phi}_{\tilde{F}_n} - \phi_{\tilde{F}_n}\right) dQ_{F_0} = o_p(n^{-1/2}),$$

as proved in Lemma S20.

Step 4. In the last step, we proceed with the following decomposition

$$-\int \bar{\phi}_{\tilde{F}_n} dQ_{F_0} = -\int \phi_{F_0} d(Q_n - Q_{F_0}) + \int \left[\bar{\phi}_{\tilde{F}_n} - \phi_{F_0}\right] d(Q_n - Q_{F_0}). \tag{S.93}$$

We show that the second term on the r.h.s. of (S.93) is $o_p(n^{-1/2})$. Following the argument in the proof of Lemma 3 in Geskus and Groeneboom (1997) and using Lemma S18, ϕ_F and its piece-wise constant approximation $\bar{\phi}_F$ are of bounded variation. Therefore, one can show that the random entropy integral as a function of δ is of order $O_p(\delta^{1/2})$ for the functional class that includes $(\bar{\phi}_{\tilde{F}_n} - \phi_{F_0})$. Then by the uniform consistency of $\tilde{F}_n$, we get

$$\int \left(\bar{\phi}_{\tilde{F}_n} - \phi_{F_0}\right)^2 dQ_{F_0} \to 0,$$

with probability 1. As a result, $\int \left[\bar{\phi}_{\tilde{F}_n} - \phi_{F_0}\right] d(Q_n - Q_{F_0}) = o_p(n^{-1/2})$ follows from the stochastic equicontinuity of the related P-Donsker class. In the end, we arrive at

$$-\int \bar{\phi}_{\tilde{F}_n} dQ_{F_0} = -\int \phi_{F_0} d(Q_n - Q_{F_0}) + o_p(n^{-1/2}). \tag{S.94}$$

□

The following lemma states the Lipschitz property for ς_F and ξ_F. It can be shown by combining the closed-form expression of ς_F in (S.86) and the proof of Lemma 4 in Geskus and Groeneboom (1996).

Lemma S18. *The derivative of ς_F at the points of continuity is bounded, uniformly over F and the points of continuity; i.e.,*

$$|\varsigma_F(u) - \varsigma_F(v)| \leqslant C_1|u - v|, \tag{S.95}$$

for u and v in the same interval between jumps and for a finite positive constant C_1. The same holds for ξ_F. Moreover, the jumps satisfy

$$|\varsigma_F(u) - \varsigma_F(u-)| \leqslant C_2|F(u) - F(u-)|, \tag{S.96}$$

for a finite positive constant C_2. Again, the same holds for ξ_F.

The next lemma controls the approximation error for the function $\xi_{\tilde{F}_n}$. Its proof resembles the one for Lemma S7 and uses Lemma S18.

Lemma S19. *Suppose our Conditions hold, then we have*

$$\| \bar{\xi}_{\tilde{F}_n}(u) - \xi_{\tilde{F}_n}(u) \|_2 \lesssim \| \tilde{F}_n(u; \alpha_0, \beta_0) - F_0(u) \|_2 \, . \tag{S.97}$$

The following lemma characterizes a smaller order term in **Step 3** of the proof of Lemma S17 while analyzing $\sqrt{n}(\kappa(\tilde{F}_n) - \kappa(F_0))$.

Lemma S20. *Under our conditions, we have*

$$\int \left(\bar{\phi}_{\tilde{F}_n} - \phi_{\tilde{F}_n}\right) dQ_{F_0} = o_p(n^{-1/2}). \tag{S.98}$$

Proof. We start by defining the function φ_n as

$$\varphi_n(u) = -\left[\bar{\phi}_{\tilde{F}_n} - \phi_{\tilde{F}_n}\right](u,1,0)F_0(u) \\ -\left[\bar{\phi}_{\tilde{F}_n} - \phi_{\tilde{F}_n}\right](u,0,1)[F_0(u+\alpha_0) - F_0(u)] + \left[\bar{\phi}_{\tilde{F}_n} - \phi_{\tilde{F}_n}\right](u,0,0)[1 - F_0(u+\alpha_0)].$$

Then we obtain

$$\begin{aligned}\varphi_n(u) =& \frac{1-\tilde{F}_n(u)}{\tilde{F}_n(u+\alpha_0) - \tilde{F}_n(u)}\left(\bar{\xi}_{\tilde{F}_n}(u) - \xi_{\tilde{F}_n}(u)\right) \\ &\times \left[F_0(u+\alpha_0)(\tilde{F}_n(u) - F_0(u)) + F_0(u)(F_0(u+\alpha_0) - \tilde{F}_n(u+\alpha_0))\right] \\ &- \frac{\tilde{F}_n(u+\alpha_0)}{\tilde{F}_n(u+\alpha_0) - \tilde{F}_n(u)}\left(\bar{\xi}_{\tilde{F}_n}(u+\alpha_0) - \xi_{\tilde{F}_n}(u+\alpha_0)\right) \\ &\times \left[(1-F_0(u+\alpha_0))(\tilde{F}_n(u) - F_0(u)) + (1-F_0(u))(F_0(u+\alpha_0) - \tilde{F}_n(u+\alpha_0))\right].\end{aligned}$$

We apply the Cauchy-Schwarz inequality to get

$$\left|\int \left(\bar{\phi}_{\tilde{F}_n} - \phi_{\tilde{F}_n}\right) dQ_{F_0}\right| \leqslant C \parallel \bar{\xi}_{\tilde{F}_n} - \xi_{\tilde{F}_n} \parallel_2 \times \parallel \tilde{F}_n - F_0 \parallel_2 . \tag{S.99}$$

Following a similar argument as in the proof of Lemma S7, we get

$$|\bar{\xi}_{\tilde{F}_n}(u) - \xi_{\tilde{F}_n}(u)| \leqslant C|\tilde{F}_n(u) - F_0(u)|.$$

Now the result follows from Lemma S13. □

The following lemma computes the Hessian matrix related to our joint estimator.

Lemma S21. *Recall that the Hessian matrix is*

$$H(\alpha,\beta) \equiv \begin{pmatrix} \mathbb{E}[-X_{-1}\frac{\partial}{\partial\alpha}F(X'\beta;\alpha,\beta)] & \mathbb{E}[-X_{-1}\frac{\partial}{\partial\beta_-}F(X'\beta;\alpha,\beta)] \\ \mathbb{E}[-\frac{\partial}{\partial\alpha}F(X'\beta+\alpha;\alpha,\beta)] & \mathbb{E}[-\frac{\partial}{\partial\beta'_-}F(X'\beta+\alpha;\alpha,\beta)] \end{pmatrix},$$

then we have

$$H_0 \equiv H(\alpha_0,\beta_0) = -\begin{pmatrix} \mathbb{E}[(X_{-1} - \mathbb{E}[X_{-1}|X'\beta_0])f_0(X'\beta_0)] & \mathbb{E}[(X_{-1} - \mathbb{E}[X_{-1}|X'\beta_0])^{\otimes 2}f_0(X'\beta_0)] \\ \mathbb{E}[f_0(X'\beta_0+\alpha_0)] & \mathbb{E}[(X_{-1} - \mathbb{E}[X_{-1}|X'\beta_0])'f_0(X'\beta_0+\alpha_0)] \end{pmatrix}.$$

Proof. In order to avoid repetition, we only show that

$$\mathbb{E}[\frac{\partial}{\partial \beta_-'} F(X'\beta+\alpha;\alpha,\beta)]|_{\alpha=\alpha_0,\beta=\beta_0} = \mathbb{E}[(X_{-1}-\mathbb{E}[X_{-1}|X'\beta_0])' f_0(X'\beta_0+\alpha_0)].$$

First of all, we have

$$F(u;\alpha,\beta) \equiv \mathbb{E}[1-\Delta_3|X'\beta+\alpha=u] = \int F_0(u+x'(\beta_0-\beta)+\alpha_0-\alpha) f_{X|(X'\beta+\alpha)}(x|X'\beta+\alpha=u)dx.$$

Because the first slope coefficient is normalized to be 1, we denote the conditional density function of $(X_2,\cdots,X_K)$ given $X'\beta+\alpha=u$ by $h_\theta(\cdot|u)$. We make the following change of variable by taking $t_1 = x'\beta+\alpha$ and $t_j = x_j$ for $j=2,\cdots,K$.

Then we can write

$$F(x'\beta+\alpha;\alpha,\beta) = \int F_0\left((x'\beta+\alpha-\sum_{j=2}^K \beta_j\tilde{x}_j)+\alpha_0+\sum_{j=2}^K \beta_{0j}\tilde{x}_j\right) h_\theta(\tilde{x}_2,\cdots,\tilde{x}_K|x'\beta+\alpha)\Pi_{j=2}^K d\tilde{x}_j.$$

Now we take partial derivative w.r.t. β_j for $j=2,\cdots,K$:

$$\begin{aligned}
&\frac{\partial}{\partial \beta_j} F(x'\beta+\alpha;\alpha,\beta) \qquad\qquad (S.100)\\
&=\int (x_j-\tilde{x}_j) f_0\left((x'\beta+\alpha-\sum_{j=2}^K \beta_j\tilde{x}_j)+\alpha_0+\sum_{j=2}^K \beta_{0j}\tilde{x}_j\right) h_\theta(\tilde{x}_2,\cdots,\tilde{x}_K|x'\beta+\alpha)\Pi_{j=2}^K d\tilde{x}_j\\
&+\int F_0\left((x'\beta+\alpha-\sum_{j=2}^K \beta_j\tilde{x}_j)+\alpha_0+\sum_{j=2}^K \beta_{0j}\tilde{x}_j\right) \frac{\partial}{\partial \beta_j} h_\theta(\tilde{x}_2,\cdots,\tilde{x}_K|x'\beta+\alpha)\Pi_{j=2}^K d\tilde{x}_j.
\end{aligned}$$

The first term on the right-hand side of (S.100) is equal to $\mathbb{E}[(X_{-1}-\mathbb{E}[X_{-1}|X'\beta_0])' f_0(X'\beta_0+\alpha_0)]$. Because the function $h_\theta(\cdot|u)$ is a conditional density function that integrates to 1, the second term on the right-hand side of (S.100) is zero, when evaluated at $\theta=\theta_0$. Therefore, the desired result follows. □

We complete this section by proving the bootstrap consistency of our joint estimator.

Proof of Theorem 3.4. The overall structure of the proof is the same as the one for Theorem 3.3. The necessary change is that one has to apply the maximal inequality with multiplier bootstrap weights to the corresponding functional classes. To avoid repetition, we only outline the main steps. We skip the steps leading to the consistency of $(\tilde{\alpha}_n^*, \tilde{\beta}_n^*)$ and directly start with

$$\mathbb{P}_n^*\zeta(Z;\tilde{\alpha}_n^*,\tilde{\beta}_n^*,\tilde{F}_n^*(\cdot;\tilde{\alpha}_n^*,\tilde{\beta}_n^*)) = 0.$$

Then we proceed with

$$
\begin{aligned}
0 =&(\mathbb{P}_n^* - P)\zeta(Z; \tilde{\alpha}_n^*, \tilde{\beta}_n^*, \tilde{F}_n^*(\cdot; \tilde{\alpha}_n^*, \tilde{\beta}_n^*)) \\
&+P[\zeta(Z; \alpha_0, \beta_0, \tilde{F}_n^*(\cdot; \alpha_0, \beta_0)) - \zeta(Z; \alpha_0, \beta_0, F_0(\cdot))] \\
&+P[\zeta(Z; \tilde{\alpha}_n^*, \tilde{\beta}_n^*, \tilde{F}_n^*(\cdot; \tilde{\alpha}_n^*, \tilde{\beta}_n^*)) - \zeta(Z; \alpha_0, \beta_0, \tilde{F}_n^*(\cdot; \alpha_0, \beta_0))].
\end{aligned}
$$

Using arguments parallel to the proof of Theorem 3.3 (the part below equation (A.4)), we have

$$
\begin{aligned}
(\mathbb{P}_n^* - P)\zeta(Z; \tilde{\alpha}_n^*, \tilde{\beta}_n^*, \tilde{F}_n^*(\cdot; \tilde{\alpha}_n^*, \tilde{\beta}_n^*)) &= \mathbb{P}_n^*\zeta(Z; \alpha_0, \beta_0, F_0) + o_p(n^{-1/2}), \\
P[\zeta(Z; \alpha_0, \beta_0, \tilde{F}_n^*(\cdot; \alpha_0, \beta_0)) - \zeta(Z; \alpha_0, \beta_0, F_0(\cdot))] &= (\mathbb{P}_n^* - P)\phi_{F_0} + o_p(n^{-1/2}),
\end{aligned}
$$

and

$$
\begin{aligned}
&P[\zeta(Z; \tilde{\alpha}_n^*, \tilde{\beta}_n^*, \tilde{F}_n^*(\cdot; \tilde{\alpha}_n^*, \tilde{\beta}_n^*)) - \zeta(Z; \alpha_0, \beta_0, \tilde{F}_n^*(\cdot; \alpha_0, \beta_0))] \\
&= P[\zeta(Z; \tilde{\alpha}_n^*, \tilde{\beta}_n^*, F_0(\cdot; \tilde{\alpha}_n^*, \tilde{\beta}_n^*)) - \zeta(Z; \alpha_0, \beta_0, F_0(\cdot))] + o_p(n^{-1/2}) \\
&= H_0 \begin{pmatrix} \tilde{\alpha}_n^* - \alpha_0 \\ \tilde{\beta}_n^* - \beta_0 \end{pmatrix} + o_p(n^{-1/2} + (\tilde{\alpha}_n^* - \alpha_0) + |\tilde{\beta}_n^* - \beta_0|).
\end{aligned}
$$

In the end, we get

$$
H_0 \begin{pmatrix} \tilde{\alpha}_n^* - \alpha_0 \\ \tilde{\beta}_{n-}^* - \beta_{0-} \end{pmatrix} = \mathbb{P}_n^*\zeta(Z; \alpha_0, \beta_0, F_0) + (\mathbb{P}_n^* - P)\phi_{F_0} + o_p(n^{-1/2} + (\tilde{\alpha}_n^* - \alpha_0) + |\tilde{\beta}_n^* - \beta_0|),
$$

which leads to

$$
\begin{pmatrix} \tilde{\alpha}_n^* - \tilde{\alpha}_n \\ \tilde{\beta}_{n-}^* - \tilde{\beta}_{n-} \end{pmatrix} = H_0^{-1} \left[(\mathbb{P}_n^* - \mathbb{P}_n)\zeta(Z; \alpha_0, \beta_0, F_0) + (\mathbb{P}_n^* - \mathbb{P}_n)\phi_{F_0}\right] + o_p(n^{-1/2}).
$$

Thereafter, the desired asymptotic normality follows from Theorem 3.6.13 in Van Der Vaart and Wellner (1996). □

References

Balabdaoui, F., P. Groeneboom, and K. Hendrickx (2019): "Score estimation in the monotone single-index model," *Scandinavian Journal of Statistics*, 46(2), 517–544.

Cameron, S. V., and J. J. Heckman (1998): "Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of American males," *Journal of Political Economy*, 106, 262–333.

Cavanagh, C., and R. P. Sherman (1998): "Rank estimators for monotonic index models," *Journal of Econometrics*, 84, 351–381.

Chen, S. (2002): "Rank estimation of transformation models," *Econometrica*, 70, 1683–1697.

Coppejans, M. (2007): "On efficient estimation of the ordered response model," *Journal of Econometrics*, 137, 577–614.

Geskus, R., and P. Groeneboom (1996): "Asymptotically optimal estimation of smooth functionals for interval censoring, I," *Statistica Neerlandica*, 50, 69–88.

——— (1997): "Asymptotically optimal estimation of smooth functionals for interval censoring, II," *Statistica Neerlandica*, 51, 201–219.

——— (1999): "Asymptotically optimal estimation of smooth functionals for interval censoring, case 2," *The Annals of Statistics*, 27, 627–674.

Groeneboom, P. (2014): "Maximum smoothed likelihood estimators for the interval censoring model," *The Annals of Statistics*, 42, 2092–2137.

Groeneboom, P., and K. Hendrickx (2017): "The nonparametric bootstrap for the current status model," *Electronic Journal of Statistics*, 11, 3446–3483.

——— (2018): "Current status linear regression," *The Annals of Statistics*, 46, 1415–1444.

——— (2019): "Estimation in monotone single-index models," *Statistica Neerlandica*, 73(1), 78–99.

Groeneboom, P., and G. Jongbloed (2014): *Nonparametric estimation under shape constraints*. Cambridge University Press.

Groeneboom, P., G. Jongbloed, and B. I. Witte (2010): "Maximum smoothed likelihood estimation and smoothed maximum likelihood estimation in the current status model," *The Annals of Statistics*, 38, 352–387.

Groeneboom, P., and J. A. Wellner (1992): *Information bounds and nonparametric maximum likelihood estimation.* Birkhauser.

Han, A. K. (1987): "Non-parametric analysis of a generalized regression model: the maximum rank correlation estimator," *Journal of Econometrics*, 35, 303–316.

Honoré, B. E., and Á. de Paula (2010): "Interdependent durations," *Review of Economic Studies*, 77, 1138–1163.

Horowitz, J. L. (1996): "Semiparametric estimation of a regression model with an unknown transformation of the dependent variable," *Econometrica*, 64, 103–137.

Klein, R. W., and R. P. Sherman (2002): "Shift restrictions and semiparametric estimation in ordered response models," *Econometrica*, 70(2), 663–691.

Klein, R. W., and R. H. Spady (1993): "An efficient semiparametric estimator for binary response models," *Econometrica*, 61(2), 387–421.

Lewbel, A. (2002): "Ordered response threshold estimation," *Working paper.*

Rubin, D. (1981): "Bayesian bootstrap," *The Annals of statistics*, 9, 130–134.

Schick, A., and Q. Yu (2000): "Consistency of the GMLE with mixed case intervalcensored data," *Scandinavian Journal of Statistics*, 27, 45–55.

Sherman, R. P. (1993): "The limiting distribution of the maximum rank correlation estimator," *Econometrica*, 61, 123–137.

Van de Geer, S. (1993): "Hellinger-consistency of certain nonparametric maximum likelihood estimators," *The Annals of Statistics*, 21, 14–44.

——— (1995): "Asymptotic normality in mixture models," *ESAIM: Probability and Statistics*, 1, 17–33.

——— (2000): *Empirical processes in M-estimation.* Cambridge University Press.

Van der Vaart, A. (1991): "On differentiable functionals," *The Annals of Statistics*, 19, 178–204.

Van Der Vaart, A. (1998): *Asymptotic statistics.* Cambridge University Press.

Van Der Vaart, A., and J. A. Wellner (1996): *Weak convergence and empirical processes.* Springer.

Varadhan, R., and P. Gilbert (2009): “BB: An R package for solving a large system of nonlinear equations and for optimizing a high-dimensional nonlinear objective function,” *Journal of Statistical Software*, 32(4), 1–26.

Wu, C. (1990): “On the asymptotic properties of the jackknife histogram,” *The Annals of Statistics*, 18, 1438–1452.