Simple Semiparametric Estimation of Ordered Response Models: With an Application to the Interdependent Durations Model

Ruixuan Liu and Zhengfei Yu
Emory University and University of Tsukuba

Abstract. We propose two simple semiparametric estimation methods for ordered response models with an unknown error distribution. The proposed methods do not require users to choose any tuning parameter and they automatically incorporate the monotonicity restriction of the unknown distribution function. Fixing finite dimensional parameters in the model, we construct nonparametric maximum likelihood estimates (NPMLE) for the error distribution based on the related binary choice data or the entire ordered response data. We then obtain estimates for finite dimensional parameters based on moment conditions given the estimated distribution function. Our semiparametric approaches deliver root-n consistent and asymptotically normal estimators of the regression coefficient and threshold parameter. We also develop valid bootstrap procedures for inference. We apply our methods to the interdependent durations model in Honoré and de Paula (2010), where the social interaction effect is directly related to the threshold parameter in the corresponding ordered response model. The advantages of our methods are borne out in simulation studies and a real data application to the joint retirement decision of married couples.

Key words: Ordered Response Models, Semiparametric Estimation, Shape Restriction, NPMLE, Social Interaction, Interdependent Durations

JEL Classification: C14, C25, C35, C41

The first draft: May 3, 2018. This version: December 27, 2018.
Corresponding Address: Zhengfei Yu, Faculty of Humanities and Social Sciences, University of Tsukuba, 1-1-1, Tennodai, Tsukuba, Ibaraki, Japan, 305-8571. E-mail: yu.zhengfei.gn@u.tsukuba.ac.jp.
We would like to thank Jason Abrevaya, Bo Honore, Hidehiko Ichimura, Hiroyuki Kasahara, Toru Kitagawa, Brendan Kline, Zhongjian Lin, Oliver Linton, Essie Maasoumi, Taisuke Otsu, Aureo de Paula, Adam Rosen, Katsumi Shimotsu, Haiqing Xu, an Associate Editor and two anonymous referees for their insightful comments. Yu gratefully acknowledges the support of JSPS KAKENHI Grant Number 17K13713.

1. Introduction

We consider the following ordered response model in which the discrete dependent variable Y_i is defined by the threshold-crossing rule given covariates X_i, a latent error term ε_i, an unknown threshold parameter α_0, and a regression coefficient β_0:

$$Y_i = \begin{cases} 1 & \text{if} \quad \varepsilon_i \leq X_i'\beta_0, \\ 2 & \text{if} \quad X_i'\beta_0 < \varepsilon_i \leq X_i'\beta_0 + \alpha_0, \\ 3 & \text{if} \quad \varepsilon_i > X_i'\beta_0 + \alpha_0, \end{cases} \tag{1.1}$$

for $i = 1, \cdots, n$. We maintain the independence assumption between X and ε and let $F_0(\cdot)$ be the true unknown distribution function of the latent error ε throughout the paper. Given independent and identically distributed (*i.i.d.*) observations of $(Y_i, X_i)_{i=1}^n$, the likelihood function takes the following form:

$$\mathbb{L}_n(\alpha, \beta, F) = \Pi_{i=1}^n \left\{ F(X_i'\beta)^{\Delta_{1i}} \left[F(X_i'\beta + \alpha) - F(X_i'\beta)\right]^{\Delta_{2i}} \left[1 - F(X_i'\beta + \alpha)\right]^{\Delta_{3i}} \right\}, \tag{1.2}$$

where the indicators Δ_{1i}, Δ_{2i}, and Δ_{3i} are defined by $\Delta_{1i} = \mathbb{I}\{Y_i = 1\}$, $\Delta_{2i} = \mathbb{I}\{Y_i = 2\}$, and $\Delta_{3i} = 1 - \Delta_{1i} - \Delta_{2i}$ for $i = 1, \cdots, n$.

The ordered response model dates back to Aitchison and Silvey (1957) where the error distribution F is parameterized and is widely used to characterize ordered categorical outcome in economics, such as consumers' demand of differentiated products (Prescott and Visscher, 1977; Shaked and Sutton, 1982), symmetric entry games (Bresnahan and Reiss, 1991), schooling choices (Cameron and Heckman, 1998; Cunha, Heckman, and Navarro, 2007), credit/liquidity constraints (Attanasio, Koujianou, and Kyriazidou, 2008; Hajivassiliou and Ioannides, 2007) and discrete time (or interval-censored) duration models (Ridder, 1990; Manski and Tamer, 2002). We refer readers to Greene and Hensher (2010) for a comprehensive review. However, the fully parametric procedure leads to an inconsistent estimate and misleading inference if the parametric model of the error distribution is misspecified. Flexible semiparametric estimation of (1.1) has been studied by Lee (1992), Melenberg and Van Soest (1996), Klein and Sherman (2002), Lewbel (2002), and Coppejans (2007), allowing for an arbitrary error distribution.[1] This literature can be roughly divided into two categories. The first branch employs either kernel or sieve based nonparametric estimation of the functional nuisance component as in Klein and Sherman (2002), Lewbel (2002), and Coppejans (2007). The implication is that one has to choose a tuning parameter, such as the bandwidth in kernel smoothing or the number of sieve basis functions, and

[1]We would like to mention Lewbel (1997, 2000) and Chen and Khan (2003) with a different emphasis on conditional heteroskedasticity, which does not require independence between the covariates and latent error in ordered response models.

there is no clear answer about the optimal choice in this context. Inevitably, this requires a considerable amount of intervention and judgment on the part of practitioners. The second approach, which estimates the finite dimensional parameter (α_0, β_0) (but not F_0) without tuning parameters, makes use of the maximum score type estimation (Lee (1992)). [2] Only the consistency result is available for Lee's (1992) estimator and it is expected to have a non-standard limiting distribution with cubic-root convergence rate which also complicates the inference in practice, not to mention that the maximum score type estimator is very hard to compute. Part of the aforementioned issues can be alleviated by smoothing the sample criterion function, as done by Melenberg and Van Soest (1996), which again introduces the kernel bandwidth. Also, the convergence rate of the smoothed maximum score estimator is known to be slower than the standard root-n rate (Horowitz, 1992, 2009).

In this paper, we propose two simple semiparametric estimation methods for ordered response models that are fully automatic and free of any tuning parameter. The resulting estimators of the slope coefficient β_0 and the threshold α_0 are root-n consistent and asymptotically normal. The first approach, which consists of two stages, starts with the likelihood function for related binary choice data (Δ_{1i}, X_i) for $i = 1, \cdots, n$:

$$\mathbb{L}_{1n}(\beta, F) = \Pi_{i=1}^{n} \left\{ F(X_i'\beta)^{\Delta_{1i}} \left[1 - F(X_i'\beta)\right]^{1-\Delta_{1i}} \right\}, \tag{1.3}$$

to get our estimated distribution function $\hat{F}_n(\cdot;\beta)$ for any given β following Cosslett (1983). This is the nonparametric maximum likelihood estimator (NPMLE) in the sense of Kiefer and Wolfowitz (1956) for the *binary choice* data. In order to differentiate from the NPMLE in our second approach, we will refer $\hat{F}_n$ as the *isotonic estimator* throughout the paper. We then estimate the regression coefficient and threshold parameter sequentially by using certain moment conditions (or solving estimation equations). In fact, the first stage is adapted from the tuning-parameter-free method in Groeneboom and Hendrickx (2018) for binary choice models, whereas in the second stage, we obtain the estimated threshold from a simple moment condition concerning the binary choice data (Δ_{3i}, X_i) for $i = 1, \cdots, n$. Our second approach directly maximizes the full likelihood in (1.2) for the *ordered response* data to obtain the NPMLE $\tilde{F}_n(\cdot;\alpha,\beta)$ for any given (α, β). Thereafter, we estimate the regression coefficient and threshold *jointly* by using the same moment conditions as in our two-stage approach. Throughout this paper, we name the first approach the (Isototic) two-stage estimation and the second one the (NPMLE-based) joint estimation.

Our estimation approaches have three main appealing features. First, both methods are free from any tuning parameter, as opposed to Klein and Sherman (2002), Lewbel

[2]Another possibility is a two-stage rank estimator that combines Cavanagh and Sherman (1998) and Chen (2002). Because the original focus of these papers are not on the ordered response model, we discuss the approach in Section 3.3. We are grateful to the knowledgeable Associate Editor who suggested this method.

(2002), Coppejans (2007), and Melenberg and Van Soest (1996). This is achieved because we estimate the latent error's distribution F in (1.3) by a well-defined NPMLE (either using the binary choice data or the ordered response data), which only exploits the shape (monotonicity) restriction. As a result, the estimator $\hat{F}_n$ or $\tilde{F}_n$ does not rely on any kernel smoothing or sieve penalization. Second, both estimators of the latent distribution F obtained from our methods are automatically non-decreasing piece-wise constant functions by construction. In contrast, the kernel-based approach in Klein and Sherman (2002) does not guarantee a monotonic estimate of the latent distribution[3] and the sieve estimator in Coppejans (2007) has to incur additional computation costs by restricting spline coefficients to accommodate monotonicity. Finally, our approach is easy to implement. The isotonic estimator $\hat{F}_n(\cdot)$ is computed using the celebrated pool-adjacent-violators algorithm (PAVA) (see Groeneboom and Jongbloed (2014) for details), whereas for the NPMLE $\tilde{F}_n(\cdot)$, we adapt the fast hybrid approach in Wellner and Zhan (1997) that combines both the E-M algorithm and the iterative greatest convex minorant algorithm. Our two-stage estimation is particularly attractive from the computational point of view, in the sense that for the given $(\hat{\beta}_n, \hat{F}_n(\cdot; \hat{\beta}_n))$, the estimating equation for the threshold parameter α is monotonic. In comparison, the second stage of Klein and Sherman (2002) involves repeating a grid search many times in order to obtain the estimator of the threshold parameter.[4]

Our interest in (1.1) stems from an important application of the interdependent durations model proposed by Honoré and de Paula (2010). The study of social interaction effects has received considerable attention in economics literature and we refer readers to Durlauf and Ioannides (2010) for fast expanding results based primarily on discrete choice models. The seminal papers of de Paula (2009), Honoré and de Paula (2010) bring strategic interactions explicitly into the optimal timing decisions of economic agents and greatly extend the scope of prior static discrete choices models by incorporating the time dimension. It is evident in de Paula (2009) and Honoré and de Paula (2010) that the social interaction effect captures notable externality and explains interesting timing coordination behaviors. Furthermore, the scalar representing the social interaction effect in Honoré and de Paula (2010) is directly related to the threshold parameter in the resulting ordered response model, which motivates us to propose the current flexible semiparametric estimation methods. Both the Monte Carlo simulation and a real data application demonstrate a good performance of our proposal. We believe that our work makes the Honoré-de Paula model a viable benchmark for analyzing interdependent durations data. Other applicable examples in economics include Park and Smith (2008), who study late rushes in market entry after a pioneer develops

[3]Figure 3 in Section 5.2 plots the estimated CDF of the latent error in a real data example.

[4]We refer readers to Stewart (2005) about detailed comparisons of existing semi-nonparametric estimators from their computational aspect.

a new product; Katz and Shapiro (1986), who examine technology adoption; Hong, Kubik, and Stein (2004), who model stock market participation; de Paula (2009), who observes group desertion in the Union Army after the American Civil War; and Honoré and de Paula (2018), who look at the joint retirement decisions of married couples.

We contribute to the literature in several ways. First, the proposed estimation methods complement the existing semiparametric approaches in the sense they free applied researchers from choosing any tuning parameter. In contrast, to implement Klein and Sherman (2002), Lewbel (2002), Coppejans (2007) or Melenberg and Van Soest (1996), one has to specify bandwidths, orders of polynomials, or trimming sequences. Unlike Lee (1992) where only the consistency result is available for the maximum score type estimator, we rigorously establish the root n consistency and asymptotic normality for our estimators of the finite dimensional parameters. Second, from a theoretical perspective, our work contributes to the literature of (two-stage) semiparametric estimation and inference that involves shape-restricted components. In the seminal works of Newey (1994) and Chen, Linton, and Van Keilegom (2003), general theorems are presented for semiparametric estimators involving some first-stage nonparametric estimation, in which the linear representation for the directional derivative of the nonparametric estimate is maintained as a high-level assumption and verified for sieve or kernel-type estimators under sufficient smoothness restrictions. A distinction of our paper is that our nonparametric estimators are only piece-wise constant functions with random jump locations determined by the data. In fact, the crux of our theoretical investigation is to prove that certain linear functional (or the directional derivative) of the shape-restricted estimate is asymptotically normal, for which we adapt the recent breakthrough made by Groeneboom and Hendrickx (2018) for our two-stage estimation. A close examination reveals that the proof in Groeneboom and Hendrickx (2018) regarding the regression coefficient β is easier, as they can utilize an orthogonal (to the nuisance tangent set) score function to account for the estimation effect of the error distribution implicitly. This orthogonal direction is well-known for single-index models (Ichimura (1993); Klein and Spady (1993)). In contrast, we have to explicitly characterize the influence of estimating the distribution through its linear functional in our two-stage estimation. Third, our joint estimation provides a new example of utilizing the NPMLE in semiparametric models. It substantially goes beyond Groeneboom and Hendrickx (2018). In our model, the NPMLE making use of information in all three categories lacks an explicit characterization in contrast to the binary choice case considered by Groeneboom and Hendrickx (2018). As a result, determining the asymptotic behavior of its linear functional becomes much more challenging. Aiming at those challenges, our proofs that make novel use of the empirical process theory and the characterization of NPMLE for the "interval censoring, case 2 model"[5] (Geskus and Groeneboom (1996, 1997, 1999) and Van de Geer (1997)) are also of independent interest. Therefore, this paper also contributes to the literature of shape-restricted estimation and inference for semiparametric models (Cosslett, 1983; Groeneboom and Wellner, 1992; Matzkin, 1991, 1993; Banerjee, Mukherjee, and Mishra, 2009; Groeneboom, Jongbloed, and Witte, 2010; Groeneboom and Hendrickx, 2017, 2018). Fourth, we propose bootstrap procedures with general exchangeable weights to facilitate inference, as the asymptotic variances of our estimators are quite involved. The bootstrap consistency does not trivially follow from Van Der Vaart and Wellner (1996) or Cheng and Huang (2010), since the bootstrap is known to fail for the point-wise distribution of the isotonic estimator or NPMLE (Abrevaya and Huang (2005)). Heuristically speaking, the bootstrap is valid for the finite dimensional parameter in our model, because the influence of the isotonic estimator or the NPMLE only involves certain linear functional (Huang and Wellner (1995),Groeneboom and Hendrickx (2017)) whose asymptotic behavior can be mimicked by the bootstrap. Last but not least, we demonstrate the usefulness of the Honoré-de Paula model by taking it to the real data based on our proposed estimation and inference methods. Indeed, it shows the interaction effect between couples is large enough to counter their age difference in explaining the joint retirement behavior. Considering the continued interest in modeling social interactions among economists and the level of sophistication of the Honoré-de Paula model, we believe our empirical illustration is worthwhile.

The rest of our paper is organized as follows. We briefly discuss the connection to the interdependent durations model in Honoré and de Paula (2010) in Section 2. In Section 3, we introduce the notation and propose two simple semiparametric estimation methods. In Section 4, we derive the asymptotic properties of our estimators for the finite dimensional parameters, proving their consistency and asymptotic normality. Given the complicated influence functions, we also develop novel bootstrap procedures so that one can easily formulate the bootstrap confidence sets. Section 5 conducts simulation studies to evaluate the finite-sample properties of the estimators and also illustrates the proposed methods using a real dataset. The final section concludes. Proofs of main theorems are in Appendix A, whereas proofs of technical lemmas are delegated to the supplemental note.

[5]To clarify the comparisons with Groeneboom and Hendrickx (2018), the binary choice model there is also known as the "interval censoring, case 1 model"; see Groeneboom and Wellner (1992).

2. The Interdependent Durations Model

Interactions among economic agents play key roles in characterizing simultaneity of multiple durations models. This endogenous interaction effect generates interesting synchronization for numerous economic phenomena. Honoré and de Paula (2010) study a game-theoretic model in which two players respectively decide (T_1, T_2) as 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) with the endogenous interaction effect are characterized by

$$
\begin{aligned}
T_1 &\equiv \inf \left\{t_1 : \phi\left(X_1\right) \Lambda\left(t_1\right) \exp\left[\alpha^* \mathbb{I}\left\{T_2 \leq t_1\right\}\right] \geq \epsilon_1\right\}, \\
T_2 &\equiv \inf \left\{t_2 : \phi\left(X_2\right) \Lambda\left(t_2\right) \exp\left[\alpha^* \mathbb{I}\left\{T_1 \leq t_2\right\}\right] \geq \epsilon_2\right\},
\end{aligned} \tag{2.1}
$$

where the unknown scalar α^* captures the social interaction effect. The common $\Lambda(\cdot)$ is a deterministic trend function. Each player has his/her own covariate X_j for $j = 1, 2$ and (ϵ_1, ϵ_2) are random utility flows from the initial state. Throughout the paper, we focus on the case where an individual's switching is complementary to each other, so the interaction effect α^* is positive.

In our empirical application, we utilize the model (2.1) to analyze the joint retirement decision of married couples so (T_1, T_2) represents the retirement timing choices of the couple. Understanding the retirement mechanism helps to guide the optimal design of employer-provided and government benefit programs (Gustman and Steinmeier (2000), Gustman and Steinmeier (2004), An, Christensen, and Gupta (2004), Honoré and de Paula (2018)). Our focus on the married couple stems from the fact that most people approaching retirement age are married and a significant portion of them choose to retire at the same time. Empirical studies documenting the joint retirement of couples abound based on different datasets (Honoré and de Paula (2018)). Standard bivariate mixed proportional hazards models do not apply to this scenario, because the simultaneous failure (when $\Pr\{T_1 = T_2\} > 0$) is ruled out from the very beginning in those models. In the current context, the complementary interaction effect is associated with the pension accrual profiles or the complementarity in leisure time; see Gustman and Steinmeier (2000) for a more detailed discussion.

Despite the sophisticated game structure and presence of multiple equilibria in (2.1), Honoré and de Paula (2010) prove the identifiability of all model primitives. Moreover, they point out a close connection with the corresponding ordered responses model, which directly suggests a natural semiparametric estimation route for the interaction effect. We start to define the auxiliary discrete dependent variable:

$$Y = \begin{cases} 1 & \text{if} \quad T_1 < T_2, \\ 2 & \text{if} \quad T_1 = T_2, \\ 3 & \text{otherwise.} \end{cases}$$

Then we parameterize the covariate effect by setting $\phi(x_j) = x_j'\beta_0$ for $j = 1, 2$, as in Honoré and de Paula (2010). Therefore, one gets

$$\begin{aligned} P\{Y = 1|X_1, X_2\} &= H\left((X_1 - X_2)'\beta_0 - \alpha^*\right), \\ P\{Y = 2|X_1, X_2\} &= H\left((X_1 - X_2)'\beta_0 + \alpha^*\right) - H\left((X_1 - X_2)'\beta_0 - \alpha^*\right), \\ P\{Y = 3|X_1, X_2\} &= 1 - H\left((X_1 - X_2)'\beta_0 + \alpha^*\right), \end{aligned} \tag{2.2}$$

where $H(w) = P\{\log \epsilon_1 - \log \epsilon_2 \le w\}$. Hence, it is straightforward to observe that (2.2) corresponds to the semiparametric ordered response model by taking $F_0(\cdot) = H(\cdot - \alpha^*)$, $X = (X_1 - X_2)$, and $\alpha_0 = 2 \times \alpha^*$.

Remark 2.1. *Note that the ordered response model does not capture all model primitives in the original interdependent durations model, in particular the common deterministic trend function $\Lambda(\cdot)$ is missing. However, the constructive identification in Theorem 3 of Honoré and de Paula (2010) suggests a straightforward nonparametric estimate for $\ln \Lambda(t)$, combined with the key insight of Horowitz (1996)[6]. To elaborate on the proposal, let*

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

then the proof of Theorem 3 in Honoré and de Paula (2010) states[7] that

$$\frac{\partial \ln \Lambda(t_1)}{\partial t_1} = \beta_{0k} \frac{\partial h / \partial t_1}{\partial h / \partial x_{1k}}. \tag{2.3}$$

Thus, one can adopt the estimator defined by equation (2.4) in Horowitz (1996) by plugging in some nonparametric kernel estimator for the partial derivative of $h(t_1, t_2; x_1, x_2)$. Given that the theoretical properties follow largely from Horowitz (1996) and $\Lambda(t)$ does not directly enter the resulting ordered response model, we will not focus on the issue.

3. Simple Semiparametric Estimation

Throughout the paper, we work with the i.i.d. data (Y_i, X_i) for $i = 1, ..., n$. It is convenient to introduce the indicators Δ_{1i}, Δ_{2i}, and Δ_{3i} defined by $\Delta_{1i} = \mathbb{I}\{Y_i = 1\}$ and

[6]We want to thank one anonymous referee who pointed out the issue.

[7]Note that in our setup, the covariate effects are parameterized by $\phi(x) = \exp(x'\beta_0)$. Hence, the ratio $\frac{\partial \phi / \partial x_{1k}}{\phi(x_1)}$ defined in Theorem 3 of Honoré and de Paula (2010) is always equal to β_{0k} in our specification.

$\Delta_{2i} = \mathbb{I}\{Y_i = 2\}$, and $\Delta_{3i} = 1 - \Delta_{1i} - \Delta_{2i}$ for $i = 1, \cdots, n$. Let K denote the dimensionality of covariates X and write $\beta_0 \equiv (\beta_{01}, \beta_{02}, ..., \beta_{0K})'$. Note that the regression coefficient β_0 is only identified up to some scale normalization for an unspecified F_0 (see Klein and Sherman (2002)). Without loss of generalization, we normalize $\beta_{01} = 1$.

We use the standard empirical process notations as follows. For a function $f(\cdot)$ of a random vector $Z = (Y, X)$ that follows distribution P, we let $Pf = \int f(z)dP(z), \mathbb{P}_n f = n^{-1}\sum_{i=1}^n f(Z_i)$, and $\mathbb{G}_n f = n^{1/2}(\mathbb{P}_n - P) f$. Function f can be replaced by a random function $z \mapsto \hat{f}_n(z; Z_1, \cdots, Z_n)$. Therefore, $P\hat{f}_n = \int f(z; Z_1, \cdots, Z_n)dP(z)$, $\mathbb{P}_n \hat{f}_n = n^{-1}\sum_{i=1}^n f(Z_i; Z_1, \cdots, Z_n)$ and $\mathbb{G}_n \hat{f}_n = n^{1/2}(\mathbb{P}_n - P)\hat{f}_n$; see Nan and Wellner (2013). Let $\eta = (\theta, F(\cdot; \theta))$ be the unknown parameter containing both finite dimensional parameter $\theta \equiv (\alpha, \beta')'$ and the distribution function F. Furthermore, we consider $\alpha \in \mathcal{A}$, $\beta \in \mathcal{B}$, and $F \in \mathcal{F}$, where $\mathcal{A} \subset \mathbb{R}_+$, $\mathcal{B} \subset \mathbb{R}^K$, and $\mathcal{F}$ is the class of distribution functions. The distance between two parameter values (η_1, η_2) is defined in terms of the following metric

$$d(\eta_1, \eta_2) = |\theta_1 - \theta_2| + \|F_1(\cdot; \theta_1) - F_2(\cdot; \theta_2)\|$$

where $|\cdot|$ is the standard Euclidean distance, and $\|\cdot\|$ is some norm for the class of distribution functions. We work with the $L_\infty-$norm related to our two-stage estimation and the L_2- norm regarding our joint estimation for technical convenience.

3.1. Two-stage Semiparametric Estimation

Our two-stage estimation procedure is inspired by Klein and Sherman (2002) and Lewbel (2002) in the sense that we obtain estimated distribution function from the related binary choice data (Δ_{1i}, X_i) and (Δ_{3i}, X_i). Unlike Klein and Sherman (2002) who resort to the kernel estimator in Klein and Spady (1993) or Lewbel (2002) who requires a preliminary nonparametric estimation of certain conditional mean function and integration over covariates' values, we adapt the important breakthrough by Groeneboom and Hendrickx (2018) to analyze ordered response models. As it will become self-evident in the sequel, our procedure delivers root-n consistent and asymptotically normal estimators of the regression coefficient β_0 and threshold parameter α_0, which does not require tuning parameters.

We now describe the two-stage semiparametric estimation for the ordered response model: **Stage 1(i).** For any β, we compute the NPMLE for $F(\cdot)$ based on the *binary choice* data:

$$\hat{F}_n(\cdot; \beta) = \arg\max_F \sum_{i=1}^n \left[\Delta_{1i} \log F(X_i'\beta) + (1 - \Delta_{1i}) \log(1 - F(X_i'\beta))\right]. \tag{3.1}$$

Stage 1(ii). Given $\hat{F}_n(\cdot;\beta)$, our estimator $\hat{\beta}_n$ for the regression coefficient is the zero-crossing point of the estimation equation:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \left[\Delta_{1i} - \hat{F}_n(X_i'\hat{\beta}_n;\hat{\beta}_n)\right] = 0. \tag{3.2}$$

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

$$\Psi_n\left(\alpha, \hat{\beta}_n, \hat{F}_n(\cdot;\hat{\beta}_n)\right) = \frac{1}{n}\sum_{i=1}^{n}\left[1 - \Delta_{3i} - \hat{F}_n(X_i'\hat{\beta}_n + \alpha;\hat{\beta}_n)\right]. \tag{3.3}$$

Here we provide a heuristic discussion of each step. The NPMLE in Stage 1 and its characterization date back to Ayer, Brunk, Ewing, Reid, and Silverman (1955) in analyzing the current status data (see Groeneboom and Wellner (1992)). The corresponding optimization problem is well-defined and it generates a piece-wise constant function $\hat{F}_n(\cdot;\beta)$, which can be characterized as follows: fixing the parameter β, we consider the values of $U_1^{(\beta)} = X_1'\beta, \cdots, U_n^{(\beta)} = X_n'\beta$. Let $U_{(1)}^{(\beta)} \leq \cdots \leq U_{(n)}^{(\beta)}$ be the order statistics with corresponding indicators $\Delta_{1,(i)}^{(\beta)}$ for $i = 1, \cdots, n$. Thereafter, $\hat{F}_n(\cdot;\beta)$ is equal to the left derivative of the convex minorant of a cumulative sum diagram consisting of the points $(0,0)$ and

$$\left(i, \sum_{j=1}^{i} \Delta_{1,(j)}^{(\beta)}\right)$$

for $i = 1, \cdots, n$; see Groeneboom and Hendrickx (2018). Within the context of binary choice models, it is utilized by Cosslett (1983) to define the tuning-parameter-free profile likelihood estimator. However, only consistency results [8] are available for Cosslett's estimator given the challenge that the estimated error distribution is neither linear nor smooth. The key to develop a root-n consistent and asymptotic normal estimator for β_0 while maintaining the tuning-parameter-free feature is the Z-estimator adapted from Groeneboom and Hendrickx (2018). Modulo the estimated latent distribution function, one makes use of the population-level moment condition

$$\mathbb{E}[X(\Delta_1 - F_0(X'\beta_0))] = 0, \tag{3.4}$$

[8] The general result in Tanaka (2008) implies the convergence rate of Cosslett's estimator is cubic root for the slope coefficient.

and plug in the NPMLE $\hat{F}_n(\cdot;\beta)$ in the sample analog.[9] In the same spirit, the last step of our procedure is based on a very simple moment condition:

$$\mathbb{E}[(1-\Delta_3 - F_0(X'\beta_0+\alpha_0))] = 0. \tag{3.5}$$

Given preliminary estimators $\hat{F}_n(\cdot;\hat{\beta}_n)$ and $\hat{\beta}_n$, the estimating equation Ψ_n is monotone with respect to α, which greatly lowers the computational burden. We emphasize that it is necessary to use both sets of moment conditions for the sake of consistency. The naive approach where one uses only the binary choice data (Δ_{2i}, X_i) and then directly applies Groeneboom and Hendrickx (2018) does not work because the intercept α_0 and the distribution function F_0 cannot be separately identified in the binary choice data alone (Ichimura (1993)). We focus on the just-identified case to be coherent with Groeneboom and Hendrickx (2018). In principle, a GMM type estimation with over-identified moment conditions could be developed[10].

As $\hat{F}_n(\cdot;\hat{\beta}_n)$ is a piece-wise constant function, the estimating equations in our procedure may not hold exactly. Therefore, we adopt the following definition from Groeneboom and Hendrickx (2018):

Definition 3.1 (Zero-crossing)**.** *We say that β_* is a zero-crossing of a function $C:\mathbf{B}\mapsto\mathbb{R}$ if each open neighborhood of β_* contains points $\beta_1,\beta_2\in\mathbf{B}$ such that $C(\beta_1)C(\beta_2)\le 0$. We say that a function $\tilde{C}:\mathbf{B}\mapsto\mathbb{R}^K$ crosses zero at point β_* if β_* is a zero-crossing in each component $\tilde{C}_j$ for $j=1,\cdots,K$.*

Figure 1 depicts the typical shape of our estimating function $\Psi_n\left(\alpha,\hat{\beta}_n,\hat{F}_n(\cdot;\hat{\beta}_n)\right)$. The data generating process is the interdependent duration model described in Section 5.1. The true value of α is 1. The joint distributions of (ϵ_1,ϵ_2) are exponential (Panel (a)) and log-normal (Panel (b)), respectively. It is obvious that these two depicted estimating functions $\Psi_n\left(\alpha,\hat{\beta}_n,\hat{F}_n(\cdot;\hat{\beta}_n)\right)$ are decreasing in α and the zero crossings are very close to the 1; i.e., they are about 0.980 in Panel (a) and 0.868 in Panel (b).

3.2. Joint Semiparametric Estimation

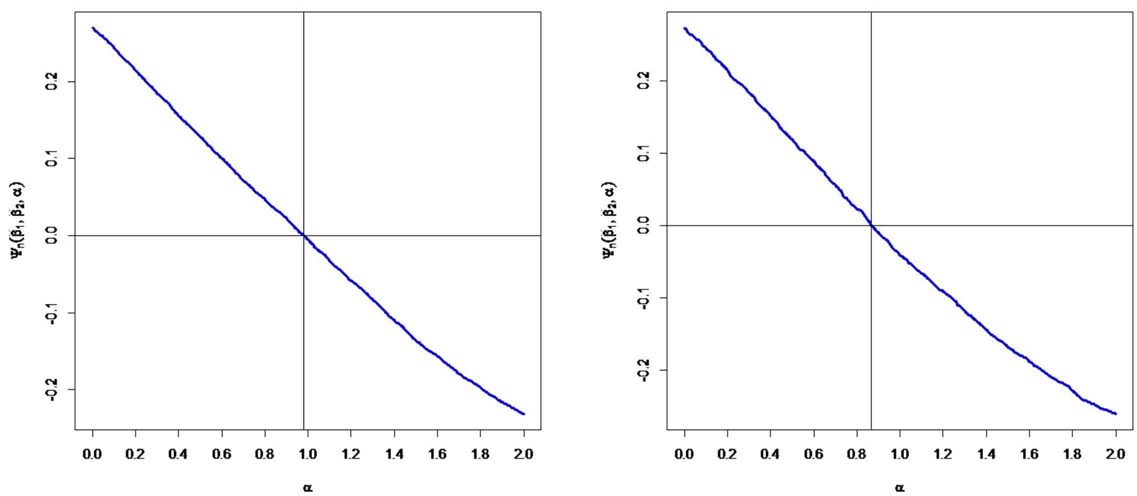
One might be wondering whether it is possible to develop a similar tuning-parameter-free estimation approach utilizing the entire ordered response data directly, instead of breaking it to two sets of binary choice data. Indeed this is feasible and will be termed as the joint

[9]The main improvement made by Groeneboom and Hendrickx (2018) over Cosslett (1983) to restore standard distributional theory for regression coefficient is that one does not need the error's density function in the moment condition (3.4). In contrast, one has to handle the error density in the likelihood based estimation appearing in the score function, whereas the NPMLE $\hat{F}_n(\cdot;\beta)$ itself is not differentiable.

[10]We want to thank an anonymous referee for helpful discussions related to the point.

Figure 1. The estimating function $\Psi_n\left(\alpha, \hat{\beta}_n, \hat{F}_n(\cdot; \hat{\beta}_n)\right)$ when the joint distribution of (ϵ_1, ϵ_2) in the interdependent duration model in Section 2 is exponential or log-normal. $\alpha_0 = 1$, sample size $= 500$.

(a). Exponential (b). Log-normal



estimation method. The starting point is the following result that guarantees the existence of the NPMLE for the ordered response model given any (α, β) in the parameter space. The proof can be found in Chapter 4 of Groeneboom and Jongbloed (2014).

Lemma 3.1. *Given α and β, the NPMLE for $\tilde{F}_n(\cdot; \alpha, \beta)$ exists based on the ordered response data; i.e.,*

$$\begin{aligned}&\tilde{F}_n(\cdot; \alpha, \beta)\\ =&\arg\max_F \sum_{i=1}^{n} \left[\Delta_{1i} \log F(X_i'\beta) + \Delta_{2i} \log\left(F(X_i'\beta + \alpha) - F(X_i'\beta)\right) + \Delta_{3i} \log\left(1 - F(X_i'\beta + \alpha)\right)\right]\end{aligned}$$

is well defined. Moreover, the NPMLE is a (sub-)distribution function and piece-wise constant with jumps over a subset of $\{X_i'\beta, X_i'\beta + \alpha : i = 1, 2, ..., n\}$.

Now we are ready to present the joint semiparametric method that makes use of information in all three categories to estimate the distribution function and returns the estimates for the regression coefficient and the threshold parameter simultaneously.

Joint Estimation. For any α and β, we employ the NPMLE for $\tilde{F}_n(\cdot; \alpha, \beta)$ based on the *ordered response* data. Given $\tilde{F}_n(\cdot; \alpha, \beta)$ from the previous step, we define $(\tilde{\alpha}_n, \tilde{\beta}_n)$ which are the zero-crossing points of the estimation equations simultaneously:

$$\Phi_n(\tilde{\alpha}_n, \tilde{\beta}_n) = 0, \tag{3.6}$$

where

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

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). The number of mass points 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:

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

And then set k be 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

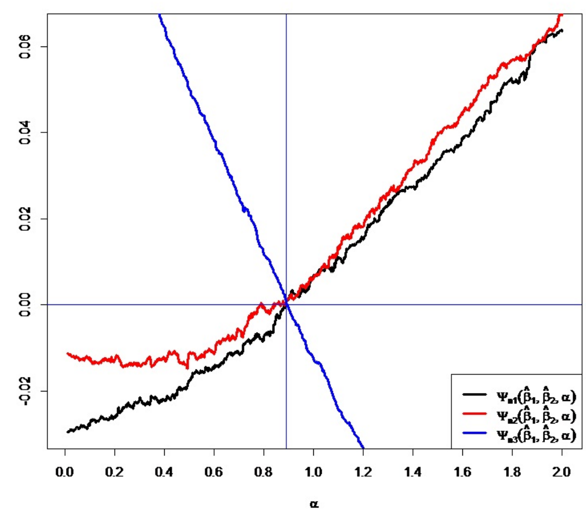
$$H_j(v) = \begin{cases} v_j^{-2} \quad \text{if} \quad j\in I_1, \\ (v_{k(j)}-v_j)^{-2} \quad \text{if} \quad j\in I_{2l}, \\ (v_j - v_{k^{-1}(j)})^{-2} \quad \text{if} \quad j\in I_{2r}, \\ (1-v_j)^{-2} \quad \text{if} \quad j \in I_3. \end{cases}$$

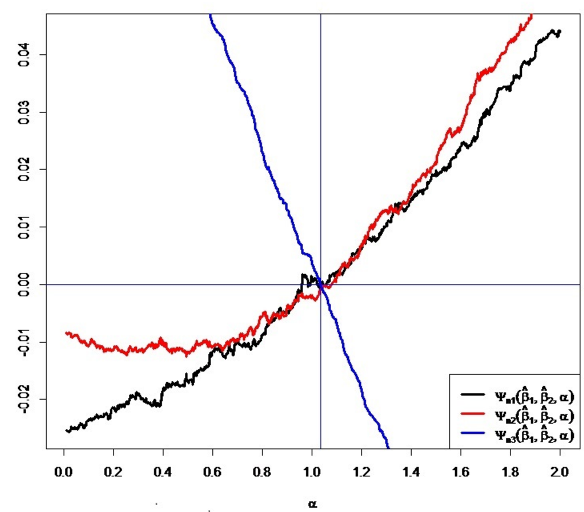
The initial value can be set as $v^{(0)} = (1/p, 2/p, ..., 1)'$, which assigns the same probability mass on each jump point. The iterative convex minorant algorithm can be implemented using the R package Icens (Gentleman and Vandal, 2018). R package Icens also provides a function for a faster hybrid algorithm proposed by Wellner and Zhan (1997), which combines the iterative convex minorant and the E-M algorithm.

Figure 2 plots the estimating functions with respect to α for fixed $\beta = \tilde{\beta}_n$. The designs follow the ones presented in Section 5.1, with $\beta_0 = (1, \beta_{02}, \beta_{03})$. Therefore, (3.6) contains three estimating equations. In Figure 2, the estimate $\tilde{\alpha}_n$ is the value where the three estimating functions cross zero, which equals to 0.890 and 1.036 for the exponential and log-normal design, respectively. Compared with the two-stage estimator, there are mainly two complications. First, all estimating equations are not guaranteed to be monotone with respect to α. Second, we need to solve for the joint zero-crossing point for all equations simultaneously. Nevertheless, as proved in our Theorem 4.3, the zero crossing points exist with probability 1.

FIGURE 2. The estimating functions $\Psi_n\left(\alpha, \hat{\beta}_n, \hat{F}_n(\cdot; \hat{\beta}_n)\right)$ when the joint distribution of (ϵ_1, ϵ_2) in the interdependent duration model in Section 2 is exponential or log-normal. $\alpha_0 = 1$, sample size =500.

(a). Exponential (b). Log-normal





3.3. Discussions and Comparisons with Alternative Methods

This section describes some alternative semiparametric estimators for the model (1.1), and also discusses the general ordered response models with more than three categories.

Remark 3.1 (The kernel-based estimator in Klein and Sherman (2002))**.** *In the first stage of Klein and Sherman (2002)'s approach, the regression coefficient β_0 is estimated by maximizing the following quasi-likelihood function with respect to β:*

$$\sum_{i=1}^{n} \hat{\tau}(X_i) \left\{ \mathbb{I}\{Y_i = 1\} \ln \hat{P}_1(X_i'\beta) + \mathbb{I}\{Y_i = 2\} \ln[\hat{P}_2(X_i'\beta) - \hat{P}_1(X_i'\beta)] + \mathbb{I}\{Y_i = 3\} \ln[1 - \hat{P}_2(X_i'\beta)] \right\},$$

where $\hat{P}_j(X_i'\beta)$ is the kernel estimator of the conditional probability $P_j(X_i'\beta) \equiv \Pr(Y_i \leq j|X_i'\beta)$ for fixed β in Klein and Spady (1993), $j = 1, 2$, and the trimming function $\hat{\tau}(x) = \mathbb{I}\left\{|x| \leq \hat{\xi}\right\}$ with $\hat{\xi}$ being certain sample quantile of $|X_i|$'s. In the second stage, the threshold parameter α_0 is estimated through the shift restriction $P_2(X_i'\beta_0 - \alpha_0) = P_1(X_i'\beta_0)$, which leads to

$$\hat{\alpha} = \frac{1}{\mathbb{I}\{i \in \mathcal{T}\}} \sum_{i \in \mathcal{T}} (\hat{V}_i - \tilde{V}_{i2}), \tag{3.7}$$

where $\hat{V}_i \equiv X_i'\hat{\beta}$ and $\tilde{V}_{i2}$ solves $\hat{P}_2(\tilde{V}_{i2}) = \hat{P}_1(\hat{V}_i)$ [11] *for each $i \in \mathcal{T}$, $\mathcal{T} = \left\{\hat{V}_i : \hat{P}_L \leq \hat{P}_1(\hat{V}_i) \leq \hat{P}_U\right\}$ and $(\hat{P}_L, \hat{P}_U)$ are determined by the pth and $(1-p)$th quantile of an a collection of estimated probabilities (see Klein and Sherman (2002), p671). Besides the choice of bandwidth for $\hat{P}_j(X_i'\beta)$, the K-S approach also requires two trimming parameters. Our simulations studies find that the performance of the threshold parameter estimator $\hat{\alpha}$ is particularly sensitive to the trimming parameter p used for constructing the target set $\mathcal{T}$. (See Table 2 in Section 5.1.) Note that the purpose of the trimming is to exclude individual estimators $\hat{V}_i - \tilde{V}_{i2}$ with poor performance.*

Remark 3.2 (The smoothed maximum-score estimator)**.** *Horowitz (1992) initially proposed the smoothed maximum-score (SMS) estimator for the binary choice model and Melenberg and Van Soest (1996) have extended it to the ordered response model. The estimator in Melenberg and Van Soest (1996) can also be viewed as smoothed version of Lee (1992)'s maximum-score estimator for the ordered response model. Under the median independence condition, the SMS approach estimates $(\alpha_0, \beta_0')'$ by maximizing the smoothed objective function:*

$$\max_{\alpha, \beta'} \sum_{i=1}^{n} (2\mathbb{I}\{Y_i \geq 2\} - 1) K\left(\frac{-X_i'\beta}{h}\right) + (2\mathbb{I}\{Y_i \geq 3\} - 1) K\left(\frac{-X_i'\beta - \alpha}{h}\right),$$

where $K(v)$ is an integral kernel function satisfying $\lim_{v \to +\infty} K(v) = 1$ and $\lim_{v \to -\infty} K(v) = 0$.[12] *Users need to specify the bandwidth h. The convergence rate of SMS is slower than*

[11] In its actual implementation, $\tilde{V}_{i2}$ is the point for which $\hat{P}_2(\tilde{V}_{i2})$ is closest to $\hat{P}_1(\hat{V}_i)$ over a grid constructed following the procedure given in [p.671-672] of Klein and Sherman (2002).

[12] Horowitz (1992) adopts $K(v) = 0.5 + (105/64)[v - (5/3)v^3 + (7/5)v^5 - (3/7)v^7]$ if $|v| \leq 1$; $K(v) = 0$ if $v < -1$ and $K(v) = 1$ if $v > 1$.

the root-n rate, even with the MSE-optimal bandwidth of order $n^{-1/5}$; see Section 4.3.3 of Horowitz (2009) for a detailed discussion. Simulation results in Section 5.1 find that SMS yields larger mean square errors and longer confidence intervals than other methods.

Remark 3.3 (A two-stage rank estimator)**.** *An alternative tuning-parameter-free method is the two-stage rank estimation combining Cavanagh and Sherman (1998) and Chen (2002)*[13]*. The original focus of Chen (2002) is the unknown link function in the transformation model; however, his method is also applicable to the ordered response model. Specifically, one can estimate β_0 following Cavanagh and Sherman (1998) first by*

$$\max_{\beta} \sum_{i=1}^{n} Y_i R_n(-X_i'\beta), \tag{3.8}$$

where $R_n(-X_i'\beta)$ denotes the rank of $-X_i'\beta$. The second stage adapts the key idea of Chen (2002) that given $P(Y_i = 1|X_i) = F_0(X_i'\beta_0)$ and $P(Y_i \leq 2|X_i) = F_0(X_i'\beta_0 + \alpha_0)$, one has

$$\mathbb{E}\left[\mathbb{I}\{Y_i = 1\} - \mathbb{I}\{Y_j \leq 2\}|X_i, X_j\right] \geq 0 \text{ whenever } X_i'\beta_0 - X_j'\beta_0 \geq \alpha_0 \text{ for } i \neq j.$$

Therefore, a maximum rank correlation estimator for the threshold is

$$\hat{\alpha} = \arg\max_{\alpha} \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \left(\mathbb{I}\{Y_i = 1\} - \mathbb{I}\{Y_j \leq 2\}\right) \mathbb{I}\{X_i'\beta - X_j'\beta \geq \alpha\}.$$

Our simulation studies demonstrate remarkable performances of this two-stage rank estimator for estimating the finite dimensional parameters, especially when the sample size is relatively large. However, unlike our two estimators or the Klein-Sherman estimator, the rank estimator itself does not provide a nonparametric estimate for the error distribution. In our empirical application, the error distribution function is still of practical interest, since it captures the (log-)difference of two heterogeneity terms for wives and husbands.

Remark 3.4. *In the density weighted estimator proposed by Lewbel (2000), one could resort to a novel ordered data estimator in (Lewbel and Schennach, 2007) without any tuning parameter, if an extra independence restriction between the special regressor and other covariates is imposed. (see Assumption A.5' [p.157] of Lewbel (2000)). However, in general Lewbel (2000)'s estimator needs a kernel or sieve type estimator of the conditional density in its first stage.*

Remark 3.5. *Finally, we discuss the general ordered response model where the dependent variable can take more than three values. Formally, the dependent variable is determined*

[13] We are grateful to the Associate Editor for this point. Given the connection is not obvious, we expand our discussion of this approach here.

by

$$
Y_i = \begin{cases} 1 \quad \text{if} \quad \varepsilon_i \le X_i'\beta_0, \\ 2 \quad \text{if} \quad X_i'\beta_0 < \varepsilon_i \le X_i'\beta_0 + \alpha_{0,1}, \\ 3 \quad \text{if} \quad X_i'\beta_0 + \alpha_{0,1} < \varepsilon_i \le X_i'\beta_0 + \alpha_{0,2}, \\ \vdots \\ J+1 \quad \text{if} \quad \varepsilon_i > X_i'\beta_0 + \alpha_{0,J}, \end{cases} \tag{3.9}
$$

for $i = 1, \cdots, n$ and $J \ge 3$, with a set of thresholds $(\alpha_{0,1}, \cdots, \alpha_{0,J})$. Our two-stage estimator can still be applied as follows. One can use the binary choice data $(\mathbb{I}(Y_i = 1), X_i)_{i=1}^n$ to implement the first stage, and then use the data $(\mathbb{I}(Y_i = j+1), X_i)_{i=1}^n$ to estimate $\alpha_{0,j}$ in the second stage, separately for $j = 1, \cdots, J$. Our joint estimator is also applicable. Note that 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 will be relevant for the computation of the NPMLE; the other intervals can be discarded (Groeneboom (2014), p2093). Therefore, the construction of the NPMLE is almost the same as the case with three categories. The consistency of the NPMLE for multiple categories can be found in Schick and Yu (2000). However, the rate of convergence or the asymptotic properties of its linear functionals remain unknown. Thereafter, we recommend practitioners use the methods in Klein and Sherman (2002)[14] or Coppejans (2007), if efficiency is the main concern. Since our empirical application in Section 5.2 involves only three categories, we will not proceed with the more general setup specified by (3.9).

4. Asymptotic Results

This section consists of three subsections. The first two subsections provide root-n consistency and asymptotic normality results for Isotonic two-stage estimator and NPMLE-based joint estimator, respectively. The last subsection provides bootstrap procedures for the construction of confidence intervals.

4.1. Asymptotic Properties of The Two-stage Estimator

A key challenge in our proof related to the two-stage estimation is to pin down the asymptotic contribution of $\hat{F}_n(\cdot, \hat{\beta}_n)$ to the finite dimensional parameter using the characterization of the isotonic estimator and empirical process theory. For the slope coefficients β_0, we rely on the recent breakthrough made by Groeneboom and Hendrickx (2018). Building on

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

that, we need substantial more efforts to determine the influence function for the threshold parameter α_0, as the unknown parameters in the estimating equation are substituted with the estimates in Groeneboom and Hendrickx (2018). The proof in Groeneboom and Hendrickx (2018) regarding the regression coefficient is easier because they can utilize an orthogonal (to the nuisance tangent set) score function to incorporate the estimation effect of the error distribution implicitly. This orthogonal direction is well-known for single-index models (Ichimura (1993) and Klein and Spady (1993)), as it involves the conditional mean of covariates X given the true linear index $U = X'\beta_0$. In contrast, we have to explicitly characterize the influence of estimating the distribution through its linear functional in our two-stage estimation.

We introduce additional notations to present our theoretical results. Following Groeneboom and Hendrickx (2018), it is straightforward to observe that the first-stage isotonic estimator $\hat{F}_n(\cdot;\beta)$ provides an estimate of

$$F_0(u;\beta) \equiv P\left\{\Delta_{1i}^{(\beta)} | U_i^{(\beta)} = u\right\} = \int F_0(u + x'(\beta_0 - \beta)) f_{X|X'\beta}(x|X'\beta = u)dx, \tag{4.1}$$

In the sequel, we let $F_0(u) = F_0(u;\beta_0)$. The density function of the random variable $X'\beta$ is denoted by $g_0(u;\beta)$. Denote the true linear index by $U_i = X_i'\beta_0$ for $i = 1, \cdots, n$, and let $g_0(u)$ be the probability density function of the random variable U. The following two terms appear in the Taylor expansion in our asymptotic analysis:

$$V_{\alpha_0} = \frac{\partial}{\partial \alpha} \mathbb{E}[F_0(X'\beta_0 + \alpha)]\big|_{\alpha=\alpha_0}, \tag{4.2}$$

$$V_{\beta_0} = \frac{\partial}{\partial \beta} \mathbb{E}[F_0(X'\beta + \alpha_0;\beta)]\big|_{\beta=\beta_0}, \tag{4.3}$$

whereas

$$H_{\beta_0} = \mathbb{E}\left[f_0(X'\beta_0)\left\{X - E[X|X'\beta_0]\right\}^{\otimes 2}\right] \tag{4.4}$$

denotes the Hessian matrix for $\hat{\beta}_n$ in Groeneboom and Hendrickx (2018).

The following regularity conditions are standard and adapted from Manski (1985), Ichimura (1993), Klein and Spady (1993), Klein and Sherman (2002), and Groeneboom and Hendrickx (2018). The only condition that we want to emphasize concerns the large support of the linear index $X'\beta_0$ in Condition 1. The assumption imposed here only requires the support to include the positive infinity, whereas the support can be bounded away from the negative infinity (so the linear index could follow exponential or gamma-type distribution). This condition is made to exclude the non-identifiable example in Manski (1985) and to ensure that the component ψ_{F_0} in the influence function of our Theorem 4.2 is well-defined. Considering the interdependent durations model, the large support condition is also assumed in Honoré and de Paula (2010) to achieve point identification.

Condition 1. We have i.i.d. data (Y_i, X_i) for $i = 1, \cdots, n$. The covariates X and latent error ε are independent. The support of X is not contained in any proper linear subspace of $\overline{\mathbb{R}}^K$. The support of $X'\beta_0$ contains $+\infty$.

Condition 2. The true regression parameter β_0 belongs to the interior of $\mathcal{B}$ where $\mathcal{B}$ is a compact set in $\mathbb{R}^K$. The true threshold parameter $\alpha_0 \in \mathcal{A} \equiv (\alpha_L, \alpha_U)$, where $[\alpha_L, \alpha_U]$ is a compact interval on the positive real line.

Condition 3. The function $F_0(\cdot; \beta)$ has a strictly positive continuous derivative, which stays away from zero for all $\beta \in \mathcal{B}$. Moreover, the function $F_0(u; \beta)$ is twice continuously differentiable on the interior of the support for $\beta \in \mathcal{B}$.

Condition 4. The density function of random variable $X'\beta$ denoted by $g_0(u; \beta)$ is continuous and also stays away from zero for all $\beta \in \mathcal{B}$.

Condition 5. The density $g_0(u; \beta)$ and conditional expectations $\mathbb{E}[X|X'\beta = u]$ and $E[XX'|X'\beta = u]$ are twice continuously differentiable w.r.t.u. The functions $\beta \mapsto g_0(u; \beta)$, $\beta \mapsto \mathbb{E}[X|X'\beta = u]$, and $\beta \mapsto \mathbb{E}[XX'|X'\beta = u]$ are continuous functions for u in the definition domain and all $\beta \in \mathcal{B}$.

Condition 6. The matrix H_{β_0} is of full rank and the scalar $V_{\alpha_0} \neq 0$, where $V_{\alpha_0} = \int f_0(u + \alpha_0) g_0(u) du$.

The asymptotic analysis of $\hat{\beta}_n$ and $\hat{F}_n(\cdot; \beta)$ follows directly from Theorem 4.1 on [p.1426] of Groeneboom and Hendrickx (2018). Specifically, the linear representation for $\hat{\beta}_n$ is as follows:

$$n^{1/2} H_{\beta_0} \left(\hat{\beta}_n - \beta_0 \right) = \int \{x - E[X|X'\beta_0 = x'\beta_0]\} \{F_0(x'\beta_0) - \delta_1\} d\mathbb{G}_n + o_p(1). \tag{4.5}$$

Thus, $\hat{\beta}_n$ is root-n consistent and asymptotically normal with the influence function equal to $H_{\beta_0}^{-1} \phi_{\beta_0}$, where

$$\phi_{\beta_0}(Z_i) = (X_i - \mathbb{E}[X_i|U_i])(F_0(U_i) - \Delta_{1i}). \tag{4.6}$$

Regarding the latent error distribution, one gets from Lemma 5.9 on [p.120] of Groeneboom and Wellner (1992) or Lemma 3.1 on [p.1423] of Groeneboom and Hendrickx (2018) the following uniform convergence at the cubic root rate (modulo the logarithm factor):

$$\| \hat{F}_n(u; \hat{\beta}_n) - F_0(u) \|_\infty = O_p(\log n \times n^{-1/3}); \tag{4.7}$$

The large sample property of $\hat{\alpha}_n$ is more complicated and it is our main focus. We consider the following random map

$$\Psi_n(\alpha) = \frac{1}{n}\sum_{i=1}^{n}\left[1 - \Delta_{3i} - \hat{F}_n(X_i'\hat{\beta}_n + \alpha; \hat{\beta}_n)\right] \tag{4.8}$$

and its probability limit

$$\Psi(\alpha) = \int [1 - \Delta_3 - F_0(X'\beta_0 + \alpha)]\, dP. \tag{4.9}$$

First, we prove the uniform convergence by the Glivenko-Cantelli Theorem (Van Der Vaart and Wellner (1996)), which leads to $\sup_\alpha |\Psi_n(\alpha) - \Psi(\alpha)| \to 0$. We then verify the population criterion function has a unique root and the sample analog has a zero-crossing point with probability tending to 1.

Theorem 4.1 (Consistency of the two-stage estimator)**.** *Suppose Conditions (1)-(6) hold. Then we have: (i). $\hat{\beta}_n$ obtained from equation (3.2) is a consistent estimator of β_0; (ii). for all large n, the unique zero-crossing $\hat{\alpha}_n$ of $\Psi_n(\alpha)$ exists with a probability tending to one and it is a consistent estimator of α_0.*

Because our estimation procedure belongs to the general Z-estimation with bundled parameter and nuisance functional components, we prove the root-n rate and asymptotic normality of $\hat{\alpha}_n$ following the route in Nan and Wellner (2013). Unlike the examples in Nan and Wellner (2013), which have nuisance nonparametric components either estimable with root-n rate or subject to certain smoothness restriction, the nonparametric part is estimated utilizing shape restriction in our model. The crux of our proof is to determine the asymptotic contribution of the estimated latent distribution to the threshold parameter.

Theorem 4.2 (Asymptotic normality of the two-stage estiamtor)**.** *Under Conditions (1)-(6), the following linear representations hold:*

$$\sqrt{n}\left(\hat{\beta}_n - \beta_0\right) = \mathbb{G}_n\left[\psi_{\beta_0}(Z_i)\right] + o_p(1), \text{ and} \tag{4.10}$$

$$\sqrt{n}\left(\hat{\alpha}_n - \alpha_0\right) = V_{\alpha_0}^{-1}\mathbb{G}_n\left[(\psi_0 + \psi_{F_0} + V_{\beta_0}\psi_{\beta_0})(Z_i)\right] + o_p(1), \tag{4.11}$$

where

$$\psi_{\beta_0}(Z_i) = H_{\beta_0}^{-1}(X_i - \mathbb{E}[X_i|U_i])(F_0(U_i) - \Delta_{1i}), \tag{4.12}$$

$$\psi_0(Z_i) = \left[1 - F_0(U_i + \alpha_0) - \Delta_{3i}\right], \tag{4.13}$$

$$\psi_{F_0}(Z_i) = \frac{g_0(U_i - \alpha_0)[\Delta_1 - F_0(U_i)]}{g_0(U_i)}. \tag{4.14}$$

Therefore,

$$\sqrt{n}\left(\hat{\beta}_n - \beta_0\right) \;\Rightarrow\; \mathbb{N}(0, \Omega_{\beta_0}), \text{ and} \tag{4.15}$$

$$\sqrt{n}\left(\hat{\alpha}_n - \alpha_0\right) \;\Rightarrow\; V_{\alpha_0}^{-1} \times \mathbb{N}(0, \Omega_{\alpha_0}), \tag{4.16}$$

where $\Omega_{\beta_0} = \mathbb{E}\left[\psi_{\beta_0}\psi_{\beta_0}'\right]$ *and* $\Omega_{\alpha_0} = \mathbb{E}\left[(\psi_0 + \psi_{F_0} + V_{\beta_0}\psi_{\beta_0})^2\right]$.

Intuitively speaking, the linear representation for the threshold estimator $\hat{\alpha}_n$ involves three parts: the oracle influence function ψ_0 given true β_0 and F_0, the effect from the estimation of F_0 encoded in ψ_{F_0}, and the effect from the estimation of β_0 collected in ψ_{β_0}. Despite the fact that we have a closed-form representation here, a brute-force estimation of the asymptotic variance involves some density function such as $g_0(\cdot)$ or $f_0(\cdot)$. This motivates us to propose a simple bootstrap approach that circumvents the obstacle in Section 4.3.

Given the linear representation for both $\hat{\alpha}_n$ and $\hat{\beta}_n$, an immediate corollary is the joint asymptotic normality for $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n')'$ as follows. To simplify the presentation, we abuse the notation somewhat by setting $\psi_{\alpha_0} \equiv V_{\alpha_0}^{-1}(\psi_0 + \psi_{F_0} + V_{\beta_0}\psi_{\beta_0})$.

Corollary 4.1. *Under Conditions (1)-(6), we have*

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \Rightarrow \mathbb{N}(0, \Sigma_0), \tag{4.17}$$

where the asymptotic covariance matrix is $\Sigma_0 = \mathbb{E}[(\psi_{\alpha_0}, \psi_{\beta_0}')'(\psi_{\alpha_0}, \psi_{\beta_0}')]$.

Remark 4.1. *In the seminal works of Newey (1994) and Chen, Linton, and Van Keilegom (2003), general theorems are presented for semiparametric estimators concerning the root-n consistency and asymptotic normality, which involves some first-stage nonparametric estimation. Maintained as a high-level assumption, Chen, Linton, and Van Keilegom (2003) assume that*

$$\sqrt{n}\left[S_n(\theta_0) + \Gamma(\theta_0)[\hat{F}_n - F_0]\right] \Rightarrow \mathbb{N}(0, \Sigma), \tag{4.18}$$

for some finite positive definite matrix Σ *without specifying how the nonparametric component* F_0 *is estimated in the first stage. In (4.18),* $S_n(\theta_0)$ *stands for the (normalized) oracle score function for the parametric part, whereas the direction derivative* $\Gamma(\theta_0)[\hat{F}_n - F_0]$ *encodes the estimation effect of the nonparametric component. In an earlier celebrated paper, Newey (1994) also directly assumes the following linear representation holds*

$$\sqrt{n}\left[\Gamma(\theta_0)[\hat{F}_n - F_0]\right] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_i + o_p(1), \tag{4.19}$$

for some zero-mean and square integrable random variables ψ_i. *The illustrative examples in Newey (1994) or Chen, Linton, and Van Keilegom (2003) are about nonparametric components with sufficient smoothness restrictions and estimated by sieve or kernel-type* *estimators. The crux of our theoretical investigation is to show that certain linear functional of the shape restricted nonparametric estimator is asymptotically normal. The verification of (4.18) or (4.19) in our context requires considerable more efforts than those examples in the aforementioned works mainly due to the fact that the isotonic estimator or the NPMLE is neither smooth nor linear.*

4.2. Asymptotic Properties of The Joint Estimator

The asymptotic analysis of our joint estimation method is more involved. There are mainly two differences compared with the results related to the two-stage estimation. First, both $\tilde{\alpha}_n$ and $\tilde{\beta}_n$ are determined in the second stage estimation together, so the calculation of the Hessian matrix is done with respect to both components; see Lemma S2.11 in the supplemntal note. More importantly, it is much more challenging to pin down the influence function capturing the effect of the NPMLE $\tilde{F}_n$ (expressed via some linear functional of $\tilde{F}_n$). Unlike the binary choice case where the isotonic estimator can be characterized as a left-continuous slope of the greatest convex minorant of certain cusum diagram, such an explicit interpretation is lacking for NPMLE using entire ordered response data. Thus, we seek an alternative characterization that builds on a sequence of research by Van de Geer (1997) and Geskus and Groeneboom (1996, 1997, 1999) for the interval censored data (case 2) and combines it with an analytic argument using empirical process theory as in Groeneboom and Hendrickx (2018).

First, we set up necessary notations. Denote the probability limit of the NPMLE as

$$F_0(u;\alpha,\beta) \equiv \mathbb{E}[\Delta_1|X'\beta+\alpha=u] = \mathbb{E}[\Delta_1|U=u-\alpha]. \tag{4.20}$$

Consider the following matrix

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

which leads to the Hessian matrix of our estimator when evaluated at the true parameter values (α_0,β_0):

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

The fundamental theory related to the linear (or more generally differentiable) functional is given by Van der Vaart (1991). We also refer readers to Chapter 3 in Groeneboom and Wellner (1992) and Chapter 25 in Van Der Vaart (1998) for more comprehensive discussions. We draw on Geskus and Groeneboom (1996, 1997, 1999) where the authors develop a systematic approach to characterize the linear functional of NPMLE for the interval censored data (case 2). For that purpose, we define $c_1(u) = -\mathbb{E}[X|u]F_0(u)f_0(u)$, $c_3(u) = -F_0(u+\alpha_0)f_0(u)$, 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

$$\tilde{\kappa}_{F_0}(u) = c(u) - \int c(v)dF_0(v). \tag{4.21}$$

The canonical gradient is also known as the "efficient influence function" and is supposed to belong to the space $L_2^0(F)$; i.e., the space of square integrable functions satisfying $\int adF = 0$, Groeneboom and Wellner (1992). 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

$$L^*\phi_F = \tilde{\kappa}_F, \tag{4.22}$$

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 (4.23). We further denote its derivative by $\tilde{\kappa}_F'$.

To present the solution ϕ_{F_0} for the true distribution function, we denote $x_0 = x'\beta_0$ 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

$$L[a](x_0,\delta_1,\delta_2) = \frac{\delta_1\int_{C_L}^{x_0} adF}{F(x_0)} + \frac{\delta_2\int_{x_0}^{x_0+\alpha_0} adF}{F(x_0+\alpha_0)-F(x_0)} - \frac{(1-\delta_1-\delta_2)\int_{x_0+\alpha_0}^{C_U} adF}{1-F(x_0+\alpha_0)} \tag{4.23}$$

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

(4.24)

$$L^*[b](x_0) = \int_{x_0}^{C_U} b(u,1,0)g_0(u)du + \int_{x_0-\alpha}^{x_0} b(u,0,1)g_0(u)du + \int_{C_L}^{x_0-\alpha} b(u,0,0)g_0(u)du.$$

Let $\phi_F(u) \equiv \int_{C_L}^{u} a(v)dF(v)$ denote the integrated score function, then we have

(4.25)

$$L^*L[a](x_0) = \int_{x_0}^{C_U} \frac{\phi(u)}{F(u)}g_0(u)du + \int_{x_0-\alpha_0}^{x_0} \frac{\phi(u+\alpha_0)-\phi(u)}{F(u+\alpha_0)-F(u)}g_0(u)du + \int_{C_L}^{x_0-\alpha_0} \frac{\phi(u+\alpha_0)}{1-F(u+\alpha_0)}g_0(u)du.$$

Then one can solve the equation and obtain

$$\begin{aligned} L[a](u,\delta_1,\delta_2) = &-\delta_1\left((1-F_0(u))\omega(u) + (1-F_0(u+\alpha_0)\omega(u+\alpha_0)\right) \\ &+\delta_2\left(F_0(u)\omega(u) - (1-F_0(u+\alpha_0)\omega(u+\alpha_0)\right) \\ &+(1-\delta_1-\delta_2)\left(F_0(u)\omega(u) + F_0(u+\alpha_0)\omega(u+\alpha_0)\right), \end{aligned}$$

where $\omega(u) \equiv \frac{c'(u)}{g_0(u)}$. Moreover, the influence function takes the following form (see Example 4.2 of Van de Geer (1997)):

(4.26)

$$\phi_{F_0}(u) = \begin{cases} -F_0(u)\left((1-F_0(u))\omega(u) + (1-F_0(u+\alpha_0)\omega(u+\alpha_0)\right), & \text{for} \quad C_L \le u \le \alpha_0 \\ (1-F_0(u))\left(F_0(u)\omega(u) + F_0(u-\alpha_0)\omega(u-\alpha_0)\right), & \text{for} \quad \alpha_0 \le u \le C_U. \end{cases}$$

We need one more set of assumptions to guarantee the asymptotic normality for the linear functional of NPMLE; see [p.31] in Van de Geer (1997).

Condition 7. We assume $\omega(u)$ is uniformly bounded for all u in the support. Moreover, the following ratios are all uniformly bounded:

$$\sup_u \left|\frac{\omega'(u)}{f_0(u)}\right| \le c, \quad \sup_u \left|\frac{\omega'(u)}{f_0(u+\alpha_0)}\right| \le c,$$
$$\sup_u \left|\frac{\omega'(u+\alpha_0)}{f_0(u)}\right| \le c, \quad \sup_u \left|\frac{\omega'(u+\alpha_0)}{f_0(u+\alpha_0)}\right| \le c,$$
$$\sup_u \left|\frac{f_0(u+\alpha_0)}{f_0(u)}\right| \le c, \quad \sup_u \left|\frac{f_0(u)}{f_0(u+\alpha_0)}\right| \le c,$$

for some universal finite constant c. Also, for any α in the parameter space, we assume $F_0(u+\alpha) - F_0(u)$ is uniformly bounded away from zero for any u in the support.

Denote the stacked estimator for the finite dimensional parameter as $\tilde{\theta}_n \equiv (\tilde{\alpha}_n, \tilde{\beta}_n')'$ and the true unknown parameter as $\theta_0 \equiv (\alpha_0, \beta_0')'$.

Theorem 4.3 (Consistency of the joint estimator)**.** *Under our Conditions (1)-(7), we have for all large n, the zero-crossing $\tilde{\theta}_n$ for $\Psi_n\left(\tilde{\theta}_n\right)$ exists with a probability tending to one and is a consistent estimator of θ_0.*

Theorem 4.4 (Asymptotic normality of the joint estimator)**.** *Under our Conditions (1)-(7), we have*

$$\sqrt{n}\left(\tilde{\theta}_n - \theta_0\right) \Rightarrow \mathbb{N}(0, \tilde{\Sigma}_0), \tag{4.27}$$

where $\tilde{\Sigma}_0 = H_0^{-1}\mathbb{E}[(\phi_0' + \phi_{F_0}')'(\phi_0' + \phi_{F_0}')]H_0^{-1}$, ϕ_{F_0} defined in equation (4.26) and

$$\phi_0 = \begin{pmatrix} [\Delta_{1i} - F_0(X_i'\beta_0)]X_i \\ \Delta_{3i} - 1 + F_0(X_i'\beta_0 + \alpha_0) \end{pmatrix}.$$

Remark 4.2. *It is known that the NPMLE is indeed more efficient than the isotonic estimator only using binary choices data in the sense that both $n^{1/3}(\tilde{F}_n(t;\alpha_0,\beta_0) - F_0(t))$ and $n^{1/3}(\hat{F}_n(t;\beta_0) - F_0(t))$ converge to the Chernoff distribution yet with different scaling constant terms. Specifically, the NPMLE has a smaller asymptotic variance than the one* *associated with the isotonic estimator; see Exercise 4.27 in Chapter 4 of Groeneboom and Jongbloed (2014). Our simulation results also confirm this theoretical claim.*

Remark 4.3. *Apropos of the asymptotic covariance matrices for finite dimensional parameters, the comparison between our two-stage and joint estimation is not obvious analytically, as both influence functions are very complicated. As the joint approach simultaneously estimates α_0 and β_0, one may naturally expect that it works better. This is supported by our simulation studies. Overall, performances of both estimators are reliable and not subject to the sensitivity of tuning parameter selection.*

Remark 4.4. *Referring to the semiparametric efficient score function in Coppejans (2007), it is inevitable to use either the kernel or sieve approach, because it involves the density function of latent error. Embedding shape restriction, we suggest an efficient estimation procedure making use of the maximum smoothed likelihood estimator (MSLE) in Groeneboom (2014), which can also be viewed as the generalization of Klein and Spady (1993) to the ordered response data. Consider a kernel density function $k(\cdot)$ and $k_{b_n}(\cdot) \equiv k(\cdot/b_n)/b_n$ with the bandwidth equal to b_n. Define $\bar{h}_{jn}(t;\beta) = \frac{1}{n}\sum_{i=1}^n k_{b_n}(X_i'\beta, t)\Delta_{ji}$, for $j = 1, 2, 3$. The MSLE $\bar{F}_{nb_n}(t;\alpha,\beta)$ is now defined by maximizing*

$$\int \bar{h}_{1n}(t;\beta)\log F(t)dt + \int \bar{h}_{3n}(t+\alpha;\beta)\log[1-F(t+\alpha)]dt + \int \bar{h}_{2n}(t;\beta)\log[F(t+\alpha)-F(t)]dt.$$

Thereafter, the estimators $(\bar{\alpha}_n, \bar{\beta}_n)$ for the finite dimensional parameters are defined by maximizing the profile (smoothed) likelihood function:

$$(\bar{\alpha}_n, \bar{\beta}_n) = \arg\max_{\alpha,\beta} \mathbb{L}_n(\bar{F}_{nb_n}(\cdot;\alpha,\beta);\alpha,\beta). \tag{4.28}$$

We conjecture that this (kernel) smoothed maximum likelihood procedure delivers semiparametric efficient estimators for (α_0, β_0), based on the results in Coppejans (2007) and Groeneboom (2014). One notable advantage over the sieve method in Coppejans (2007) is that the estimated distribution is monotone by construction, whereas Coppejans (2007) has to impose additional restrictions on spline coefficients to accommodate the monotonicity.

Remark 4.5. *For some empirical questions, one might be interested in the marginal effect*[15] *associated with the ordered response model. In this case, one needs the joint normality of the estimated finite dimensional parameter and the distribution function evaluated at a certain point. However, the point-wise asymptotic distribution of the isotonic estimator $\hat{F}_n$ or the NPMLE $\tilde{F}_n$ is not normal. Instead, they both converge to the Chernoff distribution*[16]

[15] We want to thank one anonymous referee who raised this issue.

[16] A random variable is said to follow the Chernoff distribution if it is the last time where standard two-sided Brownian motion minus the parabola t^2 reaches its maximum; see Section 3.9 in Groeneboom and Jongbloed (2014).

with the cubic root rate. In order to restore the asymptotic normality and improve the rate of convergence for the distribution function, one could resort to the smoothed maximum likelihood estimator (SMLE) or the maximum smoothed likelihood estimator (MSLE) in Groeneboom, Jongbloed, and Witte (2010) and Groeneboom (2014).

4.3. Bootstrap Inference

In this part, we propose corresponding novel bootstrap procedures for conducting inference for our semiparametric estimators. In a closely related paper for binary choice data, Groeneboom and Hendrickx (2017) first prove the bootstrap validity for the regression coefficient using Efron's multinomial weights (Efron (1979)) based on the shape-restricted estimation in Groeneboom and Hendrickx (2018). We establish the bootstrap consistency with more general exchangeable bootstrap weights $(M_{ni})_{i=1}^{n}$. Commonly used exchangeable bootstrap schemes include (i) Nonparametric bootstrap in which the bootstrap weights $M_n = (M_{n1}, ..., M_{nn})'$ follow the multinomial distribution $Mult\,(n, (n^{-1}, ..., n^{-1}))$; (ii) Bayesian bootstrap in which the bootstrap weights $M_{ni} = \omega_i / \left(\sum_{i=1}^{n} \omega_i\right)$ for $1 \leq i \leq n$ and ω_i as 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 $1 \leq i \leq n-h$ and let $M_{nj} = w_{nR_n(j)}$ where R_n is a random permutation uniformly over $\{1, ..., n\}$ (Wu, 1990). All these bootstrap procedures satisfy the following Condition 8. A detailed verification can be found in Cheng and Huang (2010).

Condition 8. **(i)** The vector $M_n = (M_{n1}, ..., M_{nn})'$ is exchangeable for all n.

(ii) $M_{ni} \geq 0$ for all n, i and $\sum_{i=1}^{n} M_{ni} = n$ for all n.

(iii) For some positive constant $C < \infty$, $\limsup_{n\to\infty} \|M_{n1}\|_{2,1} \leq C$, where $\|M_{n1}\|_{2,1} = \int_0^\infty \sqrt{\Pr\left(M_{n1} \geq u\right)} du$.

(iv) $\lim_{\lambda\to\infty} \limsup_{n\to\infty} \sup_{t\geq\lambda} t^2 \Pr\left(M_{n1} > t\right) = 0$.

(v) $(1/n) \sum_{i=1}^{n} \left(M_{ni} - 1\right)^2 \to_p 1$.

In analog with the notations from the empirical process theory, we let $\mathbb{P}_n^* f = n^{-1} \sum_{i=1}^{n} M_{ni} f\left(Z_i\right)$, and $\mathbb{G}_n^* f = n^{-1/2} \sum_{i=1}^{n} \left(M_{ni} - 1\right) f\left(Z_i\right)$. Before stating the theoretical results for the bootstrapped quantities, we should be explicit about the underlying probability space and source of randomness. There are two sources of randomness coming from the observed data and the bootstrap weight M_n. The following set of definitions and notations are adapted from Cheng and Huang (2010). We have the product probability space

$$\left(\mathcal{Z}^\infty \times \mathcal{M}, \mathcal{A}^\infty \times \Omega, P_{ZM}\right)$$

for the joint randomness from observed data and bootstrap weights. Furthermore, the bootstrap weights are independent from the sample observations, i.e., $P_{ZM} = P_Z \times P_M$.

Definition 4.1. *For a real-valued random variable Δ_n, we define (i) $\Delta_n = o_{P_M}(1)$, in P_Z-probability if for any $\varepsilon, \delta > 0$*

$$P_Z \left\{ P_{M|Z} \left(|\Delta_n| > \varepsilon \right) > \delta \right\} \to 0.$$

(ii) We define $\Delta_n = O_{P_M}(1)$, in P_Z-probability if for any $\delta > 0$ there exists a C s.t.

$$P_Z \left\{ P_{M|Z} \left(|\Delta_n| > C \right) > \delta \right\} \to 0.$$

4.3.1. **Bootstrap Two-stage Estimation.** We now describe our bootstrap estimator for the regression coefficients and threshold parameter. Note that bootstrap for regression coefficients in the binary choice model can be found in Groeneboom and Hendrickx (2017).
Stage 1(i)*. First of all, the bootstrap MLE $\hat{F}_n^*(\cdot, \beta)$ is computed using the weighted cumulative sum diagram formed by the point $(0, 0)$ and

$$\left(\sum_{j=1}^{i} M_{n(j)}^{(\beta)}, \sum_{j=1}^{i} M_{n(j)}^{(\beta)} \Delta_{1,(j)}^{(\beta)} \right),$$

where $M_{n(i)}^{(\beta)}$ corresponds to the weight attached to $U_{(i)}^{(\beta)}$.
Stage 1(ii)*. The bootstrap estimator of the regression coefficient $\hat{\beta}_n^*$ is defined as the zero-crossing point of the following estimating equations:

$$\frac{1}{n} \sum_{i=1}^{n} M_{ni} X_i \left[\Delta_{1i} - \hat{F}_n^*(X_i' \hat{\beta}_n^*; \hat{\beta}_n^*) \right] = 0. \tag{4.29}$$

Stage 2*. Finally, the bootstrap version $\hat{\alpha}_n^*$ is then defined as the zero-crossing point of the following estimating equation:

$$\frac{1}{n} \sum_{i=1}^{n} M_{ni} \left[1 - \Delta_{3i} - \hat{F}_n^*(X_i' \hat{\beta}_n^* + \hat{\alpha}_n^*; \hat{\beta}_n^*) \right] = 0. \tag{4.30}$$

Given Condition 8, the bootstrap estimate $\hat{F}_n^*$ is a step-wise monotone function. Therefore, the estimating equation for $\hat{\alpha}_n^*$ is again monotone so that the computational advantage of our approach is amplified along the bootstrap replications. Regarding the theoretical underpinning, one could easily prove that $\hat{\alpha}_n^* \to \alpha_0$ conditional on observations $(Z_1, \cdots, Z_n)$ almost surely. We characterize the conditional weak limit for $\hat{\alpha}_n^*$ in the next theorem. For completeness, we also state the result for $\hat{\beta}_n^*$, which has been established by Groeneboom and Hendrickx (2017)[p3464-3465].

Theorem 4.5 (Bootstrap validity for the two-stage estimator)**.** *Suppose Conditions (1)-(6) and Condition (8) hold. For the bootstrap estimators $\hat{\beta}_n^*$ and $\hat{\alpha}_n^*$ with exchange weights*

$(M_{n1}, ..., M_{nn})$, *we have that*

$$\left|\hat{\beta}_n^* - \hat{\beta}_n\right| = O_{P_M}\left(n^{-1/2}\right), \text{ and } |\hat{\alpha}_n^* - \hat{\alpha}_n| = O_{P_M}\left(n^{-1/2}\right),$$

in P_Z-*probability. Furthermore,*

$$\sqrt{n}\left(\hat{\beta}_n^* - \hat{\beta}_n\right) \Rightarrow \mathbb{N}(0, \Omega_{\beta_0}), \tag{4.31}$$

$$\sqrt{n}\left(\hat{\alpha}_n^* - \hat{\alpha}_n\right) \Rightarrow V_{\alpha_0}^{-1} \times \mathbb{N}(0, \Omega_{\alpha_0}), \tag{4.32}$$

conditional on observations $(Z_1, \cdots, Z_n)$ *almost surely.*

A direct consequence of the above theorem is the validity of percentile bootstrap confidence interval (see Cheng and Huang (2010)). The lower p-th quantile of the bootstrap distribution of $\hat{\alpha}_n^*$ is the quantity $\tau_{n,\alpha}^*(p)$ satisfying $\tau_{n,\alpha}^*(p) \equiv \inf\{\tau : P_{M|Z}(\hat{\alpha}_n^* \leq \tau) \geq p\}$. Similar, we can define $\tau_{n,\beta_k}^*(p) \equiv \inf\{\tau : P_{M|Z}(\hat{\beta}_{n,k}^* \leq \tau) \geq p\}$ for the kth component of $\hat{\beta}_n^*$, $k = 2, ..., K$. Then the percentile-type bootstrap confidence intervals can be constructed as

$$\widehat{BC}_{\alpha}(p) = \left[\tau_{n,\alpha}^*(p/2), \tau_{n,\alpha}^*(1-p/2)\right], \tag{4.33}$$

and

$$\widehat{BC}_{\beta,k}(p) = \left[\tau_{n,\beta_k}^*(p/2), \tau_{n,\beta_k}^*(1-p/2)\right], \text{ for } k = 2, \cdots, K. \tag{4.34}$$

Corollary 4.2. *Suppose Conditions (1)-(8) hold, then we have*

$$P_{ZM}\left(\alpha_0 \in \widehat{BC}_{\alpha}(p)\right) \to 1-p, \tag{4.35}$$

and

$$P_{ZM}\left(\beta_{0k} \in \widehat{BC}_{\beta,k}(p)\right) \to 1-p, \quad \text{for } k = 2, \cdots, K, \tag{4.36}$$

as $n \to \infty$.

4.3.2. **Bootstrap Joint Estimation.** Now we describe the bootstrap inference for our joint estimation approach.

Joint Estimation*. For any α and β, we can derive the bootstrap NPMLE for $F(\cdot; \alpha, \beta)$ based on the *ordered response* data:

$$\begin{aligned} &\tilde{F}_n^*(\cdot; \alpha, \beta) \\ =& \arg\max_F \sum_{i=1}^n M_{ni}\left[\Delta_{1i}\log F(X_i'\beta) + \Delta_{2i}\log\left(F(X_i'\beta+\alpha) - F(X_i'\beta)\right) + \Delta_{3i}\log\left(1 - F(X_i'\beta+\alpha)\right)\right], \end{aligned}$$

which exists. Given $\tilde{F}_n^*(\cdot;\alpha,\beta)$ from the previous step, we define $(\tilde{\alpha}_n^*, \tilde{\beta}_n^*)$ which are the zero-crossing points of the estimating equations simultaneously:

$$\frac{1}{n}\sum_{i=1}^{n} M_{ni}X_i\left[\Delta_{1i} - \tilde{F}_n^*(X_i'\tilde{\beta}_n;\tilde{\alpha}_n,\tilde{\beta}_n)\right] = 0, \tag{4.37}$$

$$\frac{1}{n}\sum_{i=1}^{n} M_{ni}\left[1 - \Delta_{3i} - \tilde{F}_n^*(X_i'\tilde{\beta}_n + \tilde{\alpha}_n;\tilde{\alpha}_n,\tilde{\beta}_n)\right] = 0.$$

Theorem 4.6 (Bootstrap validity for the joint estimator)**.** *Suppose Conditions (1)-8) hold. For the bootstrap estimator $\tilde{\theta}_n^*$ with exchange weights $(M_{n1},...,M_{nn})$, we have that*

$$\left|\tilde{\theta}_n^* - \tilde{\theta}_n\right| = O_{P_M}\left(n^{-1/2}\right),$$

in P_Z-probability. Moreover, we have

$$\sqrt{n}\left(\tilde{\theta}_n^* - \tilde{\theta}_n\right) \Rightarrow \mathbb{N}(0,\tilde{\Sigma}_0), \tag{4.38}$$

conditional on observations $(Z_1,\cdots,Z_n)$, almost surely.

Similarly as the two-stage estimator (part 4.3.1), we can construct the percentile-type bootstrap confidence intervals $\widetilde{BC}_\alpha$ and $\widetilde{BC}_{\beta,k}$ of the regression coefficients and threshold parameter for the joint estimator. The coverage properties are summarized in the following corollary.

Corollary 4.3. *Suppose Conditions (1)-(8) hold, then we have*

$$P_{ZM}\left(\alpha_0 \in \widetilde{BC}_\alpha(p)\right) \to 1-p, \tag{4.39}$$

and

$$P_{ZM}\left(\beta_{0k} \in \widetilde{BC}_{\beta,k}(p)\right) \to 1-p, \quad \textit{for } k = 2,\cdots,K, \tag{4.40}$$

as $n \to \infty$.

5. Numerical Results

In this section, we conduct Monte Carlo simulations to evaluate the finite sample performances of our proposals and other semiparametric approaches described in Section 3.3. The simulation designs follow the interdependent duration model of Honoré and de Paula (2010). We also illustrate the usefulness of all these methods by applying them to the Danish joint retirement data in An, Christensen, and Gupta (2004).

5.1. Monte Carlo Studies

Recall the model setup in Section 2 of the interdependent durations model proposed by Honoré and de Paula (2010): two players respectively decide (T_1, T_2) as the time between switching from an initial activity to an alternative activity. The utility flows from the initial activity for players 1 and 2 are random variables (ϵ_1, ϵ_2) with a joint CDF $G(\cdot,\cdot)$. The utility flow from the alternative activity at time t gives player j an amount $t^a \exp(X_j'\beta_0)\exp[\alpha^* \mathbb{I}\{T_k \le t\}]$ for $j, k \in \{1, 2\}$ and $j \ne k$. Here we consider individuals with covariates denoted by $X_j = (1, X_{1j}, X_{2j}, X_{3j}, X_4)$, where the last coordinate is assumed to be the same for both players, and write the corresponding regression coefficient as $\beta_0 = (\beta_{0o}, \beta_{01}, \beta_{02}, \beta_{03}, \beta_{04})'$.

The main complication of the model is the characterization of (T_1, T_2) as the Nash equilibrium outcomes. Based on the value of ϵ_1/ϵ_2, the equilibrium (T_1, T_2) can be categorized into three cases: $T_1 < T_2$, $T_1 = T_2$, and $T_1 > T_2$. Specifically, Honoré and de Paula (2010) characterize the equilibrium (T_1, T_2) as follows: if $\epsilon_1/\epsilon_2 < \exp[(X_1 - X_2)'\beta_0 - \alpha^*]$, then $T_1 = [\epsilon_1/\exp(X_1'\beta_0)]^{1/a}$ and $T_2 = [\epsilon_2/\exp(X_2'\beta_0 + \alpha^*)]^{1/a}$; if $\exp[(X_1 - X_2)'\beta_0 - \alpha^*] < \epsilon_1/\epsilon_2 < \exp[(X_1 - X_2)'\beta_0 + \alpha^*]$, then $T_1 = T_2$ and can take any value between

$$\left[\max\left\{\left[\epsilon_1/e^{(X_1'\beta_0+\alpha^*)}\right]^{1/a}, \left[\epsilon_2/e^{(X_2'\beta_0+\alpha^*)}\right]^{1/a}\right\}, \min\left\{\left[\epsilon_1/e^{(X_1'\beta_0)}\right]^{1/a}, \left[\epsilon_2/e^{(X_2'\beta_0)}\right]^{1/a}\right\}\right];$$

if $\epsilon_1/\epsilon_2 > \exp[(X_1-X_2)'\beta_0+\alpha^*]$, then $T_1 = [\epsilon_1/\exp(X_1'\beta_0 + \alpha^*)]^{1/a}$ and $T_2 = [\epsilon_2/\exp(X_2'\beta_0)]^{1/a}$. When $\exp[(X_1-X_2)'\beta_0-\alpha^*] < \epsilon_1/\epsilon_2 < \exp[(X_1-X_2)'\beta_0+\alpha^*]$, the model produces multiple equilibria so that in case of the simultaneous failure $(T_1 = T_2)$ there exists a range of possible values. However, the equilibrium selection rule does not matter here when it comes to estimating the regression coefficient β_0 or the interaction effect α^*, because the conditional probability $P\{T_1 = T_2|X_1, X_2\}$ remains the same. Therefore, one can transform the above interdependent durations model into the ordered responses model with $Y = 1$ if $T_1 < T_2$, $Y = 2$ if $T_1 = T_2$ and $Y = 3$ if $T_1 > T_2$. Then the conditional probabilities become:

$$P\{Y \le 1|X_1, X_2\} = H(\beta_{01}(X_{11} - X_{12}) + \beta_{02}(X_{21} - X_{22}) + \beta_{03}(X_{31} - X_{32}) - \alpha^*), \tag{5.1}$$

$$P\{Y \le 2|X_1, X_2\} = H(\beta_{01}(X_{11} - X_{12}) + \beta_{02}(X_{21} - X_{22}) + \beta_{03}(X_{31} - X_{32}) + \alpha^*), \tag{5.2}$$

where $H(w) = P\{\log(\epsilon_1/\epsilon_2) \le w\}$. Furthermore, $H(\cdot)$ is an arbitrary distribution function if no assumption is made regarding (ϵ_1, ϵ_2). Note that the common effect from covariates' effect $\beta_{0o}+\beta_{04}X_4$ is differenced out here. Also, we normalize $\beta_{01} = 1$ for the identification purpose. In our simulations, we set $(a, \beta_o, \beta_1, \beta_{02}, \beta_{03}, \beta_{04}, \alpha^*) = (1.35, -4.00, 1.00, 1.00, 1.00, 0.50, 1.00)$. We consider two cases for $G(\cdot,\cdot)$: the joint distribution of two independent unit exponential variables or two independent log normal variables. Variables $(X_{11}, X_{12}, X_{31}, X_{32}, X_4)$

are independent standard normal random variables and (X_{21}, X_{22}) are independent $\chi^2(1)$ random variables (normalized to a mean zero and variance one). Our Monte-Carlo study involves 1,000 simulations of the foregoing models. The sample sizes are $n = 250, 500$, and 750.

Table 1 presents the finite sample bias, root of the mean square error (RMSE), and median absolute deviation (MAE) of our Isotonic two-stage estimator and NPMLE-based joint estimator for $(\beta_{02}, \beta_{03}, \alpha^*)$. The error terms (ϵ_1, ϵ_2) are two independent unit exponential or log-normal random variables. The initial values of $(\beta_2, \beta_3, \alpha^*)'$ are set as $(0.2, 0.2, 0.1)'$ and the estimating equations are solved using R package BB. As a comparison, Table 2 presents the finite sample results for three alternative semiparametric estimators described in Remarks 3.1 to 3.3; i.e., Klein and Sherman (2002)'s kernel-based approach (K-S), Melenberg and Van Soest (1996)'s smoothed maximum score estimator (SMS), and the rank estimator combining Cavanagh and Sherman (1998) and Chen (2002). For the K-S approach, the bandwidths (including a bandwidth, a pilot bandwidth, and a smoothing parameter in the damping function) are chosen following the guidelines of Klein and Sherman (2002).[17] The trimming parameter $\hat{\xi}$ for the quasi-likelihood function is set to be the 0.95th sample quantile of the Euclidean norm of covariates. We consider two values for trimming proportion for the target set: $p = 0.05$ and $p = 0.20$. For the SMS approach, we experiment two bandwidths $n^{-1/4}$ and $n^{-1/5}$. The former reduces bias by under-smoothing and the later is chosen according to the MSE-optimal rate. Table 3 examines point-wise estimates for the error distribution function $H(\cdot)$ evaluated at 5 equally spaced points in the support of $(\log \epsilon_1 - \log \epsilon_2)$.

We first evaluate the performances of our two-stage estimator and joint estimator. Table 1 shows that bias, RMSE, and MAE for both estimators decrease when the sample size increases. NPMLE-based joint estimation yields smaller bias than the two-stage estimator. The RMSE's are close between two estimators, with the two-stage estimator slightly better for $(\beta_{02}, \beta_{03})'$ and the joint estimator slightly better for α^*. The lower variability of two-stage estimator is likely due to its computational simplicity. As shown in Figures 1 and 2, the estimating functions exhibit smoother variation in the two-stage procedure than in the joint procedure. Nevertheless, joint estimator still yields smaller RMSE for α^* due to the fact that it utilizes all the information across three categories by the simultaneous estimation. Also, the MAE of the joint estimator is slightly smaller than that of two-stage estimator. In terms of the nonparametric estimate of $H(\cdot)$, Table 3 shows that the

[17]The value δ in Klein and Sherman (2002) p669 is set to $1/6$. The rate of bandwidth α is set as the middle point of the allowed range $((3+\delta)/20, 1/6)$, see p670. The rate of the pilot bandwidth follows Lemma 5A of Klein and Sherman (2002). Finally, ϵ in the damping function is set as the middle point of the allowed range $(0, 1/40 - \delta/20)$, see p670.

NPMLE-based approach performs better in general than the isotonic estimator, especially for $H(1)$ and $H(2)$. This confirms the efficiency gain from utilizing additional information that differentiates the category with $Y = 2$ and the one with $Y = 3$ (see Remark 4.2).

We further compare our two-stage and joint estimator with other semiparametric estimators as reported in Table 2. The K-S estimator for α^* is sensitive to the trimming parameter p, which decides the subset used for averaging individual estimators (see Remark 3.1). It is clear that $p = 0.05$ does not sufficiently exclude individual estimators that perform poorly and thus leads to unreliable estimate for α^*. The performance for α^* greatly improves when $p = 0.20$; however, the RMSE and MAE are still larger than our two-stage and joint estimators. The SMS estimator has significantly larger RMSE and MAE than all other methods. This can not be fully explained by the under-smoothing bandwidth $n^{-1/4}$. When bandwidth takes the MSE-optimal rate $n^{-1/5}$, the RMSE and MAE of SMS decrease, yet still larger than other methods. The rank estimator described in Remark 3.3 exhibits remarkable finite sample performances in our simulations. It yields the smallest bias for regression coefficients and threshold parameter among all methods. However, referring to the RMSE for α^*, our two-stage and joint estimators still have an edge on the rank estimator. In sum, our simulation studies show that the isotonic two-stage estimator and the NPMLE-based joint estimator perform well and stable. Their tuning-parameter-free advantage is further enhanced by the robust finite sample performances. Namely, in terms of the RMSE, they clearly outperform the SMS in all scenarios we considered and are better than the K-S for most cases. Moreover, our estimators, especially the joint estimator, yield the smallest RMSE for α^*, which is the parameter of primary interest in the interdependent durations model.

Table 1. Finite sample performances of the isotonic two-stage estimator and the NPMLE-based joint estimator for $(\beta_{02}, \beta_{03}, \alpha^*)$.

		(ϵ_1, ϵ_2)	Exponential			Log-normal		
Methods	n		Bias	RMSE	MAE	Bias	RMSE	MAE
Two-stage	250	β_{02}	-.0658	.1985	.1448	-.0482	.1701	.1202
		β_{03}	-.0692	.1717	.1234	-.0547	.1582	.1199
		α^*	-.0663	.1470	.1092	-.0496	.1364	.0982
	500	β_{02}	-.0470	.1370	.0937	-.0345	.1255	.0841
		β_{03}	-.0485	.1259	.0883	-.0333	.1113	.0798
		α^*	-.0476	.1132	.0810	-.0342	.0978	.0715
	750	β_{02}	-.0425	.1163	.0822	-.0297	.0980	.0679
		β_{03}	-.0423	.1033	.0719	-.0294	.0861	.0607
		α^*	-.0376	.0907	.0643	-.0299	.0791	.0556
Joint	250	β_{02}	-.0337	.2079	.1412	-.0211	.1786	.1197
		β_{03}	-.0394	.1774	.1258	-.0285	.1647	.1166
		α^*	-.0516	.1463	.1031	-.0378	.1357	.0968
	500	β_{02}	-.0282	.1372	.0916	-.0184	.1272	.0835
		β_{03}	-.0316	.1276	.0889	-.0184	.1120	.0789
		α^*	-.0389	.1099	.0758	-.0281	.0953	.0689
	750	β_{02}	-.0280	.1190	.0828	-0160	.0998	.0694
		β_{03}	-.0300	.1048	.0717	-.0162	.0894	.0623
		α^*	-.0322	.0889	.0618	-.0246	.0776	.0542

Table 2. Finite sample performances of other semiparametric estimators, described in Remarks 3.1 to 3.3. Trimming parameter p in the K-S approach is used to exclude individual estimators for α^* that performs poorly; h is the bandwidth used in SMS approach.

		(ϵ_1, ϵ_2)	Exponential			Log-normal		
Methods	n		Bias	RMSE	MAE	Bias	RMSE	MAE
K-S	250	β_{02}	.0374	.2430	.1593	.0267	.1956	.1256
		β_{03}	.0373	.1943	.1251	.0201	.1585	.1000
$p = .05$		α^*	.1568	.2797	.1549	.1647	.2681	.1541
$p = .20$		α^*	.0697	.2104	.1214	.0570	.1791	.1058
	500	β_{02}	.0205	.1435	.0942	.0142	.1258	.0849
		β_{03}	.0164	.1267	.0794	.0121	.1075	.0671
$p = .05$		α^*	.1682	.2501	.1477	.1642	.2232	.1484
$p = .20$		α^*	.0446	.1405	.0824	.0378	.1160	.0738
	750	β_{02}	.0194	.1202	.0825	.0166	.0958	.0655
		β_{03}	.0090	.1003	.0622	.0080	.0847	.0563
$p = .05$		α^*	.1718	.2422	.1452	.1688	.2143	.1507
$p = .20$		α^*	.0347	.1130	.0680	.0356	.0937	.0593
SMS	250	β_{02}	.0776	.3093	.1658	.0574	.2683	.1419
$(h = n^{-1/4})$		β_{03}	.0728	.2791	.1645	.0559	.2326	.1359
		α^*	.0638	.3046	.1939	.0427	.2648	.1658
	500	β_{02}	.0534	.2100	.1310	.0379	.1960	.1193
		β_{03}	.0381	.1882	.1138	.0318	.1693	.1094
		α^*	.0365	.2152	.1468	.0250	.1847	.1201
	750	β_{02}	.0428	.1896	.1111	.0255	.1564	.1057
		β_{03}	.0267	.1574	.1001	.0236	.1410	.0869
		α^*	.0215	.1823	.1222	.0156	.1638	.1049
$(h = n^{-1/5})$	750	β_{02}	.0468	.1823	.1086	.0273	.1419	.0910
		β_{03}	.0303	.1449	.0877	.0243	.1277	.0764
		α^*	.0273	.1628	.1065	.0174	.1470	.0905
Rank	250	β_{02}	.0108	.2082	.1296	.0111	.1685	.1108
		β_{03}	.0112	.1725	.1142	.0118	.1478	.0977
		α^*	.0087	.1639	.1103	.0087	.1462	.0905
	500	β_{02}	.0160	.1311	.0900	.0090	.1090	.0711
		β_{03}	.0137	.1179	.0738	.0094	.0988	.0648
		α^*	.0114	.1171	.0750	.0054	.0974	.0651
	750	β_{02}	.0071	.1041	.0696	.0073	.0869	.0581
		β_{03}	.0042	.0909	.0609	.0056	.0790	.0522
		α^*	.0053	.0921	.0580	.0037	.0797	.0513

TABLE 3. Finite sample performances of point-wise estimators for the function $H(w)$, Isotoic estimator uses binary data, NPMLE uses entire ordered response data.

		(ϵ_1, ϵ_2)	Exponential			Log-normal		
Methods	n		Bias	RMSE	MAE	Bias	RMSE	MAE
Isotonic	250	$H(-2)$	-.0186	.0578	.0417	-.0223	.0528	.0430
		$H(-1)$	-.0053	.0829	.0580	-.0088	.0842	.0602
		$H(0)$	.0089	.1043	.0726	.0131	.1014	.0720
		$H(1)$	.0294	.1018	.0698	.0296	.1078	.0728
		$H(2)$	.0416	.0834	.0673	.0340	.0706	.0785
	750	$H(-2)$	-.0091	.0386	.0274	-.0089	.0354	.0260
		$H(-1)$	.0003	.0558	.0391	-.0041	.0551	.0359
		$H(0)$	.0083	.0682	.0483	.0103	.0714	.0474
		$H(1)$	.0154	.0693	.0476	.0137	.0679	.0462
		$H(2)$	.0172	.0550	.0380	.0189	.0482	.0368
NPMLE	250	$H(-2)$	-.0084	.0498	.0342	-.0073	.0466	.0328
		$H(-1)$	.0053	.0660	.0445	.0052	.0734	.0495
		$H(0)$	.0183	.0847	.0571	.0168	.0873	.0611
		$H(1)$	.0298	.0823	.0562	.0239	.0872	.0627
		$H(2)$	.0216	.0588	.0435	.0190	.0540	.0421
	750	$H(-2)$	-.0051	.0335	.0243	-.0032	.0312	.0221
		$H(-1)$	.0044	.0469	.0332	.0012	.0449	.0302
		$H(0)$	.0094	.0569	.0387	.0412	.0618	.0428
		$H(1)$	.0148	.0542	.0367	.0110	.0535	.0372
		$H(2)$	.0117	.0393	.0277	.0096	.0337	.0237
K-S	250	$H(-2)$	.0029	.0412	.0279	-.0021	. 0346	.0234
		$H(-1)$	.0040	.0574	.0404	.0019	.0561	.0376
		$H(0)$	.0004	.0690	.0479	.0075	.0744	.0518
		$H(1)$	-.0118	.0688	.0463	-.0131	.0703	.0431
		$H(2)$	-.0352	.0683	.0426	-.0425	.0662	.0399
	750	$H(-2)$	-.0003	.0240	.0161	-.0014	.0211	.0133
		$H(-1)$	.0032	.0337	.0231	.0010	.0343	.0230
		$H(0)$	.0041	.0412	.0280	.0081	.0453	.0298
		$H(1)$	-.0029	.0440	.0301	-.0042	.0401	.0265
		$H(2)$	-.0223	.0411	.0268	-.0300	.0407	.0285

Table 4 reports the empirical coverage proportion (CP) and the median length (ML) of the confidence intervals (CIs) based on the nonparametric bootstrap described in Section 4.3, for the our two-stage and the joint estimators. We observe that the coverage proportions of their nonparametric bootstrap based CIs are close to the nominal 95% level. For the threshold parameter α^*, the CIs of joint estimator has better coverage proportions and shorter median lengths than the two-stage version. Table 4 also summarizes the nonparametric bootstrap based CIs for other semiparemetric estimators: the K-S, the SMS, and the rank estimator. The under-smoothing bandwidth $n^{-1/4}$ is used for the SMS to reduce the bias in the asymptotic distribution (see Section 4.3.3 of Horowitz (2009)). The trimming proportion p is set to 0.2 in the K-S estimator. We observe that the SMS exhibits over-coverage in all scenarios and the median length of CIs are substantially longer than other methods. The K-S leads to under-coverage for α^* when the $n = 750$. The rank estimator performs well when $n = 750$; however, it suffers from under-coverage for the regression coefficients $(\beta_{02}, \beta_{03})'$ when $n = 250$ and 500.

TABLE 4. Coverage proportions of 95% nonparametric bootstrap based confidence intervals, sample size = n, CP = coverage proportion, ML = median length of the confidence interval, number of bootstrap replications = 200.

		(ϵ_1, ϵ_2) Exponential						(ϵ_1, ϵ_2) log-normal					
		$n = 250$		$n = 500$		$n = 750$		$n = 250$		$n = 500$		$n = 750$	
Methods		CP	ML	CP	ML	CP	ML	CP	ML	CP	ML	CP	ML
Two-stage	β_{02}	.974	.779	.955	.506	.945	.415	.978	.683	.958	.457	.941	.368
	β_{03}	.961	.677	.940	.454	.943	.373	.963	.597	.947	.409	.952	.332
	α^*	.961	.577	.929	.386	.926	.317	.956	.522	.945	.363	.930	.290
Joint	β_{02}	.931	.737	.945	.520	.948	.431	.952	.681	.929	.470	.944	.386
	β_{03}	.945	.664	.942	.470	.934	.386	.949	.609	.948	.424	.951	.347
	α^*	.950	.514	.920	.372	.934	.307	.952	.484	.930	.344	.944	.279
K-S	β_{02}	.988	.864	.985	.590	.987	.476	.986	.713	.970	.484	.983	.402
	β_{03}	.983	.804	.980	.548	.983	.448	.994	.648	.979	.459	.979	.380
	α^*	.952	.748	.924	.463	.910	.358	.942	.613	.905	.384	.884	.295
SMS	β_{02}	.965	1.317	.977	.849	.976	.709	.980	1.102	.976	.736	.975	.625
	β_{03}	.973	1.182	.975	.755	.980	.632	.985	.990	.975	.666	.975	.566
	α^*	.978	1.281	.971	.869	.980	.716	.984	1.067	.975	.744	.970	.648
Rank	β_{02}	.831	.512	.907	.431	.936	.373	.843	.452	.915	.374	.946	.323
	β_{03}	.869	.499	.935	.410	.939	.343	.884	.434	.929	.348	.947	.297
	α^*	.939	.565	.939	.423	.950	.355	.935	.512	.952	.387	.956	.316

On the basis of our simulation results, we recommend the NPMLE-based estimator and the isotonic two-stage estimator in combination with the nonparametric bootstrap inference. The main convenience of the recommended estimators is that applied researchers do not need to choose any tuning parameter. Moreover, our simulations show that NPMLE-based joint estimator is more efficient (smaller RMSE and shorter CI median length) than the K-S and SMS approaches. The rank estimator performs quite well when the sample size is relatively large, but is outperformed by our estimators when the sample size is relatively small.

5.2. An Empirical Application Using Joint Retirement Data

We utilize the interdependent durations model to study the joint retirement decisions of Danish couples (An, Christensen, and Gupta (2004)) so that T_1 and T_2 represent the timing choices of retirement dates for the wives and husbands, respectively. The unique feature in this type of data is the couples' synchronization of retirement dates (so $T_1 = T_2$), despite the difference in their observable characteristics such as age, education level, or working skill. This posts new challenges to the econometric analysis, because the probability of simultaneous retirement is assumed to be zero in the traditional (mixed) proportional hazards models in duration analysis. In sharp contrast, the Honoré-de Paula model accommodates the presence of simultaneous retirement, and this is generated by strategic interaction of both players in the noncooperative game. In the literature, noncooperative games have been widely used to model joint retirement behavior of couples in Gustman and Steinmeier (2000) and Gustman and Steinmeier (2004), among others. Also see Lundberg and Pollak (1994) for a general discussion of noncooperative games for within-marriage decisions.

The data of elderly couples are drawn from a random 0.5% sample of the Danish population and are provided by An, Christensen, and Gupta (2004). We restrict the sample to be couples satisfying the following conditions: (1) Both husband and wife were working in the base year (defined as the year in which the oldest spouse is 54 in November); (2) At least one spouse retired by the end of the survey, which allows us to determine the time order of retirement for each couple; (3) Couples remained married until either the death of the spouse or the end of the survey; (4) The base year was 1990 or earlier. The resulting sample consists of 146 couples. For 29% of the couples in our sample, wives retired at least one year earlier than husbands; 35% of couples retired in the same year; for the remaining 36% couples, wives retiring later than husbands. As explained in Section 2, the interdependent durations model can be estimated via an ordered response model. We consider two covariates: age and a skill dummy variable indicating the job category. Table 5 presents the summary statistics of the covariates for husbands and wives. Note that in the order response model, we use the covariate of the wife minus the covariate of the husband. The coefficient on age difference is normalized to one.

Table 5. Descriptive statistics.

Variables	Mean	Std.
Age-wife	50.10	3.82
Age-husband	53.41	2.05
Skill-wife	0.60	0.49
Skill-husband	0.68	0.47

Table 6 presents the estimates of the coefficient β_{02} on the job category difference and the interaction parameter α^*. The minus sign of the point estimate of β_{02} suggests that if a wife was categorized as a skilled employee while her husband was not, then the probability that she retired earlier was smaller, which can be driven by the family-wise income effect. Our two estimators and the rank estimator yield similar point estimates for β_{02}. Nevertheless, such an effect of job skill difference on the retirement timing turns out to be insignificant at the 5% significance level except for the SMS estimator with an under-smoothed bandwidth ($n^{-1/4}$) and the rank estimator. On the other hand, the interaction parameter α^* is significantly positive in all approaches except for K-S estimator, suggesting that a spouse gained extra utility from life after retirement together with her/his counterpart. In particular, the Isotonic two-stage estimator indicates that couples to synchronize their retirement timing with a force large enough to compensate on average 4.76 years of age difference in the SSE estimate. That number rises to 6.17 years for the NPMLE-based joint estimator. Other approaches yield comparable point estimates. In terms of the 95% confidence intervals, our two-stage and joint estimators yield results close to the rank estimator. The vast majority of values in their confidence intervals are large enough to counter the average age difference 3.31 of the couples[18]. This evidence is also consistent with previous empirical results based on various structural models (Gustman and Steinmeier (2000, 2004)) where the essence is to argue the interaction effect is significant enough to counter the age difference when couples retire simultaneously. It is notable that SMS estimator produces substantially larger upper bound for α^* and smaller lower bound for β_{02}. This is not surprising given the over-coverage reported in our Monte-Carlo results. As a matter of fact, in a different empirical application, Bellemare, Melenberg, and Van Soest (2002) also found that the SMS estimates are far away from other methods including the ordered Probit, the partial linear, and the semi-parametric least square, among others[19].

[18] Recall that the coefficient on the age difference is normalized to one.

[19] See p194-195 of Bellemare, Melenberg, and Van Soest (2002) for details.

Table 6. Estimates of an interdependent duration model applied to joint retirement in Denmark, $n = 146$, β_{01} (coef. on age difference) is normalized to 1. The 95% confidence intervals are constructed from 200 bootstrap samples.

		β_{02} (job category diff.)	α^* (interaction)
Two-stage	Coef.	-2.35	4.76
	95% CI	[-12.69, 2.32]	[3.06, 14.31]
Joint	Coef.	-3.21	6.17
	95% CI	[-9.54, 3.32]	[2.90, 15.96]
K-S	Coef.	-0.14	5.76
($p = 0.05$)	95% CI	[-12.24, 4.02]	[-1.71, 9.55]
K-S	Coef.	-0.14	7.60
($p = 0.20$)	95% CI	[-12.24, 4.02]	[-0.57, 9.34]
SMS	Coef.	-5.20	4.70
($h = n^{-1/5}$)	95% CI	[-23.90, 0.01]	[4.16, 23.60]
SMS	Coef.	-5.22	4.83
($h = n^{-1/4}$)	95% CI	[-24.14, -0.32]	[3.99, 23.94]
Rank	Coef.	-2.99	8.20
	95% CI	[-10.56, -0.40]	[3.15, 14.40]

Figure 3. The estimated CDF of $\log(\epsilon_1/\epsilon_2)$: the distribution of the log ratio of wife-husband unobservables.

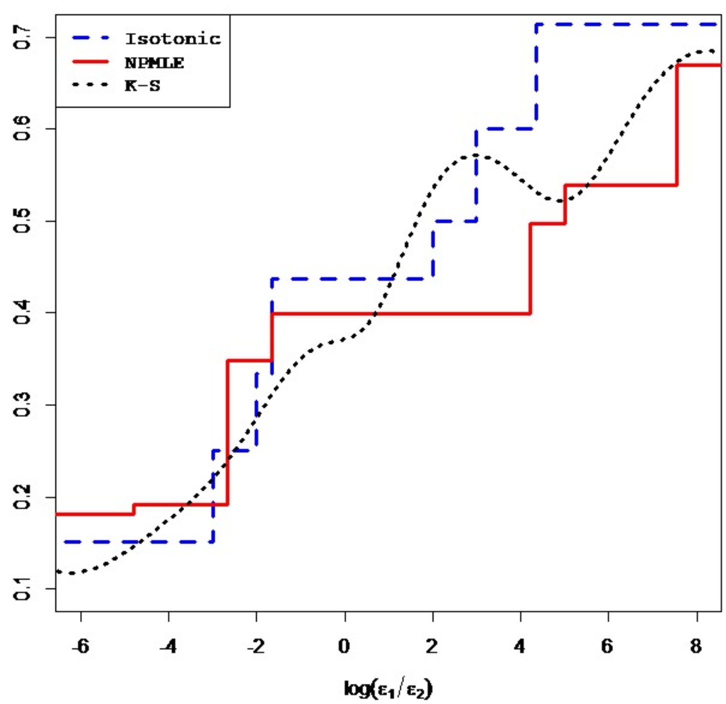


Figure 3 plots the estimated CDF of the error term in the ordered response model, which corresponds to the distribution of the log ratio of unobservables for wives and husbands, $H(\cdot)$. By construction, both our isotonic and NPMLE estimates for the CDF are non-decreasing step functions. In contrast, the K-S approach yields an estimated CDF that fluctuates and has some local maximum and minimum points.

6. Conclusion and Extension

In this paper, we have proposed two simple semiparametric estimation methods for ordered response models with an unknown latent error distribution. We rigorously establish the asymptotic properties of finite dimensional parameters, tackling the challenging issues related to the nonparametric components based on NPMLE. Our methods are easy to implement and free of any tuning parameter. Also, the methodology is directly applicable to estimate the social interaction effect in the interdependent durations model by Honoré and de Paula (2010). Complementing several important contributions as in Lee (1992), Klein and Sherman (2002), Lewbel (2002), and Coppejans (2007), our work makes the Honoré-de Paula model a viable benchmark in analyzing multiple durations data with strategically-interacting agents. Both the Monte Carlo simulation and a real data application demonstrate the utility of our approach.

In conclusion, we briefly draw the reader's attention to several interesting questions that can provide future research problems in this area. First of all, a rigorous investigation of an efficient procedure proposed in Section 4.2 need to be delegated to another paper, considering the amount of details required. A second extension involves modeling the joint retirement phenomenon by cooperative games as considered by Honoré and de Paula (2018). The Nash bargaining solution in Honoré and de Paula (2018) can not be written as the standard ordered response model, unlike the non-cooperative game model in Honoré and de Paula (2010). It remains a challenge to develop a flexible semiparametric estimation procedure for the model in Honoré and de Paula (2018). Third, one could apply our estimation methods in estimating a sample selection model where the selection equation involves making ordered choices; see Section VI.A on [p.147] of Vella (1998). An interesting example is presented by Attanasio, Koujianou, and Kyriazidou (2008) in which they study credit constraints faced by consumers in the market for durable goods. Last but note least, it is worthwhile to study the ordered response model with panel data setting (Muris (2017), Abrevaya and Muris (2018)). A thorough treatment of all these extensions is outside the scope of this paper and will be pursued in the future.

7. Appendix A: Proofs of Main Results

This appendix provides proofs for the theorems. The lemmas used for proofs (denoted as Lemma S1's and S2's) are collected in the supplemental note.

We denote some positive constants by c or C whose value might change line by line.

Proof of Theorem 4.1. Part (i) has been shown in Theorem 4.1 [p1426] of Groeneboom and Hendrickx (2018). Here we proof part (ii). Note that Remark 3.1 [p1423] of Groeneboom and Hendrickx (2018) showed that $\hat{F}_n(u;\hat{\beta}_n)$ converges uniformly to $F_0(u)$. We start with the following estimating equation in Stage 2:

$$\Psi_n(\alpha) = \frac{1}{n}\sum_{i=1}^{n}\left[1-\Delta_{3i}-\hat{F}_n(X_i'\hat{\beta}_n+\alpha;\hat{\beta}_n)\right], \tag{7.1}$$

and its probability limit

$$\Psi(\alpha) = \mathbb{E}\left[1-\Delta_3-F_0(X'\beta_0+\alpha)\right]. \tag{7.2}$$

We first show the following uniform convergence result

$$\sup_{\alpha}|\Psi_n(\alpha)-\Psi(\alpha)| \to_p 0. \tag{7.3}$$

Observe that

$$\begin{aligned}\sup_{\alpha}|\Psi_n(\alpha)-\Psi(\alpha)| \le & \sup_{\alpha}|(\mathbb{P}_n-P)\left[\Delta_3+\hat{F}_n(X'\hat{\beta}_n+\alpha;\hat{\beta}_n)\right]| \\ & +\sup_{\alpha}|P\left[\hat{F}_n(X'\hat{\beta}_n+\alpha;\hat{\beta}_n)-F_0(X'\beta_0+\alpha)\right]|.\end{aligned} \tag{7.4}$$

Hence, the uniform convergence (7.3) hold by Glivenko-Cantelli property in Lemma (S1.4) and the fact that $\hat{F}_n(u;\hat{\beta}_n)$ converges to $F_0(u)$ in sup-norm.
Next we show the existence of a unique zero-crossing point 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:

$$(\hat{\alpha}_n-\alpha)\Psi_n(\alpha) \ge 0, \tag{7.5}$$

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

$$(\alpha_1-\alpha_2)\Psi_n(\alpha_2) \le -c < 0, \tag{7.6}$$

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

$$(\alpha_0 - \alpha_2)\Psi_n(\alpha_2) \leq -c, \tag{7.7}$$

By (7.3), we get

$$(\alpha_0 - \alpha_2)\Psi(\alpha_2) \leq -c/2, \tag{7.8}$$

with a 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 w.r.t. α, given the monotonicity and absolute continuity of F_0. Thus, the zero-crossing point $\hat{\alpha}_n$ exists with a probability tending to 1.

Finally considering the monotone estimating equation $\Psi_n(\alpha)$ given the monotonic NPMLE $\hat{F}_n$, the consistency result is a direct consequence of Lemma 5.10 in Van Der Vaart (1998). □

Proof of Theorem 4.2. The linear representation of $\hat{\beta}_n$ and its asymptotic normality have been shown in Theorem 4.1 of Groeneboom and Hendrickx (2018). Here we focus on $\hat{\alpha}_n$. Following Groeneboom and Hendrickx (2018), we can define the value of Ψ_n at $\hat{\alpha}_n$ by setting $\Psi_n(\hat{\alpha}_n) = 0$. Note that $\Psi_n(\hat{\alpha}_n)$ is the convex combination of the left and right limit at $\hat{\alpha}_n$:

$$\Psi_n(\hat{\alpha}_n) = \lambda\Psi_n(\hat{\alpha}_n-) + (1-\lambda)\Psi_n(\hat{\alpha}_n+) = 0, \tag{7.9}$$

where we can choose $\lambda \in [0, 1]$ such that (7.9) holds. To prove the root-n consistency and asymptotic normality of $\hat{\alpha}_n$, we start with the estimating equation $\Psi_n(\hat{\alpha}_n) = 0$ and we decompose the l.h.s. as

$$\frac{1}{n}\sum_{i=1}^{n}\left[1 - \hat{F}_n(X_i'\hat{\beta}_n + \hat{\alpha}_n; \hat{\beta}_n) - \Delta_{3i}\right] = I_{1n} + I_{2n} + I_{3n}, \tag{7.10}$$

where

$$I_{1n} = \frac{1}{n}\sum_{i=1}^{n}\left[1 - F_0(X_i'\beta_0 + \alpha_0) - \Delta_{3i}\right], \tag{7.11}$$

$$I_{2n} = \frac{1}{n}\sum_{i=1}^{n}\left[F_0(X_i'\beta_0 + \alpha_0) - \hat{F}_n(X_i'\beta_0 + \alpha_0; \beta_0)\right], \tag{7.12}$$

$$I_{3n} = \frac{1}{n}\sum_{i=1}^{n}\left[\hat{F}_n(X_i'\beta_0 + \alpha_0; \beta_0) - \hat{F}_n(X_i'\hat{\beta}_n + \hat{\alpha}_n; \hat{\beta}_n)\right]. \tag{7.13}$$

Here $\hat{F}_n(X_i'\beta_0 + \alpha_0; \beta_0)$ is the (infeasible) NPMLE computed using the true unknown β_0. Apparently, the term I_{1n} is of $O_p(n^{-1/2})$ with its influence function equal to ψ_0 as defined in our Theorem 4.2.

Referring to I_{2n}, we get $I_{2n} = I_{2n}^a + I_{2n}^b$ where

$$I_{2n}^a = P\left[F_0(U+\alpha_0) - \hat{F}_n(U+\alpha_0;\beta_0)\right] \quad \text{and} \tag{7.14}$$

$$I_{2n}^b = (\mathbb{P}_n - P)\left[F_0(U+\alpha_0) - \hat{F}_n(U+\alpha_0;\beta_0)\right].$$

We shall utilize P-Donsker property (Van Der Vaart and Wellner (1996)) to show $I_{2n}^b = o_p(n^{-1/2})$ in Lemma (S1.4), and we obtain the linear representation for I_{2n}^a as follows

$$\sqrt{n} I_{2n}^a = \mathbb{G}_n \psi_{F_0} + o_p(1), \tag{7.15}$$

in our Lemma (S1.9). When it comes to I_{3n}, we decompose it into three terms $I_{3n} = I_{3n}^a + I_{3n}^b + I_{3n}^c$ where

(7.16)

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

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

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

In Lemma S1.11 and Lemma S1.4 respectively, we prove that $I_{3n}^b = o_p(n^{-1/2})$, $I_{3n}^c = o_p(n^{-1/2})$ using P-Donsker Property of related functional classes. Also we have the following expansion:

$$I_{3n}^a = V_{\alpha_0}(\hat{\alpha}_n - \alpha_0) + V_{\beta_0}(\hat{\beta}_n - \beta_0) + o_p(n^{-1/2} + \hat{\alpha}_n - \alpha_0 + |\hat{\beta}_n - \beta_0|). \tag{7.17}$$

In sum, the desired linear representation for $\hat{\alpha}_n$ follows after collecting the leading terms in I_{1n}, I_{2n}^a and I_{3n}^a and substituting the linear representation for $\hat{\beta}_n$ based on (4.5). □

Proof of Theorem 4.3. Given the compactness of the parameter space, the sequence of estimator $\tilde{\theta}_n$ has a subsequence $\tilde{\theta}_{n_k}$ converging to some element $\theta^* = (\alpha^*, \beta^{*\prime})'$. In Appendix C of the supplemental note, we apply Theorem 7.4 in Van de Geer (2000) to show the following convergence in terms of the Hellinger distance (Van de Geer (1993)):

$$\sup_{\theta} \mathbf{h}(\hat{p}_{n,\theta}, p_{0,\theta}) = O_p(n^{-1/3}\log^2 n),$$

where the underlying density function is

$$p_{0,\theta} \equiv F_0^{\Delta_{1i}}(X_i'\beta;\theta) \times \left(F_0(X_i'\beta + \alpha;\theta) - F_0(X_i'\beta;\theta)\right)^{\Delta_{2i}} \times \left(1 - F_0(X_i'\beta + \alpha;\theta)\right)^{\Delta_{3i}},$$

and $\hat{p}_{n,\theta}$ is the corresponding maximum likelihood estimator given θ.
Since the Hellinger distance (Van de Geer (1993)) is equal to

$$\begin{aligned} 2\mathbf{h}(\hat{p}_{n,\theta}, p_{0,\theta}) = & \int \left(\tilde{F}_n^{1/2}(u;\theta) - F_0^{1/2}(u;\theta)\right)^2 dQ \\ & + \int \left((\tilde{F}_n(u+\alpha;\theta) - \tilde{F}_n(u;\theta))^{1/2} - (F_0(u+\alpha;\theta) - F_0(u;\theta))^{1/2}\right)^2 dQ \\ & + \int \left((1-\tilde{F}_n(u+\alpha;\theta))^{1/2} - (1-F_0(u+\alpha;\theta))^{1/2}\right)^2 dQ, \end{aligned}$$

it is obvious that

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

where the L_2 norm is defined as $\parallel \tilde{F}_n(u;\theta) - F_0(u;\theta) \parallel_2^2 \equiv \int \left(\tilde{F}_n(u;\theta) - F_0(u;\theta)\right)^2 g_0(u)du$.
As a consequence of the uniform convergence for NPMLE, we have

$$\tilde{F}_{n_k}(\tilde{\alpha}_{n_k} + x'\tilde{\beta}_{n_k}; \tilde{\alpha}_{n_k}, \tilde{\beta}_{n_k}) \to F_0(\alpha^* + x'\beta^*; \alpha^*, \beta^*).$$

Thereafter, the following uniform convergence is immediate.

$$|\Phi_{n_k}(\tilde{\theta}_{n_k}) - \Phi(\theta^*)| \to_p 0. \tag{7.18}$$

Given the true unknown θ_0 is the unique root of the continuous limiting function $\Phi(\cdot)$, we must have $\theta^* = \theta_0$, which leads to the consistency of $\tilde{\theta}_n$.

The proof of $\tilde{\theta}_n$'s existence is more tedious (see Theorem 4.1 in Groeneboom and Hendrickx (2018)) and is relegated to Appendix C in the supplemental note. □

Proof of Theorem 4.4. Denote the stacked moment conditions by

$$\zeta(Z_i;\alpha,\beta,F(;\alpha,\beta)) = \begin{pmatrix} [\Delta_{1i} - F(X_i'\beta;\alpha,\beta)]X_i \\ \Delta_{3i} - 1 + F(\alpha + X_i'\beta;\alpha,\beta) \end{pmatrix}$$

We set the value of Φ_n to be zero at the point of zero-crossing $\tilde{\theta}_n = (\tilde{\alpha}_n, \tilde{\beta}_n)$ as

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

With this definition, we use the representation of the components as a convex combination of the left and right limit at $\tilde{\theta}_n$:

$$\Phi_{n,j}(\tilde{\theta}_n) = \lambda_j\Phi_{n,j}(\tilde{\theta}_n-) + (1-\lambda_j)\Phi_{n,j}(\tilde{\theta}_n+) = 0, \tag{7.19}$$

where $\Phi_{n,j}$ denotes the jth coordinate of Φ_n and where we can choose λ_j from the unit interval in such a way that (7.19) holds since we have a crossing of zero component-wise. Note that this does not change the location of zero-crossing point either.

Then we proceed with

$$\begin{aligned}0 =& (\mathbb{P}_n - P)\zeta(Z; \tilde{\alpha}_n, \tilde{\beta}_n, \tilde{F}_n(\cdot; \tilde{\alpha}_n, \tilde{\beta}_n)) \\ &+P[\zeta(Z; \alpha_0, \beta_0, \tilde{F}_n(\cdot; \alpha_0, \beta_0)) - \zeta(Z; \alpha_0, \beta_0, F_0(\cdot))] \\ &+P[\zeta(Z; \tilde{\alpha}_n, \tilde{\beta}_n, \tilde{F}_n(\cdot; \tilde{\alpha}_n, \tilde{\beta}_n)) - \zeta(Z; \alpha_0, \beta_0, \tilde{F}_n(\cdot; \alpha_0, \beta_0))]\end{aligned} \tag{7.20}$$

In the technical lemma (see Lemma S2.3) of our Appendix C, we prove that by the consistency of our estimator and the P-Donsker property of the corresponding functional class, one has

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

Also, by applying the P-Donsker property and taking a Taylor expansion, one has

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

When it comes to the second term in (7.20), we integrate by parts to get

$$P[\zeta(Z; \alpha_0, \beta_0, \tilde{F}_n(\cdot; \alpha_0, \beta_0)) - \zeta(Z; \alpha_0, \beta_0, F_0(\cdot))] = \int c(u)d[(\tilde{F}_n - F_0)(u)] = \kappa(\tilde{F}_n) - \kappa(F_0),$$

which reduces the problem to characterizing the asymptotic property of the linear functional for the NPMLE. Thus, we proceed in Appendix C to show that

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

In sum, we obtain

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

Hence, the desired conclusion follows given the short-hand notation $\phi_0 \equiv \zeta(Z; \alpha_0, \beta_0, F_0)$. □

Proof of Theorem 4.5. 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 we decompose it into

$$\frac{1}{n}\sum_{i=1}^{n} M_{ni}\left[1 - \hat{F}_n^*(X_i'\hat{\beta}_n^* + \hat{\alpha}_n^*; \hat{\beta}_n^*) - \Delta_{3i}\right] = I_{1n}^* + I_{2n}^* + I_{3n}^*, \tag{7.21}$$

where

$$I^*_{1n} = \frac{1}{n}\sum_{i=1}^{n} M_{ni}\left[1 - F_0(X_i'\beta_0 + \alpha_0) - \Delta_{3i}\right], \tag{7.22}$$

$$I^*_{2n} = \frac{1}{n}\sum_{i=1}^{n} M_{ni}\left[F_0(X_i'\beta_0 + \alpha_0) - \hat{F}^*_n(X_i'\beta_0 + \alpha_0; \beta_0)\right], \tag{7.23}$$

$$I^*_{3n} = \frac{1}{n}\sum_{i=1}^{n} M_{ni}\left[\hat{F}^*_n(X_i'\beta_0 + \alpha_0; \beta_0) - \hat{F}^*_n(X_i'\hat{\beta}^*_n + \hat{\alpha}^*_n; \hat{\beta}^*_n)\right]. \tag{7.24}$$

The general scheme is in analog with our proof of Theorem 4.2. First of all, note that $I^*_{1n} = O_{P_M}(n^{-1/2})$ in P_Z-probability.
Referring to I^*_{2n}, we get $I_{2n} = I^{*a}_{2n} + I^{*b}_{2n}$ where
(7.25)
$$I^{*a}_{2n} = P\left[F_0(U + \alpha_0) - \hat{F}^*_n(U + \alpha_0; \beta_0)\right] \quad \text{and} \quad I^{*b}_{2n} = (\mathbb{P}^*_n - P)\left[F_0(U + \alpha_0) - \hat{F}^*_n(U + \alpha_0; \beta_0)\right].$$

We shall utilize P-Donsker property (Van Der Vaart and Wellner (1996)) to show $I^b_{2n} = o_p(n^{-1/2})$ in Lemma S1.4 in the supplemental note. We also state the linear representation of I^{*a}_{2n} in Lemma S1.9.
When it comes to I^*_{3n}, we decompose it into three terms $I^*_{3n} = I^{*a}_{3n} + I^{*b}_{3n} + I^{*c}_{3n}$ where

(7.26)
$$\begin{aligned}
I^{*a}_{3n} &= P\left[F_0(X'\hat{\beta}^*_n + \hat{\alpha}^*_n; \hat{\beta}^*_n) - F_0(X'\beta_0 + \alpha_0)\right], \\
I^{*b}_{3n} &= P\left[\hat{F}^*_n(X'\hat{\beta}^*_n + \hat{\alpha}^*_n; \hat{\beta}^*_n) - \hat{F}^*_n(X'\beta_0 + \alpha_0; \beta_0) - F_0(X'\hat{\beta}^*_n + \hat{\alpha}^*_n; \hat{\beta}^*_n) + F_0(X'\beta_0 + \alpha_0)\right], \\
I^{*c}_{3n} &= (\mathbb{P}^*_n - P)\left[\hat{F}^*_n(X'\hat{\beta}^*_n + \hat{\alpha}^*_n; \hat{\beta}^*_n) - \hat{F}^*_n(X'\beta_0 + \alpha_0; \beta_0)\right].
\end{aligned}$$

In Lemma S1.11 and Lemma S1.4 respectively, we prove that $I^{*b}_{3n} = o_p(n^{-1/2})$, $I^{*b}_{3n} = o_p(n^{-1/2})$ using P-Donsker Property of related functional classes. Also we have the following expansion:

$$I^{*a}_{3n} = V_{\alpha_0}(\hat{\alpha}^*_n - \alpha_0) + V_{\beta_0}(\hat{\beta}^*_n - \beta_0) + o_p(n^{-1/2} + \hat{\alpha}^*_n - \alpha_0 + |\hat{\beta}^*_n - \beta_0|). \tag{7.27}$$

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

$$\sqrt{n}\,(\hat{\alpha}^*_n - \hat{\alpha}_n) = V^{-1}_{\alpha_0}\mathbb{G}^*_n\left[\psi_0 + \psi_{F_0} + \psi_{\beta_0}\right] + o_p(1), \tag{7.28}$$

and the claim follows from Theorem 3.6.13 in Van Der Vaart and Wellner (1996). □

Proof of Theorem 4.6. The overall structure of the proof is similar as the one for Theorem 4.4. The only necessary change is that one has to apply the maximal inequality with multiplier bootstrap weights to relevant functional classes. To avoid repetition, we only sketch 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

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

Then one can show that

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

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

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

□

References

Abrevaya, J., and J. Huang (2005): "On the bootstrap of the maximum score estimator," *Econometrica*, 73(4), 1175–1204.

Abrevaya, J., and C. Muris (2018): "Interval censored regression with fixed effects," *working paper*.

Aitchison, J., and S. Silvey (1957): "The generalization of probit analysis to the case of multiple responses," *Biometrika*, 44, 131–140.

An, M. Y., B. J. Christensen, and N. D. Gupta (2004): "Multivariate mixed proportional hazard modelling of the joint retirement of married couples," *Journal of Applied Econometrics*, 19(6), 687–704.

Attanasio, O. P., G. P. Koujianou, and E. Kyriazidou (2008): "Credit constraints in the market for consumer durables: Evidence from micro data on car loans," *International Economic Review*, 49(2), 401–436.

Ayer, M., H. Brunk, G. Ewing, W. Reid, and E. Silverman (1955): "An empirical distribution function for sampling with incomplete information," *Annals of Mathematical Statistics*, 26, 641–647.

Banerjee, M., D. Mukherjee, and S. Mishra (2009): "Semiparametric binary regression models under shape constraints with an application to Indian schooling data," *Journal of Econometrics*, 149(2), 101–117.

Bellemare, C., B. Melenberg, and A. Van Soest (2002): "Semi-parametric models for satisfaction with income," *Portuguese Economic Journal*, 1, 181–203.

Bresnahan, T. F., and P. C. Reiss (1991): "Entry and competition in concentrated markets," *Journal of Political Economy*, 99, 977–1009.

Cameron, S. V., and J. J. Heckman (1998): "Life cyce 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.

Chen, S., and S. Khan (2003): "Rates of convergence for estimating regression coefficients in heteroskedastic discrete response models," *Journal of Econometrics*, 117, 245–278.

Chen, X., O. Linton, and I. Van Keilegom (2003): "Estimation of semiparametric models when the criterion function is not smooth," *Econometrica*, 71, 1591–1608.

Cheng, G., and J. Huang (2010): "Bootstrap consistency for general semiparametric M-estimate," *Annals of Statistics*, 38, 2884–2915.

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

Cosslett, S. R. (1983): "Distribution-free maximum likelihood estimator of the binary choice model," *Econometrica*, 51, 765–782.

Cunha, F., J. Heckman, and S. Navarro (2007): "The identification and economic content of ordered choice models with stochastic thresholds," *International Economic Review*, 48, 1273–1309.

de Paula, Á. (2009): "Inference in a synchronization game with social interactions," *Journal of Econometrics*, 148, 56–71.

Durlauf, S. N., and Y. M. Ioannides (2010): "Social interactions," *Annual Review of Economics*, 2, 451–478.

Efron, B. (1979): "Bootstrap methods: Another look at the jackknife," *The Annals of statistics*, 7, 1–26.

Gentleman, R., and A. Vandal (2018): "Icens: NPMLE for censored and truncated data," *R package version 1.53.0.*

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.

Greene, W. H., and D. A. Hensher (2010): *Modeling ordered choices: A primer.* Cambridge University Press.

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.

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.

Gustman, A. L., and T. L. Steinmeier (2000): "Retirement in dual-career families: a structural model," *Journal of Labor economics*, 18, 503–545.

——— (2004): "Social security, pensions and retirement behaviour within the family," *Journal of Applied Econometrics*, 19, 723–737.

Hajivassiliou, V. A., and Y. M. Ioannides (2007): "Unemployment and liquidity constraints," *Journal of Applied Econometrics*, 22, 479–510.

Hong, H., J. D. Kubik, and J. C. Stein (2004): "Social interaction and stock-market participation," *The Journal of Finance*, 59(1), 137–163.

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

——— (2018): "A new model for interdependent durations with an application to joint retirement," *Quantitative Economics*, 9, 1299–1333.

Horowitz, J. L. (1992): "A smoothed maximum score estimator for the binary response model," *Econometrica*, pp. 505–531.

——— (1996): "Semiparametric Estimation of a Regression Model with an Unknown Transformation of the Dependent Variable," *Econometrica*, 64, 103–137.

——— (2009): *Semiparametric and nonparametric methods in econometrics*. Springer.

Huang, J., and J. A. Wellner (1995): "Asymptotic normality of the NPMLE of linear functionals for interval censored data, Case 1," *Statistica Neerlandica*, pp. 153–163.

Ichimura, H. (1993): "Semiparametric least squares (SLS) and weighted SLS estimation of single-index models," *Journal of Econometrics*, 58.

Katz, M. L., and C. Shapiro (1986): "Technology adoption in the presence of network externalities," *Journal of Political Economy*, 94(4), 822–841.

Kiefer, J., and J. Wolfowitz (1956): "Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters," *The Annals of Mathematical Statistics*, 27(4), 887–906.

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.

Lee, M.-j. (1992): "Median regression for ordered discrete response," *Journal of Econometrics*, 51, 59–77.

Lewbel, A. (1997): "Semiparametric estimation of location and other discrete choice moments," *Econometric Theory*, 13, 32–51.

——— (2000): "Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables," *Journal of Econometrics*, 97, 145–177.

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

Lewbel, A., and S. M. Schennach (2007): "A simple ordered data estimator for inverse density weighted expectations," *Journal of Econometrics*, 136(1), 189–211.

Lundberg, S., and R. A. Pollak (1994): "Noncooperative bargaining models of marriage," *The American Economic Review*, 84, 132–137.

Manski, C. F. (1985): "Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator," *Journal of Econometrics*, 27, 313–333.

Manski, C. F., and E. Tamer (2002): "Inference on regressions with interval data on a regressor or outcome," *Econometrica*, 70(2), 519–546.

Matzkin, R. L. (1991): "Semiparametric estimation of monotone and concave utility functions for polychotomous choice models," *Econometrica*, 59, 1315–1327.

——— (1993): "Nonparametric identification and estimation of polychotomous choice models," *Journal of Econometrics*, 58, 137–168.

Melenberg, B., and A. Van Soest (1996): "Measuring the costs of children: parametric and semiparametric estimators 1," *Statistica Neerlandica*, 50, 171–192.

Muris, C. (2017): "Estimation in the fixed effects ordered Logit model," *Review of Economics and Statistics*, 99, 465–477.

Nan, B., and J. A. Wellner (2013): "A general semiparametric Z-estimation approach for case-cohort studies," *Statistica Sinica*, 23(3), 1155.

Newey, W. K. (1994): "The asymptotic variance of semiparametric estimators," *Econometrica*, 62, 1349–1382.

Park, A., and L. Smith (2008): "Caller number five and related timing games," *Theoretical Economics*, 3, 231–256.

Prescott, E., and M. Visscher (1977): "Sequential location among firms with foresight," *Bell Journal of Economics*, 8, 378–893.

Ridder, G. (1990): "The non-parametric identification of generalized accelerated failure-time models," *Review of Economic Studies*, 57, 167–181.

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.

Shaked, A., and J. Sutton (1982): "Relaxing price competition through product differentiation," *Review of Economic Studies*, 49, 3–13.

Stewart, M. B. (2005): "A comparison of semiparametric estimators for the ordered response model," *Computational statistics & data analysis*, 49, 555–573.

Tanaka, H. (2008): "Semiparametric least squares estimation of monotone single index models and its application to the iterative least squares estimation of binary choice models," *Working Paper*.

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

——— (1997): "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.

Vella, F. (1998): "Estimating models with sample selection bias: a survey," *Journal of Human Resources*, 33, 127–169.

Wellner, J. A., and Y. Zhan (1997): "A hybrid algorithm for computation of the non-parametric maximum likelihood estimator from censored data," *Journal of the American Statistical Association*, 92, 945–959.

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

Supplemental Note for "Simple Semiparametric Estimation of Ordered Response Models: With an Application to the Interdependent Durations Model"

Ruixuan Liu and Zhengfei Yu
Emory University and University of Tsukuba

S1. Appendix B: Technical Proofs Related to Two-stage Estimation

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 $|f(x)| \leq F(x)$ for all x and $f \in \mathcal{F}$.

Lemma S1.1 (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

$$\mathbb{E} \sup_{f \in \mathcal{F}} |\mathbb{G}_n f| \lesssim J(1, \mathcal{F}) \parallel F \parallel_{P_0,2} \tag{S1.1}$$

Lemma S1.2. [Lemma 3.6.7 in (Van Der Vaart and Wellner (1996))] Let $Z_{n1}, ..., Z_{nn}$ be arbitrary stochastic processes and $(M_{n1}, ..., M_{nn})'$ be any exchangeable random vector independent of $Z_{n1}, ..., Z_{nn}$. For any $n_0 > 0$ and $n > n_0$, we have

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

We need to apply the following well-known entropy bounds concerning monotone functions or functions of bounded variation repeatedly. The bounds actually hold for the bracketing entropy uniformly over the underlying probability measure, which will be used in Appendix C 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 S1.3 (Entropy Bounds)**.** Let $\mathcal{A}_C$ be the class of monotone functions with values in $[0, C]$, then for all $\delta > 0$,

$$J(\delta, \mathcal{A}_C) \lesssim \sqrt{\delta}. \tag{S1.2}$$

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

$$J(\delta, \mathcal{B}_C) \lesssim \sqrt{\delta}. \tag{S1.3}$$

Now we obtain the entropy bounds for the key functional class in our context and prove the asymptotic characterizations for terms appearing in our Theorem 4.1 and Theorem 4.2.

Lemma S1.4. The functional class $\mathcal{G}$ defined by

$$\mathcal{G} \equiv \left\{(x, \delta_3) \mapsto (1 - \delta_3 - F(x'\beta + \alpha)) : (\alpha, \beta) \in \Theta, F(\cdot) \in \mathcal{A}\right\} \tag{S1.4}$$

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

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

$$(\mathbb{P}_n^* - P)\left[\hat{F}_n\left(X'\hat{\beta}_n^* + \hat{\alpha}_n^*; \hat{\beta}_n^*\right)\right] = o_{P_M}(1),$$

in P_Z-probability. Moreover, we obtain the stochastic equicontinuity as

$$\mathbb{G}_n\left[\hat{F}_n\left(X'\hat{\beta}_n + \hat{\alpha}_n; \hat{\beta}_n\right) - F_0\left(X'\beta_0 + \alpha_0\right)\right] = o_p(1), \quad \text{and} \tag{S1.5}$$

$$\mathbb{G}_n^*\left[\hat{F}_n\left(X'\hat{\beta}_n^* + \hat{\alpha}_n^*; \hat{\beta}_n^*\right) - F_0\left(X'\beta_0 + \alpha_0\right)\right] = o_{P_M}(1), \tag{S1.6}$$

in P_Z-probability.

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 an monotonically increasing link/ridge function and a linear index $x'\beta+\alpha$ with parameters (α,β) belonging to a compact Euclidean space. Hence, by Lemma 2.3 in Baladbaoui, Groeneboom, and Hendrickx (2017) we get the following bound on the uniform entropy $\log N(\varepsilon,\mathcal{G}) \lesssim 1/\varepsilon$, 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 properties.

Regarding the stochastic equicontinuity, consider the following class:

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

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

$$\begin{aligned}(\text{S1.7})\quad & \| \hat{F}_n\left(X'\hat{\beta}_n+\hat{\alpha}_n;\hat{\beta}_n\right) - F_0\left(X'\beta_0+\alpha_0\right) \|_\infty \\ &\leq \| \hat{F}_n\left(X'\hat{\beta}_n+\hat{\alpha}_n;\hat{\beta}_n\right) - F_0\left(X'\hat{\beta}_n+\hat{\alpha}_n;\hat{\beta}_n\right) \|_\infty + \| F_0\left(X'\hat{\beta}_n+\hat{\alpha}_n;\hat{\beta}_n\right) - F_0\left(X'\beta_0+\alpha_0\right) \|_\infty \\ &\to 0.\end{aligned}$$

The first term on the right hand side of the inequality follows from the uniform consistency of NPMLE as in (S1.16), 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 consistency of $\hat{\alpha}_n$ and $\hat{\beta}_n$. Thereafter, the desired stochastic equicontinuity follows from applying (S1.1) to the class $\mathcal{G}_\epsilon$.

When it comes to the bootstrap version, applying the multiplier inequality in Lemma S1.2 to $\mathcal{G}_\epsilon$, we get

$$\mathbb{E}_{ZM} \||| \mathbb{G}_n^* \||| \lesssim \mathbb{E}_Z|G_\epsilon|\frac{1}{\sqrt{n}}\mathbb{E}_M\left|\max_i M_{ni}\right| + \mathbb{E}_Z\left|\max_{n_0\leq k\leq n} \| \mathbb{G}_k \|\right|, \tag{S1.8}$$

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:

$$P\{\max_{k\leq n} \| \mathbb{G}_k \| > \lambda\} \leq 2P\{\| \mathbb{G}_n \| > \lambda\}$$

which makes the second term negligible. □

An immediate consequence of Lemma S1.4 is Lemma S1.5 showing the negligibility of related terms in our Theorem 4.2 and Theorem 4.5. The claims on I_{3n}^c and I_{3n}^{c*} directly follow from results (S1.5) and (S1.6). Regarding I_{2n}^b and I_{2n}^{b*}, the proofs are even easier because both α_0 and β_0 are fixed in those two terms.

Lemma S1.5. Suppose Conditions (1)-(8) hold. We have the characterization of following smaller order terms:

$$\sqrt{n} I_{2n}^b = o_p(1) \quad \sqrt{n} I_{3n}^c = o_p(1), \quad \text{and} \tag{S1.9}$$

$$\sqrt{n} I_{2n}^{b*} = o_{P_M}(1) \quad \sqrt{n} I_{3n}^{c*} = o_{P_M}(1), \tag{S1.10}$$

in P_Z-probability.

Now we prove several preparatory lemmas related to the linear representation of I_{2n}^a.

Lemma S1.6. Suppose Conditions (1)-(8) hold. The following representations hold:

$$I_{2n}^a = -\int \phi_{\alpha_0}(u)(\hat{F}_n(u;\beta_0) - \delta_1) dP(u, \delta_1), \tag{S1.11}$$

where

$$\phi_{\alpha_0}(u) = g_0(u - \alpha_0)/g_0(u). \tag{S1.12}$$

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

$$\begin{aligned} I_{2n}^a &= -\int \Big(\hat{F}_n(u + \alpha_0; \beta_0) - F_0(u + \alpha_0)\Big) g_0(u) du \\ &= -\int \Big(\hat{F}_n(u; \beta_0) - F_0(u)\Big) g_0(u - \alpha_0) du \\ &= -\int \phi_{\alpha_0}(u) \Big(\hat{F}_n(u; \beta_0) - F_0(u)\Big) dG_0(u) \\ &= -\int \phi_{\alpha_0}(u)(\hat{F}_n(u;\beta_0) - \delta_1) dP(u, \delta_1), \end{aligned} \tag{S1.13}$$

where in the last equality we have used 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

$$p(u, \delta_1) = F_0(u)^{\delta_1}(1 - F_0(u))^{1-\delta_1} g_0(u). \tag{S1.14}$$

□

We consider the piece-wise constant version of ϕ_{F_0} which is constant on the same intervals where the NPMLE $\hat{F}_n(\cdot;\beta)$ remains constant. Denote those intervals by $[\tau_i, \tau_{i+1})$. We define

$$\bar{\phi}_{\alpha_0}(u) = \phi_{\alpha_0}(\hat{A}_n(u;\beta)), \tag{S1.15}$$

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 uniform convergence results are available in Lemma 5.9 in Groeneboom and Wellner (1992), Lemma 3.1 of Groeneboom and Hendrickx (2018), and the corollary of Lemma 3.1 in Groeneboom and Hendrickx (2017).

Lemma S1.7. Suppose Conditions (1)-(8) hold, then we have

$$P\left(\lim_{n\to\infty} \sup_{\beta\in\mathcal{B},u} \left|\hat{F}_n(u;\beta) - F_0(u;\beta)\right| = 0\right) = 1, \tag{S1.16}$$

and

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

Furthermore, for the bootstrap version, we have

$$P_M\left(\lim_{n\to\infty} \sup_{\beta\in\mathcal{B},u} \left|\hat{F}^*_n(u;\beta) - F_0(u;\beta)\right| = 0|Z\right) = 1, \tag{S1.17}$$

and

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

in P_Z-probability.

Lemma S1.8. Suppose Conditions (1)-(8) hold, then we have

$$\| \bar{\phi}_{\alpha_0} - \phi_{\alpha_0} \|_\infty \lesssim \| \hat{F}_n(u;\beta_0) - F_0(u) \|_\infty . \tag{S1.18}$$

Proof. Our proof adapts the argument in Lemma A.4 of Groeneboom, Jongbloed, and Witte (2010). Let ξ to denote some intermediate value in Taylor expansions whose value may change from line to line. For given u in the support of $F_0(\cdot;\beta)$, then the interval J_i it belongs to is of one of the following three types:

(i) $F_0(x;\beta) > \hat{F}_n(\tau_i;\beta)$ for all $x \in J_i$;

(ii) $F_0(x;\beta) > \hat{F}_n(x;\beta)$ for some $x \in J_i$;

(iii) $F_0(x;\beta) < \hat{F}_n(\tau_i;\beta)$ for all $x \in J_i$;

see Figure 5 in Groeneboom, Jongbloed, and Witte (2010). First of all, when $F_0(u;\beta) = \hat{F}_n(u;\beta)$, then by definition of $\bar{\phi}_{\alpha_0}(u) = \phi_{\alpha_0}(u)$ so both the left-hand side and right-hand side of (S1.18) are zero.

Next, for the case where $F_0(u;\beta) \neq \hat{F}_n(u;\beta)$. For $v, \xi \in J_i$, we get by Taylor expansion:

$$|\hat{F}_n(u;\beta)-F_0(u;\beta)| = |\hat{F}_n(v;\beta)-F_0(u;\beta)| = |\hat{F}_n(v;\beta)-F_0(v;\beta)-(u-v)f_0(\xi;\beta)|. \tag{S1.19}$$

There are three possibilities. If $\hat{A}_n(u;\beta) = \tau_i$, then $F_0(\tau_i;\beta) - \hat{F}_n(\tau_i;\beta) > 0$ which gives rise to

$$|\hat{F}_n(u;\beta) - F_0(u;\beta)| = |(u-\tau_i)f_0(\xi;\beta) + F_0(\tau_i;\beta) - \hat{F}_n(\tau_i;\beta)| \geq |u-\tau_i| f_0(\xi;\beta), \tag{S1.20}$$

where we apply (S1.19) while letting $v = \tau_i$.

If $\hat{A}_n(u;\beta) = v$ for some $v \neq u \in J_i$, then $\hat{F}_n(v;\beta) = F_0(v;\beta)$ so

$$|\hat{F}_n(u;\beta) - F_0(u;\beta)| = |\hat{F}_n(v;\beta) - F_0(v;\beta) - (u-v)f_0(\xi;\beta)| = |u-v| f_0(\xi;\beta). \tag{S1.21}$$

If $\hat{A}_n(u;\beta) = \tau_{i+1}$, then $\hat{F}_n(\tau_{i=1}-;\beta) > F_0(\tau_{i+1}-;\beta) \geq 0$ giving us

$$|\hat{F}_n(u;\beta) - F_0(u;\beta)| = |(\tau_{i+1}-u)f_0(\xi;\beta) + \hat{F}_n(\tau_{i+1}-;\beta) - F_0(\tau_{i+1};\beta)| \geq |\tau_{i+1}-u| f_0(\xi;\beta). \tag{S1.22}$$

In sum, we obtain

$$|\hat{F}_n(u;\beta) - F_0(u;\beta)| \geq |v-u| f_0(\xi;\beta) \geq c|u-v|, \tag{S1.23}$$

where we use the fact that the density function is bounded below by a positive constant over the compact interval.

Finally, the imposed smoothness condition on ϕ_{α_0} leads to

$$|\bar{\phi}_{\alpha_0}(u) - \phi_{\alpha_0}(u)| = |\phi_{\alpha_0}(v) - \phi_{\alpha_0}(u)| \leq c|v-u|, \tag{S1.24}$$

where $v \in [\tau_i, \tau_{i+1}]$. Hence, the desired claim follows by combining (S1.23) and (S1.24). □

Given the above lemmas, we get the following characterization of I_{2n}^a and its bootstrapped version I_{2n}^{*a}.

Lemma S1.9. Suppose Conditions (1)-(8) hold, then we have the following linear representations:

$$\sqrt{n} I_{2n}^a = \sqrt{n} P \left[F_0 - \hat{F}_n(\cdot, \beta_0) \right] = \mathbb{G}_n \psi_{F_0} + o_p(1). \tag{S1.25}$$

and

$$\sqrt{n} I_{2n}^{*a} = \sqrt{n} P \left[F_0 - \hat{F}_n^*(\cdot, \beta_0) \right] = \sqrt{n} \left[\mathbb{P}_n^* - P \right] \psi_{F_0} + o_{p_M}(1), \tag{S1.26}$$

in P_Z-probability.

Proof. We only proof the claim regarding I^a_{2n} to avoid repetition. Given the characterization of our isotonic estimate $\hat{F}_n(u;\beta_0)$ and the piece-wise constant nature of $\bar{\phi}_{\alpha_0}$, we get

$$\int \bar{\phi}_{\alpha_0}[\hat{F}_n(u;\beta_0) - \delta_1]d\mathbb{P}_n = 0, \tag{S1.27}$$

by equality (8.15) in Groeneboom and Jongbloed (2014). Therefore, starting with the alternative representation of I^a_{2n} we get

$$I^a_{2n} = \int \bar{\phi}_{\alpha_0}(\hat{F}_n(u;\beta_0) - \delta_1)d(\mathbb{P}_n - P) \tag{S1.28}$$

$$+ \int [\bar{\phi}_{\alpha_0} - \phi_{\alpha_0}](\hat{F}_n(u;\beta_0) - \delta_1)dP(u,\delta_1). \tag{S1.29}$$

In the next lemma, we show that

$$\int \bar{\phi}_{\alpha_0}(\hat{F}_n(u;\beta_0) - \delta_1)d(\mathbb{P}_n - P) = \int \phi_{\alpha_0}(F_0(u) - \delta_1)d(\mathbb{P}_n - P) + o_p(n^{1/2}), \tag{S1.30}$$

and

$$\int [\bar{\phi}_{\alpha_0} - \phi_{\alpha_0}](\hat{F}_n(u;\beta_0) - \delta_1)dP(u,\delta_1) = O_p(n^{-2/3}), \tag{S1.31}$$

which lead to the desired conclusion. □

Lemma S1.10. Suppose Conditions (1)-(8) 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

$$S_n \equiv \int [\bar{\phi}_{\alpha_0} - \phi_{\alpha_0}](\hat{F}_n(u;\beta_0) - \delta_1)dP(u,\delta_1) = o_p(n^{-1/2}). \tag{S1.32}$$

Proof. We first handle the term S_n as follows.

$$S_n = \int [\bar{\phi}_{\alpha_0} - \phi_{\alpha_0}](\hat{F}_n(u;\beta_0) - F_0(u))dG(u) \tag{S1.33}$$

$$\lesssim \| \hat{F}_n(u;\beta_0) - F_0(u) \|^2_\infty = O_p(n^{-2/3} \times \log^2 n), \tag{S1.34}$$

where in the second step we used (S1.8).
Referring to the term R_n, we introduce some notations adapted from Groeneboom, Jongbloed, and Witte (2010). Define

$$\xi_B(u) = \bar{\phi}_{\alpha_0}(u)B(u) \tag{S1.35}$$

where B is a function of bounded variation and with bounded superior norm C. And let

$$\mathcal{G}_C \equiv \{\xi_B(u) : B \in \mathcal{B}_C\}. \tag{S1.36}$$

Given the result in (S1.7), for any small $\gamma > 0$ we can find finite constant term C such that for all n sufficiently large:

$$\begin{aligned} Pr\{\Upsilon_{n,C}\} \equiv Pr\{&\sup_{u,\beta} |\hat{F}_n(u;\beta) - F_0(u;\beta)| \leq Cn^{-1/3}\log n\} \\ &\geq 1 - \gamma/2. \end{aligned}$$

Now for the vanishing sequence ν_n to be specified later, it is straightforward to arrive at

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

for any small γ. Again by (S1.7), we have

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

The rest of our proof is to utilize 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., $J(1, \mathcal{G}_C) < \infty$. And the L_2-norm of the envelope function is also bounded. Thus applying (S1.1), we have

$$E|R_n| \leq n^{-5/6}\log n \times E \sup_{\xi\in\mathcal{G}_C} \left| \int \xi(u) d\mathbb{G}_n(u)\right| \lesssim n^{-5/6}\log n, \tag{S1.37}$$

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

□

Lemma S1.11. Suppose Conditions (1)-(8) hold. We have the characterization of following smaller order terms:

$$\sqrt{n} I_{3n}^b = o_P(1) \quad \text{and} \quad \sqrt{n} I_{3n}^{b*} = o_{P_M}(1), \tag{S1.38}$$

in P_Z-probability.

Proof. Again we only prove the part involving I_{3n}^b as the other one follows analogously. First of all, 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 S1.9, we get

(S1.39)

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

The smoothness assumption in Condition (5) implies that the function in the bracket of (S1.39) belongs to a P-Donsker class by Example 19.7 in Van Der Vaart (1998). Now the convergence of $\hat{\alpha}_n$ and $\hat{\beta}_n$ leads to the desired conclusion that $\sqrt{n}I_{3n}^b = o_P(1)$. □

S2. Appendix C: Technical Proofs Related to Joint Estimation

We first prove the main theorem which shows the root-n consistency and asymptotic normality of our estimators for the finite dimensional parameters. The overall structure of the proof is similar to the one of Theorem 4.2. Thus, we skip some straightforward intermediate steps and highlight the changes that we make. Some technical details are separately verified in lemmas of this section.

Related to the P-Glivenko-Cantelli or P-Donsker property, it turns out that 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 $N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|_2)$ for subclass $\mathcal{F}$ is defined to be the minimum of m such that $\exists\ f_1^L, f_1^U, ..., f_m^L, f_m^U$ for $\forall f \in \mathcal{F}$, $f_j^L \le f \le f_j^U$ for some j, and $\left\|f_j^U - f_j^L\right\|_2 \le \epsilon$. Denote $H_{[]}(\epsilon, \mathcal{F}, \|\cdot\|_2) \equiv \log N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|_2)$.

We need a few more notations in order to characterize the consistency and rate of convergence. Let p_0 be the true density function and p be a member of the class of densities, denoted by $\mathcal{P}$. We denote

$$\bar{q} = \frac{q+q_0}{2}, \tag{S2.1}$$

$$\bar{\mathcal{P}} \equiv \{\bar{q} : q \in \mathcal{P}\}, \tag{S2.2}$$

$$\bar{\mathcal{P}}^{1/2} \equiv \{\bar{q}^{1/2} : \bar{q} \in \bar{\mathcal{P}}\}. \tag{S2.3}$$

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

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

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

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

Similarly as in Geskus and Groeneboom (1997), we define the following density functions: (S2.4)

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

We record Theorem 7.4 in Van de Geer (2000) which is needed to obtain the rate of convergence in our context.

Lemma S2.1 (Theorem 7.4 in Van de Geer (2000))**.** Given a upper bound $\Psi(\delta)$ for the entropy integral function in such a way that $\Psi(\delta)/\delta^2$ is a non-increasing function of δ. Then for a universal constant c and for $\delta_n \to 0$ such that

$$\sqrt{n}\delta_n^2 \geq c\Psi(\delta_n),$$

then the likelihood estimator $\tilde{q}_n$ converges with rate $O_p(\delta_n)$. In fact, there exist constants L_0, C_0, s.t. for all $L \geq L_0$

$$\Pr\{\mathbf{h}(\bar{q}, q_0) \geq L\delta_n\} \leq C_0 \exp\left(-L^2\delta_n^2 n\right).$$

The next lemma delivers the rate of convergence for the NPMLE in terms of L_2 norm uniformly over the finite dimensional parameter.

Lemma S2.2. Regarding the convergence by L_2 norm, we have

$$\sup_{\theta} \| \tilde{F}_n(u; \theta) - F_0(u; \theta) \|_2 = O_p\left(\log^2 n \times n^{-1/3}\right). \tag{S2.5}$$

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

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

The only complication comes from the term

$$\mathcal{F}_D \equiv \left\{\sqrt{F(x'\beta + \alpha; \theta) - F(x'\beta; \theta)} : (\alpha, \beta, F)\right\}. \tag{S2.6}$$

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

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

In sum, for any probability measure Q, we get

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

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 S2.1 to get

$$\sup_{\theta}\mathbf{h}(q_{\tilde{F}_n,\theta},q_{F_0,\theta})=O_p(\log^2 n\times n^{-1/3}). \tag{S2.7}$$

Also, not that

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

we get that

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

□

Next, we fill the details related to our proofs of the consistency and asymptotic normality for our joint estimation method. Specifically, we prove the existence of zero-crossing points and show the stochastic equicontinuity of negligible terms related to our estimating equations.

Existence of Zero-crossing Points. 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

$$\Phi_n(\theta)=\dot{\Phi}_{\theta_0}(\theta-\theta_0)+r_n(\theta), \tag{S2.8}$$

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

$$\Phi_{n,h}(\theta)=\dot{\Phi}_{\theta_0}(\theta-\theta_0)+\tilde{r}_{n,h}(\theta), \tag{S2.9}$$

with

$$\tilde{r}_{n,h}(\theta)=h^{-d}\int k_h(u_1-\alpha)\cdots k_h(u_K-\beta_K)r_n(u_1,\cdots,u_K)du_1\cdots du_K, \tag{S2.10}$$

where $k(\cdot)$ is a standard kernel density function supported on $[-1,1]$ and $\beta'=(\beta_1,\cdots,\beta_d)'$. Note that $\lim_{h\to 0}\tilde{r}_{n,h}(\theta)=r_n(\theta)$.

We re-parameterize as follows by defining

$$\gamma = \dot{\Phi}_{\theta_0}\theta, \quad \text{and} \quad \gamma_0 = \dot{\Phi}_{\theta_0}\theta_0. \tag{S2.11}$$

This gives

$$\Phi_{n,h}(\theta) = \gamma - \gamma_0 + \tilde{r}_{n,h}(\dot{\Phi}_{\theta_0}^{-1}\gamma). \tag{S2.12}$$

Given the result in S2.8, the mapping

$$\gamma \mapsto \gamma_0 - \tilde{r}_n(\dot{\Phi}_{\theta_0}^{-1}\gamma) \tag{S2.13}$$

maps, for each $\delta > 0$, the ball $B_\delta(\gamma_0) = \{\gamma : |\gamma - \gamma_0| \le \delta\}$ into $B_{\delta/2}(\gamma_0) = \{\gamma : |\gamma - \gamma_0| \le \delta/2\}$ with probability approaching to 1. Therefore by Brouwer's fixed point theorem (Groeneboom and Hendrickx (2018)), the mapping

$$\gamma \mapsto \gamma_0 - \tilde{r}_{n,h}(\dot{\Phi}_{\theta_0}^{-1}\gamma), \tag{S2.14}$$

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

$$\Phi_{n,h}(\theta_{n,h}) = 0 \tag{S2.15}$$

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 as in Groeneboom and Hendrickx (2018) that $\Phi_n(\theta)$ has a crossing of zero at $\bar{\theta}_n$ by contradiction.
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 $\Phi_n^j(\theta) \ge c > 0$ for all $\theta \in B_\delta(\bar{\theta}_n)$ and some constant $c > 0$. Arguing as Groeneboom and Hendrickx (2018), the j-th component of $\Phi_{n,h}^j$ of $\Phi_{n,h}$ satisfies

$$\Phi_{n,h}^j(\theta) \ge \frac{c}{2}, \tag{S2.16}$$

for sufficiently small h and all $\theta \in B_\delta(\bar{\theta}_n)$, which contradicting S2.15, since $\theta_{n,h}$ for $h = 1/k_l$ belongs to $B_\delta(\bar{\theta}_n)$ for large k_l. □

Lemma S2.3. Under our conditions, we get

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

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 S1.4 by the stochastic equicontinuity of the related P-Donsker classes. The only minor 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 S1.3 does not depend on the range of the function, as long as it is finite. Hence, the results follow. □

From now on, we drop the additional index θ in defining the NPMLE or its probability limit with out confusion. Because the remaining proofs reduce to characterizing the asymptotic property of the linear functional for the NPMLE given the data and the true unknown parameter θ_0. Also, 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 sample; i.e., $\tilde{F}_n(u) < 1$ for any u in the support. However, this plays a minor role regarding the large sample properties we study because the defectiveness does not occur with probability 1 as sample size goes to infinity.

Lemma S2.4 (Proposition 1 in Geskus and Groeneboom (1997))**.** We have

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

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

Lemma S2.5. Any function σ that is constant at the same intervals as $\tilde{F}_n$ satisfies

$$\text{(S2.17)} \qquad \int \sigma(u) \left[\frac{\delta_1}{\tilde{F}_n(u)} - \frac{1-\delta_1-\delta_2}{1-\tilde{F}_n(u+\alpha_0)} + \frac{\delta_2}{\tilde{F}_n(u+\alpha_0)-\tilde{F}_n(u)} \right] dQ_n(u,\delta_1,\delta_2) = 0.$$

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

$$\text{(S2.18)} \qquad \bar{\xi}(u) = \xi_{\tilde{F}}(\tilde{A}_n(u;\alpha,\beta)),$$

where

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

for $u \in J_i$.

A key element in determining the asymptotic distribution of $(\tilde{\alpha}_n, \tilde{\beta}_n)$ is the linear functional of NPMLE.

Lemma S2.6. Under our conditions, we have

$$\sqrt{n}(\kappa(\tilde{F}_n) - \kappa(F_0)) = \int \phi_{F_0} d(Q_n - Q_{F_0}) + o_p(n^{-1/2}). \tag{S2.19}$$

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

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

$$\sqrt{n}(\kappa(\tilde{F}_n) - \kappa(F_0)) = \sqrt{n} \int \tilde{\kappa}_{F_0} d(\tilde{F}_n - F_0). \tag{S2.20}$$

Step 2. The second step is similar as in the proof of our Lemma S1.6 where we apply the integration-by-parts. Now we have

$$\int \tilde{\kappa}_{F_0} d(\tilde{F}_n - F_0) = -\int \phi_{\tilde{F}_n} dQ_{F_0}, \tag{S2.21}$$

in our Lemma S2.9, utilizing the relationship between the linear operator L and its adjoint L^*.

Step 3. Next, we need to consider the piecewise approximation of ϕ_F which has the same jump locations as the NPMLE. By the characterization lemma S2.5, one gets

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

Thus, we have

$$-\int \phi_{\tilde{F}_n} dQ_{F_0} = -\int \bar{\phi}_{\tilde{F}_n} d(Q_n - Q_{F_0}) + o_p(n^{-1/2}), \tag{S2.22}$$

in which we have used that

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

as proved in our Lemma S2.10.

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

$$-\int \bar{\phi}_{\tilde{F}_n} dQ_{F_0} = -\int \phi_{F_0} d(Q_n - Q_{F_0}) + \int \left[\bar{\phi}_{\tilde{F}_n} - \phi_{F_0}\right] d(Q_n - Q_{F_0}), \tag{S2.23}$$

where the second term is shown to be negligible asymptotically. Note that ϕ_F is the integrated score function, therefore it is absolutely continuous with respect to F; also see our Lemma S2.7. Given that all F are (sub) distribution functions, then ϕ_F is of bounded variation, so is its piece-wise constant approximation $\bar{\phi}_F$. 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 $(\bar{\phi}_{\tilde{F}_n} - \phi_{F_0})$. And 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. Now, $\int \left[\bar{\phi}_{\tilde{F}_n} - \phi_{F_0}\right] d(Q_n - Q_{F_0}) = o_p(n^{-1/2})$ follows from the stochastic equicontinuity of the related P-Donsker class. In the end, we arrive at

$$-\int \bar{\phi}_{\tilde{F}_n} dQ_{F_0} = -\int \phi_{F_0} d(Q_n - Q_{F_0}) + o_p(n^{-1/2}). \tag{S2.24}$$

□

Next, we list a few intermediate lemmas that are similar as the ones in Appendix B, so instead of presenting the full proofs we only highlight the necessary changes. First, the following lemma states that the piece-wise constant function $\bar{\xi}(u)$ is absolutely continuous w.r.t the NPMLE $\tilde{F}_n$, combining the closed-form expression of ϕ_F in Van de Geer (1997) and the proof of Lemma 4 in Geskus and Groeneboom (1996).

Lemma S2.7. The derivative of ϕ_F at the points of continuity is bounded, uniformly over F and the points of continuity; i.e.,

$$|\phi_F(y) - \phi_F(x)| \leq C_1|y - x|, \tag{S2.25}$$

for y and x in the same interval between jumps and a finite positive constant C_1. Moreover, the jumps satisfy

$$|\phi_F(x) - \phi_F(x-)| \leq C_2|F(x) - F(x-)|, \tag{S2.26}$$

with a finite positive constant C_2.

The next lemma controls the approximation error for a function ξ with respect to its piece-wise version $\bar{\xi}$ determined by the NPMLE. The proof is identical to the one of Lemma S1.8.

Lemma S2.8. Suppose our Conditions hold, then we have

$$\| \bar{\xi}_{\tilde{F}_n}(u) - \xi_{\tilde{F}_n}(u) \|_2 \lesssim \| \tilde{F}_n(u; \alpha_0, \beta_0) - F_0(u) \|_2 . \tag{S2.27}$$

Lemma S2.9. Under our conditions, We have

$$\int \tilde{\kappa}_{F_0} d(\tilde{F}_n - F_0) = -\int \phi_{\tilde{F}_n} dQ_{F_0}. \tag{S2.28}$$

Proof. The idea of the proof follows from Lemma 1 in Geskus and Groeneboom (1997)[p.251]. Let $1 \in L_2(F)$ denote the constant function $1(u) \equiv 1$ for any $u \in [C_L, C_U]$. Note that for this constant function, $L(1) = 1$ after applying the L transform. Thus, we have

$$\int \phi_F dQ_{F_0} = \langle \phi_F, L(1) \rangle_{Q_{F_0}}$$
$$= \langle L^* \phi_F, 1 \rangle_{F_0} = \int L^*(\phi_F) dF_0.$$

The desired conclusion immediately follows if we can show

$$L^*(\phi_F) = \tilde{\kappa}_{F_0} - \int \kappa_{F_0} dF, \tag{S2.29}$$

and then set F equal to $\tilde{F}_n$. To see this, recall the solution ϕ_{F_0} is obtained by differentiating Ł$^*\phi_F(u) = \tilde{\kappa}_F(u)$ with $\phi_F = \phi_{F_0}$. Then by integrating, we get

$$Ł^* \phi_F(u) = \tilde{\kappa}_{F_0}(u) + C,$$

for some constant C. To pin down this constant term, we use the fact that F is non-defective so that

$$C = \int C dF$$
$$= \int L^* \phi_F dF - \int \tilde{\kappa}_{F_0} dF = \langle L^* \phi_F, 1 \rangle_F - \int \tilde{\kappa}_{F_0} dF.$$

Now it is immediate that ϕ_F is contained in $L_2^0(Q_F)$. We also have

$$\langle L^* \phi_F, 1 \rangle_F = \langle \phi_F, L(1) \rangle_F = \langle \phi_F, 1 \rangle_F = 0,$$

which completes the proof. □

The following lemma characterizes some smaller order term in our Step 3 while analyzing $\sqrt{n}(\kappa(\tilde{F}_n) - \kappa(F_0)$.

Lemma S2.10. Under our conditions, we have

$$\int \left(\bar{\phi}_{\tilde{F}_n} - \phi_{\tilde{F}_n} \right) dQ_{F_0} = o_p(n^{-1/2}). \tag{S2.30}$$

Proof. We start by defining the function φ_n

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

and the function ξ_F through the relationship $\xi_F(u) \equiv \phi_F(u)F(u)(1-F(u+\alpha_0))$.
Then we obtain

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

We apply the Cauchy-Schwarz inequality to get

$$\left|\int\left(\bar{\phi}_{\tilde{F}_n}-\phi_{\tilde{F}_n}\right)dQ_{F_0}\right| \le C \parallel \bar{\xi}_{\tilde{F}_n}-\xi_{\tilde{F}_n}\parallel_2 \times \parallel \tilde{F}_n - F_0 \parallel_2 . \tag{S2.31}$$

Following the analogous argument as in the proof of our Lemma S1.8, we get

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

Now the result follows from the convergence result that $\parallel \tilde{F}_n(u) - F_0(u) \parallel_2 = O_p(\log^2 n \times n^{-1/3})$. □

We complete the section by computing the Hessian matrix related to our joint estimation method.

Lemma S2.11. Recall that the Hessian matrix is

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

then we have

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

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

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

First of all, we have

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

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 $j=2,\cdots,K$:

(S2.32)

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

It is evident that the first term on the right-hand side of (S2.32) is equal to $\mathbb{E}[(X - E[X|X'\beta_0])'f_0(X'\beta_0+\alpha_0)]$. 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 (S2.32) is zero, when evaluated at $\theta=\theta_0$. Therefore, the desired result follows. □

References

BALADBAOUI, F., P. GROENEBOOM, AND K. HENDRICKX (2017): "Score estimation in the monotone single index model," *working paper*.

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.

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.

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.

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

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

——— (1997): "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. (1998): *Asymptotic statistics*. Cambridge University Press.

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