Econometric Theory, **40**, 2024, 1–36.
doi:10.1017/S0266466622000317

SIMPLE SEMIPARAMETRIC ESTIMATION OF ORDERED RESPONSE MODELS

Ruixuan Liu
Chinese University of Hong Kong

Zhengfei Yu
University of Tsukuba

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 parameters, 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 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. The advantages of our methods are borne out in simulation studies and a real data application.

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:

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

We are grateful to the Editor (Peter C.B. Phillips), the Co-Editor (Simon Lee), and two anonymous referees for their comments and suggestions that greatly improved the paper. We would like to thank Jason Abrevaya, Bo Honoré, Hidehiko Ichimura, Hiroyuki Kasahara, Toru Kitagawa, Brendan Kline, Zhongjian Lin, Oliver Linton, Essie Maasoumi, Taisuke Otsu, Aureo de Paula, Adam Rosen, Katsumi Shimotsu, and Haiqing Xu; seminar participants at UT Austin and SUNY Albany; and participants at the 2018 North American Econometric Society Meetings, the 2018 CEME Conference for Young Econometricians, and the Workshop on Advances in Econometrics 2018 for their insightful comments and suggestions. Yu gratefully acknowledges the support of JSPS KAKENHI Grant Numbers 19K13666 and 21K01419. All errors are ours. Address correspondence to Zhengfei Yu, Faculty of Humanities and Social Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8571, Japan; e-mail: yu.zhengfei.gn@u.tsukuba.ac.jp.

for $i = 1,\ldots,n$. We maintain the independence assumption between X and ε throughout the paper. Let $F_0(\cdot)$ be the true unknown distribution function of the latent error ε. 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 are $\Delta_{ji} = \mathbb{I}\{Y_i = j\}$, for $j \in \{1,2,3\}$.

The ordered response model dates back to Aitchison and Silvey (1957) where the error distribution F_0 is parameterized and has been widely used to characterize ordered categorical outcomes in economics. 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 has been studied by Lee (1992), Melenberg and Van Soest (1996), Klein and Sherman (2002), Lewbel (2000, 2002), Chen and Khan (2003), and Coppejans (2007), allowing for an arbitrary error distribution. 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), Chen and Khan (2003), and Coppejans (2007). The implication is that the user has to choose a tuning parameter, such as the bandwidth in kernel smoothing or the number of sieve basis functions, and there is no clear answer about the optimal choice in this context.[1] Inevitably, this requires a considerable amount of intervention and judgment on the part of practitioners. The second branch, which does not require tuning parameters, includes the maximum score and maximum rank estimation. For the maximum score estimator (Lee, 1992), only the consistency result is available, and it is expected to have a nonstandard limiting distribution with cubic-root convergence rate. The rate of convergence can be improved by smoothing the sample criterion function, as done by Melenberg and Van Soest (1996), which once again introduces a smoothing parameter (bandwidth). Moreover, the convergence rate of the smoothed maximum-score (SMS) estimator remains slower than the standard root-n rate (Horowitz, 1992, 2009). Using the maximum rank estimation, it is possible to develop a two-stage estimator for the ordered response model. The first stage estimates the regression coefficient β_0 by recasting the ordered response model as a generalized regression model (Han, 1987; Sherman, 1993; Cavanagh and Sherman, 1998), and the second stage estimates the threshold parameter α_0 by adapting the rank estimator for the transformation model (Chen, 2002). We describe such a two-stage rank estimator in Section 2.4.1 and treat it as an alternative to our main proposal.

[1] In Lewbel (2000), under an additional independence assumption between the special regressor and other covariates (see Assumption A.5' on page 157 of Lewbel (2000)), one can apply the ordered data estimator (Lewbel and Schennach, 2007) without a tuning parameter. However, in general, the Lewbel (2000) estimator needs a kernel- or sieve-type estimator of the conditional density in its first stage.

https://doi.org/10.1017/S0266466622000317 Published online by Cambridge University Press

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 method consists of two stages. The first stage starts with the likelihood function for related binary choice data $(\Delta_{1i}, X_i)_{i=1}^n$:

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

and follows Groeneboom and Hendrickx (2018) to get estimated $\hat{\beta}_n$ and $\hat{F}_n$. Using the monotonicity of the distribution function, the estimated $\hat{F}_n(\cdot;\beta)$ for any given β can be obtained by the isotonic regression of Δ_{1i} on $X_i'\beta$ or, equivalently, as the nonparametric maximum likelihood estimator (NPMLE) in the sense of Kiefer and Wolfowitz (1956) for the *binary choice* data.[2] We then estimate the regression coefficients by using a set of moment conditions. In the second stage, we obtain the estimated threshold from a simple moment condition concerning the binary choice data (Δ_{3i}, X_i), for $i = 1, \ldots, n$. Our second method 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 coefficients and threshold *jointly* by using the moment conditions as in our two-stage approach. Throughout this paper, we name the first approach (isotonic) two-stage estimation and the second one (NPMLE-based) joint estimation.

Our estimation approaches have three main appealing features. First, both methods are free from any tuning parameter. This is because we estimate the error distribution F in (1.3) or (1.2) by a well-defined isotonic estimator $\hat{F}_n$ or NPMLE $\tilde{F}_n$ (either using the binary choice data or the ordered response data), which exploits the monotonicity of the distribution function. As a result, the estimator $\hat{F}_n$ or $\tilde{F}_n$ does not rely on any kernel smoothing or sieve penalization. Second, our estimators of the error distribution are automatically nondecreasing functions by construction. In contrast, the kernel-based approach in Klein and Sherman (2002) may not yield a monotonic estimate of the error 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 $\hat{F}_n$ can be easily computed using the pool-adjacent-violators algorithm (PAVA) (see Robertson, Wright, and Dykstra (1988, Chap. 1) for details). For the NPMLE $\tilde{F}_n$, we adapt the hybrid approach in Wellner and Zhan (1997) that combines both the expectation-maximization (EM) 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

[2] In order to differentiate from the NPMLE in our second method, we will refer $\hat{F}_n$ as the *isotonic estimator* throughout the paper and reserve the NPMLE for $\tilde{F}_n$. Indeed, $\hat{F}_n$ can be obtained by an isotonic regression, whereas $\tilde{F}_n$ cannot.

[3] Figure 5 in Section 4.2 plots the estimated error distribution in a real data example.

monotonic. In our simulations, the two-stage estimator is computationally fastest among five semiparametric estimation methods under consideration.

We contribute to the literature in several ways. First, we propose a new tuning-parameter-free semiparametric method for practitioners to estimate the ordered response model. The Monte Carlo results also confirm robust finite sample performances of our proposals. Second, our interest in model (1.1) stems from the interdependent durations model proposed by Honoré and de Paula (2010), in which the scalar representing the social interaction effect is directly related to the threshold parameter in the resulting ordered response model. We demonstrate the usefulness of the Honoré–de Paula model by estimating the derived ordered response model in simulation experiments and real data. Third, our work contributes to the literature of semiparametric estimation that involves shape-restricted components (Groeneboom and Jongbloed, 2014). In the seminal works of Newey (1994), Chen, Linton, and Van Keilegom (2003)), and Ichimura and Lee (2010), general theorems are presented for semiparametric estimators involving some first-stage nonparametric estimation under high-level assumptions and then verified for sieve- or kernel-type estimators under smoothness restrictions. In our setting, however, the isotonic estimate or NPMLE is not smooth. The crux of our theoretical investigation is to prove that certain linear functionals (or the directional derivatives) of the estimated distribution function are asymptotically normal, combining the characterization of shape-restricted estimation and empirical process theory. Fourth, we prove bootstrap consistency to facilitate inference. Note that the bootstrap is known to fail for the pointwise distribution of the isotonic estimator or NPMLE (Abrevaya and Huang, 2005). However, similar to Groeneboom and Hendrickx (2017), the bootstrap is valid for our finite-dimensional parameters because the influence of the isotonic estimator or the NPMLE is carried over by linear functionals.

Our technical analysis is built on Groeneboom and Hendrickx (2018); however, there are distinctions. A close examination reveals that the proof in Groeneboom and Hendrickx (2018) regarding the regression coefficient β is relatively 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). On the other hand, we have to explicitly characterize the influence of estimating the distribution through its linear functional in our two-stage estimation. The corresponding issue related to our joint estimator goes beyond Groeneboom and Hendrickx (2018). Unlike the binary choice case, in our model, the NPMLE making use of information in all three categories lacks an explicit characterization. As a result, determining the asymptotic behavior of its linear functional becomes much more challenging, and we adapt the proof of NPMLE for the "interval censoring, case 2 model"[4] (Van de Geer, 1995; Geskus and Groeneboom, 1996,

[4]To clarify the comparison 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).

https://doi.org/10.1017/S0266466622000317 Published online by Cambridge University Press

1997, 1999). From a different perspective, the linear functional of the isotonic distribution estimator can be analyzed under the framework of Beare and Fang (2017) using their least concave majorant (LCM) operator, because the isotonic estimator can be equivalently defined as a left derivative of a proper LCM (see Mukherjee and Sen (2019) for an application to the integral functional of the isotonic density estimator). Unfortunately, the theory of Beare and Fang (2017) does not apply to the NPMLE of our joint estimation. In response, we have to follow a more abstract route by analyzing the score operator and its adjoint (Van der Vaart, 1991).

The rest of the paper is organized as follows. Section 2 verifies the identification, and introduces two simple semiparametric estimation methods that are based on isotonic estimation/NPMLE of the error distribution. Section 3 investigates the asymptotic properties of our estimators for the finite-dimensional parameters, proving their consistency and asymptotic normality. We also develop bootstrap procedures for the confidence intervals. Section 4 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 the Appendix, whereas other theorems and technical lemmas are proved in the Supplementary Material. The Supplementary Material also collects auxiliary results and additional simulation evidence.

2. SIMPLE SEMIPARAMETRIC ESTIMATION

Throughout the paper, we work with the i.i.d. data (Y_i, X_i) for $i = 1, \ldots, n$. It is convenient to introduce the indicators $\Delta_{ji} = \mathbb{I}\{Y_i = j\}$, for $i = 1, \ldots, n$ and $j = 1, 2, 3$. Responses with more than three categories are discussed in Section S1.3 of the Supplementary Material. Let K denote the dimensionality of covariates X, and write $\beta_0 \equiv (\beta_{01}, \beta_{02}, \ldots, \beta_{0K})'$. Note that the regression coefficient β_0 is only identified up to some scale normalization for an unspecified F_0 (Klein and Sherman, 2002). Without loss of generalization, we normalize $\beta_{01} = 1$ and denote $\beta_0' = (1, \beta_{0-}')$. In accordance, covariates are partitioned as $X' = (X_1, X_{-1}')$. Furthermore, our coefficient estimators are denoted by $\hat{\beta}_n' = (1, \hat{\beta}_{n-}')$ and $\tilde{\beta}_n' = (1, \tilde{\beta}_{n-}')$ in the sequel.

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 $\mathcal{A} \subset \mathbb{R}_+$, $\mathcal{B} \subset \mathbb{R}^{K-1}$, 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_∞-norm in the consistency proof and the L_2-norm in showing the rate of convergence, as well as killing smaller-order terms for technical convenience.

https://doi.org/10.1017/S0266466622000317 Published online by Cambridge University Press

2.1. Identification

The identification of the finite-dimensional parameters $(\beta_{0-}', \alpha_0)'$ and the nonparametric component F_0 can be achieved in two stages. First, the coefficients β_{0-} and the cdf F_0 can be identified using the binary choice data $(\Delta_{1i}, X_i')'$. Then, the threshold α_0 can be identified using another set of binary choice data $(\Delta_{3i}, X_i')'$, in particular, from the moment restriction $\mathbb{E}[\Delta_{3i}|X'\beta_0] = 1 - F_0(X'\beta_0 + \alpha_0)$. This idea of identification is inspired by Klein and Sherman (2002) and Lewbel (2002). For completeness, we state the identification conditions and results as follows.

Condition 1. (i) We observe i.i.d. data (Y_i, X_i) for $i = 1, \ldots, n$. (ii) The covariates X and latent error ε are independent. (iii) The distribution of first element X_1 conditional on other elements X_{-1} has an everywhere positive Lebesgue density. (iv) The support of X is not contained in any proper linear subspace of $\overline{\mathbb{R}}^K$.

Condition 2. (i) The distribution function F_0 is differentiable. (ii) F_0 is not constant over the support of $X'\beta_0$ and over the support of $X'\beta_0 + \alpha_0$.

Condition 3. Let $X_{-1}' \equiv (X_c', X_d')$, where $X_c \in \overline{\mathbb{R}}^{K_c}$ and $X_d \in \overline{\mathbb{R}}^{K_d}$ be the continuous and discrete components of X_{-1}. Write $\beta_-' \equiv (\beta_c', \beta_d')$ to denote the corresponding coefficients. Denote the supports of X_1, X_c, and X_d as $\mathcal{X}_1$, $\mathcal{X}_c$, and $\mathcal{X}_d$, respectively. Then, for any $\beta_- \in \mathcal{B}$, there exist $K_d + 1$ vectors $c_0, c_1, \ldots, c_{K_d} \in \mathcal{X}_d$ satisfying the following two conditions: (i) the vectors $c_l - c_0$ for $l = 1, \ldots, K_d$ are linearly independent; and (ii) the following set

$$\bigcap_{l=0}^{K_d} \left\{ x_1 + x_c'\beta_c + c_l'\beta_d : x_1 \in \mathcal{X}_1, x_c \in \mathcal{X}_c \right\}$$

contains an open interval.

Condition 4. The support of $X'\beta_0$ contains the support of ε.

Conditions 1–3 are adapted from the literature on the single-index model (Ichimura, 1993; Horowitz, 2009) and the binary choice model (Manski, 1985; Klein and Spady, 1993). For the ordered response model, the independence between ε and X in Condition 1 is also imposed by the kernel- or sieve-based estimators (Klein and Sherman, 2002; Coppejans, 2007) and the rank estimator described in Section 2.4.1. Condition 3 corresponds to Assumption 4.2(4) in Ichimura (1993), which guarantees the overlap of the support of $X'\beta$ when the discrete components of X_{-1} vary over $K_d + 1$ different values. Condition 4 is a support assumption on the linear index $X'\beta_0$, which allows us to identify the entire distribution of F_0 and thus facilitates the identification of α_0.

THEOREM 2.1. *Under Conditions 1–4, the regression coefficients β_{0-}, the threshold parameter α_0, and the distribution function F_0 in the model (1.1) are identified.*

Remark 2.1. Lewbel (2002) studied the identification and estimation of the ordered response model using the special regressor, which, in our notations, corresponds to X_1. In his approach, the independence and the support assumptions are imposed on the special regressor X_1 conditional on other covariates X_{-1}. This special regressor approach permits heteroskedasticity of the error term ε, and it can also be conveniently extended to deal with the random thresholds. Because our primary focus is on new estimation methods, we follow the classical approach from Ichimura (1993), which is also adopted in Klein and Sherman (2002) and Coppejans (2007).

2.2. Two-Stage Semiparametric Estimation

Following the identification strategy in the previous section, our two-stage estimation method first estimates the coefficients β_0 and the distribution F_0 using the binary choice data $(\Delta_{1i}, X_i)_{i=1}^n$ and then estimates the threshold parameter α_0 using another set of binary choice data $(\Delta_{3i}, X_i)_{i=1}^n$. 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 the conditional mean function, we estimate the nonparametric component by an isotonic regression which does not require tuning parameters. This is possible when the monotonicity restriction is imposed on the nonparametric component, which is natural in our setup as the distribution function is nondecreasing. We now describe the two-stage semiparametric estimation for the ordered response model. The first stage is adapted from Groeneboom and Hendrickx (2018).

Stage 1(i). For any β, we compute the estimator for $F(\cdot)$ based on the *binary choice* data $(\Delta_{1i}, X_i)_{i=1}^n$:

$$\hat{F}_n(\cdot;\beta) = \arg\max_{F\in\mathcal{F}} \sum_{i=1}^{n} \left[\Delta_{1i}\log F(X_i'\beta) + (1-\Delta_{1i})\log(1-F(X_i'\beta))\right], \tag{2.1}$$

where $\mathcal{F}$ is the class of all distribution functions.

Stage 1(ii). Given $\hat{F}_n(\cdot;\beta)$, our estimator $\hat{\beta}_n$ for the regression coefficients is the zero-crossing point[5] of the estimating equation

$$\Upsilon_n(\beta) \equiv \frac{1}{n}\sum_{i=1}^{n} X_{i,-1}\left[\Delta_{1i} - \hat{F}_n(X_i'\beta;\beta)\right] = 0. \tag{2.2}$$

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 estimating equation $\Psi_n\left(\hat{\alpha}_n, \hat{\beta}_n, \hat{F}_n(\cdot;\hat{\beta}_n)\right) = 0$, where

[5] As $\hat{F}_n(\cdot;\hat{\beta}_n)$ is a piecewise constant function, the estimating equations may not hold exactly. Therefore, we adopt Definition 4.1 from Groeneboom and Hendrickx (2018) so that the estimators are defined as zero-crossing points.

https://doi.org/10.1017/S0266466622000317 Published online by Cambridge University Press

$$\Psi_n\left(\alpha, \hat{\beta}_n, \hat{F}_n(\cdot;\hat{\beta}_n)\right) \equiv \frac{1}{n}\sum_{i=1}^{n}\left[1-\Delta_{3i}-\hat{F}_n(X_i'\hat{\beta}_n+\alpha;\hat{\beta}_n)\right]. \tag{2.3}$$

The estimator $\hat{F}_n(\cdot;\beta)$ in Stage 1(i) and its characterization date back to Ayer et al. (1955) in analyzing the current status data (Groeneboom and Wellner, 1992). The corresponding optimization problem is well defined, and it generates a piecewise constant function $\hat{F}_n(\cdot;\beta)$, which can be characterized as follows: fixing the parameter β, we consider the values of $U_i^{(\beta)} = X_i'\beta$, for $i = 1,\dots,n$. Let $U_{(1)}^{(\beta)} \le \cdots \le U_{(n)}^{(\beta)}$ be the order statistics and $\Delta_{1,(i)}^{(\beta)}, i = 1,\dots,n$, be the corresponding indicators. 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 $i = 1,\dots,n$.

In practice, $\hat{F}_n(\cdot;\beta)$ can be computed fast by the PAVA (Robertson et al., 1988, Chap. 1), which determines a nondecreasing sequence $r^* = (r_1^*, r_2^*, \dots, r_n^*)$ in the following way with $r_i^* = \hat{F}_n(U_{(i)}^{(\beta)};\beta)$.[6] Starting with the initial sequence $r^{[0]} = (\Delta_{1,(1)}^{(\beta)},\dots,\Delta_{1,(n)}^{(\beta)})$ and the unity weights $w^{[0]} = (1,\dots,1)$, in the kth step, if $r^{[k]}$ is a nondecreasing sequence, then we set $r^* = r^{[k]}$ and stop. Otherwise, there must exist an index j such that $r_{j-1}^{[k]} > r_j^{[k]}$ (i.e., the violators). Then, we update these two elements (violators) $r_{j-1}^{[k]}$ and $r_j^{[k]}$ by their weighted average: $r_{j-1}^{[k+1]} = r_j^{[k+1]} = (w_{j-1}^{[k]}r_{j-1}^{[k]} + w_j^{[k]}r_j^{[k]})/(w_{j-1}^{[k]} + w_j^{[k]})$, and also replace the two weights $w_{j-1}^{[k]}$ and $w_j^{[k]}$ by $w_{j-1}^{[k+1]} = w_j^{[k+1]} = w_{j-1}^{[k]} + w_j^{[k]}$. This process of "pool the adjacent violators" is repeated until we reach a nondecreasing sequence. Given $\hat{F}_n(\cdot;\beta)$, the computation of our Stage 1(ii) can be carried out by a spectral method (La Cruz, Martínez, and Raydan, 2006), which avoids matrix computation and efficiently solves a nonlinear system of equations. To be specific, the kth iteration of the algorithm is defined as $\beta^{[k+1]} = \beta^{[k]} - l^{[k]}\Upsilon_n(\beta^{[k]})$, where the step length $l^{[k]} = s_{k-1}'s_{k-1}/s_{k-1}'y_{k-1}$ with $s_k = \beta^{[k]} - \beta^{[k-1]}$ and $y_k = \Upsilon_n(\beta^{[k]}) - \Upsilon_n(\beta^{[k-1]})$. Note that $1/l^{[k]}$, which is called the spectral coefficient, is the least-squares solution to the equation $bs_{k-1} = y_{k-1}$ for a scalar b. In comparison, a typical iteration of the quasi-Newton method looks for an approximation of the matrix B that satisfies the equations $Bs_{k-1} = y_{k-1}$.[7] This spectral method can be implemented using the R package BB (Varadhan and Gilbert, 2009). Our Stage 2 is easy to compute, as there is a single equation and the function Ψ_n is monotone with respect to α.

Within the context of binary choice models, the isotonic estimator $\hat{F}_n(\cdot;\beta)$ is used by Cosslett (1983) to define the tuning-parameter-free profile likelihood estimator. However, only consistency results are available for Cosslett's estimator. The

[6] We call $\hat{F}_n(\cdot;\beta)$ the isotonic estimator because r^* is the solution to the constrained least-squares problem: $\min_{r_1,\dots,r_n}\sum_{i=1}^{n}\left(r_i - \Delta_{1,(i)}^{(\beta)}\right)^2$, such that $r_1 \le r_2 \le \cdots \le r_n$.

[7] To achieve the global convergence, the described spectral iteration needs to be combined with a descent condition (see (La Cruz et al., 2006, pp. 1431–1432) for details).

https://doi.org/10.1017/S0266466622000317 Published online by Cambridge University Press

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 in Stage 1(ii) adapted from Groeneboom and Hendrickx (2018). Modulo the estimated latent distribution function, one makes use of the population-level moment condition

$$\mathbb{E}[X_{-1}(\Delta_1 - F_0(X'\beta_0))] = 0, \tag{2.4}$$

and plugs the isotonic estimator $\hat{F}_n(\cdot;\beta)$ in the sample analog of (2.4). In the same spirit, Stage 2 of our procedure is based on a very simple moment condition:

$$\mathbb{E}[(1 - \Delta_3 - F_0(X'\beta_0 + \alpha_0))] = 0. \tag{2.5}$$

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 consistent with Groeneboom and Hendrickx (2018). In principle, a generalized method of moments estimator based on overidentified moment conditions could be developed.

Remark 2.2. Aside from avoiding any bandwidth selection, our procedure also sidesteps the trimming or truncation. Note that some other semiparametric methods require trimming mainly for two reasons (Ichimura and Todd, 2007): (i) trimming is sometimes needed to establish the uniform consistency of the nonparametric estimator, which in turn is required for establishing the asymptotic properties of the semiparametric estimator, and (ii) the statistic or the moment condition itself may not be well behaved without the trimming. Echoing these two points, we observe that the trimming can be avoided for our estimation method because (i) the isotonic estimator for the distribution function is uniformly consistent without trimming the support,[8] and (ii) the moments in (2.4) and (2.5) and their variances are generally well defined without any trimming. Referring to point (i), the NPMLE is consistent in terms of the Hellinger distance without any trimming (Van de Geer, 1993). This implies the pointwise consistency if the true error distribution function is absolutely continuous. Since both NPMLE and the true distribution are monotone, pointwise consistency implies uniform consistency (see Example 3.3(a) in Van de Geer (1993)). Regarding point (ii), note that the efficient score function derived from the smoothed maximum likelihood estimator of Klein and Spady (1993) is

$$\tilde{l}_\beta \equiv \frac{f_0(X'\beta_0)}{F_0(X'\beta_0)(1 - F_0(X'\beta_0))} X_{-1}(\Delta_1 - F_0(X'\beta_0)), \tag{2.6}$$

[8] When the dependent variable is unbounded, the isotonic regression is not consistent at the boundary in general. This also motivates the recent works on regularizing the isotonic estimator by the *bounded* isotonic regression (see (Chen, Lin, and Sen, 2020)). However, this is not a concern for our problem, as the dependent variable is always bounded between $[0,1]$ herein (see also the first equation on page 79 of Groeneboom and Wellner (1992)).

https://doi.org/10.1017/S0266466622000317 Published online by Cambridge University Press

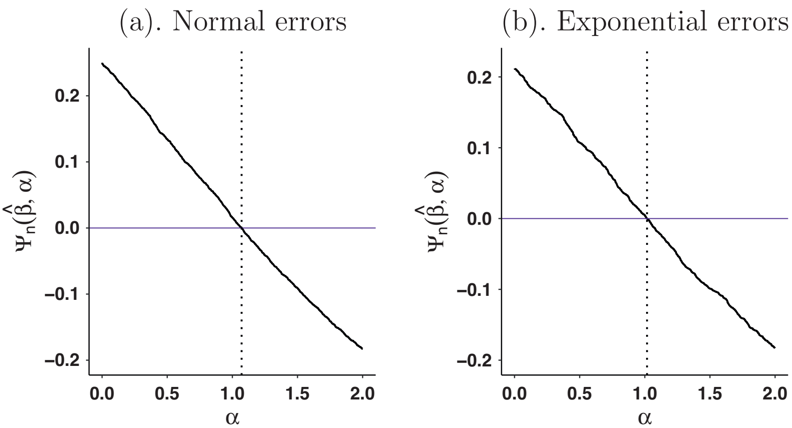

FIGURE 1. The estimating function $\Psi_n(\alpha, \hat{\beta}_n, \hat{F}_n(\cdot; \hat{\beta}_n))$ in α, generated by the interdependent duration model in Section 4.1, $\alpha^* = 1$, $n = 500$.

which may not be well defined due to the additional weighting factor $\frac{f_0}{F_0(1-F_0)}$ (see the example in Remark 2.1 in Groeneboom and Hendrickx (2018)). As specified in Condition C.7 in Klein and Spady (1993), a proper trimming is required for the Klein–Spady estimator. The efficient shape-restricted estimator in Section 4.2 of Groeneboom and Hendrickx (2018) also needs trimming as it uses the efficient score function in (2.6). We also conduct Monte Carlo simulations to evaluate the effect of trimming (see Section S2.1 of the Supplementary Material). We find that the bias and mean square error (MSE) are similar between the two-stage estimator with and without trimming. A similar observation can be made for the joint estimator proposed below.

Figure 1 depicts a typical shape of the estimating function $\Psi_n(\alpha, \hat{\beta}_n, \hat{F}_n(\cdot; \hat{\beta}_n))$ in (2.3) with respect to α. The data are generated by the interdependent duration model described in Section 4.1, with the true value $\alpha^* = 1$. The plotted estimating function $\Psi_n(\alpha, \hat{\beta}_n, \hat{F}_n(\cdot; \hat{\beta}_n))$ is decreasing in α, and the zero-crossing points are close to the true value: they are about 1.07 in Panel (a) and 1.02 in Panel (b).

2.3. Joint Semiparametric Estimation

It is natural to ask whether it is possible to develop a similar tuning-parameter-free estimation approach utilizing the entire ordered response data altogether, instead of breaking it into two sets of binary choice data. The answer is affirmative, and such an approach, termed as the joint estimation method, is introduced in this section. Our joint semiparametric method makes use of information in all three categories to estimate the distribution function and returns the estimates for the regression coefficients and the threshold parameter simultaneously.

For any α and β, we employ the NPMLE for $\tilde{F}_n(\cdot; \alpha, \beta)$ based on the *ordered response* data:

$$\tilde{F}_n(\cdot;\alpha,\beta) = \arg\max_{F\in\mathcal{F}} \prod_{i=1}^{n} \left\{ F(X_i'\beta)^{\Delta_{1i}} \left[F(X_i'\beta+\alpha) - F(X_i'\beta)\right]^{\Delta_{2i}} \left[1 - F(X_i'\beta+\alpha)\right]^{\Delta_{3i}} \right\},$$

where $\mathcal{F}$ is the class of all distribution functions. The above NPMLE is well defined, and it is a (sub)distribution function and piecewise constant with jumps over a subset of $\{X_i'\beta, X_i'\beta+\alpha : i = 1,2,\ldots,n\}$. Given $\tilde{F}_n(\cdot;\alpha,\beta)$ from the previous step, we obtain the joint estimator $(\tilde{\alpha}_n, \tilde{\beta}_n')$ as the zero-crossing point of the estimating equations simultaneously:

$$\Phi_n(\tilde{\alpha}_n, \tilde{\beta}_n) = 0, \tag{2.7}$$

where

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

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 iterative convex minorant algorithm can be implemented using the R package Icens (Gentleman and Vandal, 2018).[9] Then, the estimators $(\tilde{\alpha}_n, \tilde{\beta}_n')$ are solved from the estimating equations (2.7). The computational details are described in Section S1.2 of the Supplementary Material.

Figure 2 plots an example of the estimating functions $\Phi_n(\alpha, \tilde{\beta}_n)$ with respect to α for fixed $\beta = \tilde{\beta}_n$. The design once again follows the one in Section 4.1, with five covariates. Therefore, there are five estimating equations, as plotted in Figure 2, where the downward sloping solid line corresponds to the last row of (2.7). The estimate $\tilde{\alpha}_n$ is the value where all the estimating functions cross zero, which equals to 1.06 in Panel (a) and 1.04 in Panel (b).

2.4. Alternative Semiparametric Estimators

This section summarizes three alternative semiparametric estimators for the model (1.1). Among them, the two-stage rank estimator is also tuning-parameter-free, whereas the kernel-based estimator and the SMS estimators require the users to choose the smoothing parameter (and possibly other tuning parameters).

2.4.1. *The Two-Stage Rank Estimator.* An alternative tuning-parameter-free method is a two-stage rank estimator that combines Cavanagh and Sherman (1998) and Chen (2002). Although the original focus of Chen (2002) is estimating the unknown link function in the transformation model, his method also applies to

[9] The 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 EM algorithm.

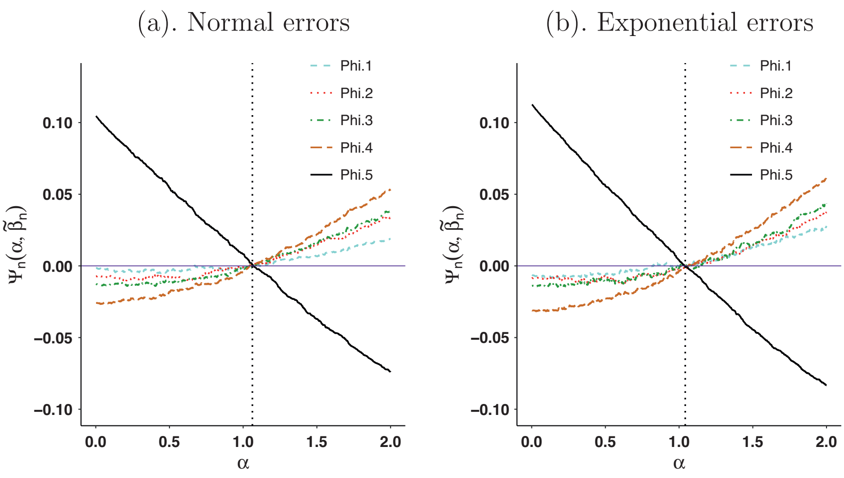

FIGURE 2. (Color online) The estimating functions $\Psi_n(\alpha, \tilde{\beta}_n, \tilde{F}_n(\cdot; \tilde{\beta}_n))$ in α, generated by the interdependent duration model in Section 4.1, Ψ_n contains five components, $\alpha^* = 1$, $n = 500$.

the estimation of the threshold parameter in the ordered response model.[10] We describe the two-stage rank estimator as follows. Note that model (1.1) implies that $\mathbb{E}[Y_i|X_i] = 3 - F(X_i'\beta_0 + \alpha_0) - F(X_i'\beta_0)$ is a decreasing function in $X_i'\beta_0$, so that one can apply the rank estimator of Cavanagh and Sherman (1998) to estimate β_{0-} in the first stage. The rank estimator $\hat{\beta}_{R-}$ is given by

$$\hat{\beta}_{R-} = \arg\max_{\beta_-} \sum_{i=1}^{n} Y_i R_n(-X_i'\beta), \tag{2.8}$$

where $R_n(-X_i'\beta)$ denotes the rank of $-X_i'\beta$. We denote $\hat{\beta}_R' = (1, \hat{\beta}_{R-}')$. Furthermore, note that $\Pr(Y_i = 1|X_i) = F_0(X_i'\beta_0)$ and $\Pr(Y_i \leq 2|X_i) = F_0(X_i'\beta_0 + \alpha_0)$. Applying the key idea of Chen (2002) yields the relationship

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

A maximum rank correlation estimator for α_0 can be obtained in the second stage:

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

Our simulation studies demonstrate a stable performance of this two-stage rank estimator for estimating the finite-dimensional parameters. In particular, the bias of the rank estimator is small. On the other hand, its standard error and the overall

[10]This approach is suggested by an anonymous associate editor in an early submission of the paper to another journal. Given that this two-stage rank estimator has not been proposed for estimating the ordered response model, we expand our discussion here. The duality between the ordered response model and the transformation model is also mentioned in Klein and Sherman (2002, p. 665).

https://doi.org/10.1017/S0266466622000317 Published online by Cambridge University Press

MSE are larger than those of the isotonic two-stage estimator and the NPMLE-based joint estimator. This gap is larger for the estimation of β_{0-} than for α_0.

2.4.2. *The Kernel-Based Estimator of Klein and Sherman (2002).* This approach also consists of two stages. In the first stage, the regression coefficient is estimated by maximizing the following quasi-likelihood function:

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

where $\hat{P}_j(X_i'\beta)$ is the kernel estimator of the conditional probability $P_j(X_i'\beta) \equiv \Pr(Y_i \le j|X_i'\beta)$ for fixed β in Klein and Spady (1993), $j=1,2$, and the trimming function $\hat{\tau}(x) = \mathbb{I}\left\{|x| \le \hat{\xi}\right\}$ with $\hat{\xi}$ being a 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

$$\hat{\alpha} = \frac{1}{\mathbb{I}\{i \in \mathcal{T}\}} \sum_{i \in \mathcal{T}} (\hat{V}_i - \tilde{V}_{i2}), \tag{2.9}$$

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 $i \in \mathcal{T}$, $\mathcal{T} = \left\{\hat{V}_i : \hat{P}_L \le \hat{P}_1(\hat{V}_i) \le \hat{P}_U\right\}$, and $(\hat{P}_L, \hat{P}_U)$ are determined by the pth and $(1-p)$th quantiles of a collection of estimated probabilities (Klein and Sherman (2002, p. 671)). In addition to the choice of bandwidth for $\hat{P}_j(X_i'\beta)$, this approach relies on the trimming scheme in the construction of the target set $\mathcal{T}$, which excludes individual estimators $\hat{V}_i - \tilde{V}_{i2}$ with poor performance. Our simulations study in Section 4.1 finds that the performance $\hat{\alpha}$ is sensitive to the trimming parameter p.

2.4.3. *The Smoothed Maximum-Score Estimator.* Horowitz (1992) initially proposed the SMS estimator for the binary choice model. Melenberg and Van Soest (1996) extended it to the ordered response model. The estimator in Melenberg and Van Soest (1996) can also be viewed as the smoothed version of Lee's (1992) maximum-score estimator. Under the median independence condition, the SMS approach estimates (α_0, β_{0-}') by maximizing the smoothed sample criterion function:

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

[11] In the 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 pages 671 and 672 of Klein and Sherman (2002).

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 the root-n rate, even with the MSE-optimal bandwidth of order $n^{-1/9}$ (for the fourth-order kernel; see Section 4.3.3 of Horowitz (2009) for a detailed discussion). Simulation results in Section 4.1 find that SMS yields larger MSEs than our two-stage and joint estimators, especially for the estimation of the threshold.

In the simulation experiments, our two-stage and joint estimators are faster and easier to compute than Klein and Sherman's (2002) kernel estimator and SMS estimator. The first reason is that the latter two require the choice of bandwidth, which costs considerable computational resources when the data-driven bandwidth selector such as cross-validation is implemented. The second reason concerns Klein and Sherman's (2002) estimator: solving $\tilde{V}_{i2}$ from the shift restriction $\hat{P}_2(\tilde{V}_{i2}) = \hat{P}_1(\hat{V}_i)$ turns out computationally expensive.

3. ASYMPTOTIC RESULTS

This section consists of two subsections that provide asymptotic results for our isotonic two-stage estimator and NPMLE-based joint estimator, respectively.

3.1. Asymptotic Properties of the Two-Stage Estimator

The crux of our investigation 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. For the slope coefficients β_{0-}, we apply Groeneboom and Hendrickx (2018). The proof in Groeneboom and Hendrickx (2018) regarding the regression coefficient directly utilizes an orthogonal (to the nuisance tangent set) score function to incorporate the estimation effect of the error distribution. This orthogonal direction involves the conditional mean of covariates X given the true linear index $U = X'\beta_0$ (Ichimura, 1993; Klein and Spady, 1993). In comparison, we have made efforts to determine the influence function for the threshold parameter α_0 by explicitly characterizing the effect of estimating the unknown distribution through certain linear functional that represents the directional derivative of the estimated distribution (Ichimura and Lee, 2010).

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

$$F_0(u;\beta) \equiv P\left\{\Delta_{1i}^{(\beta)} | U_i^{(\beta)} = u\right\} = \int F_0(u + x'(\beta_0 - \beta)) f_{X|X'\beta}(x|X'\beta = u)dx. \tag{3.1}$$

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

https://doi.org/10.1017/S0266466622000317 Published online by Cambridge University Press

In the sequel, we let $F_0(u) = F_0(u;\beta_0)$. Denote the true linear index by $U_i = X_i'\beta_0$ for $i = 1,\dots,n$. Let $G_0(u)$ and $g_0(u)$ be the distribution and density functions of the random variable U. The following two terms appear in the Taylor expansion in our asymptotic analysis:

$$V_{\alpha_0} = \frac{\partial}{\partial\alpha}\mathbb{E}[F_0(X'\beta_0+\alpha)]\big|_{\alpha=\alpha_0}, \tag{3.2}$$

$$V_{\beta_0} = \frac{\partial}{\partial\beta_-}\mathbb{E}[F_0(X'\beta+\alpha_0;\beta)]\big|_{\beta_-=\beta_{0-}}, \tag{3.3}$$

whereas

$$H_{\beta_0} = \mathbb{E}\left[f_0(X'\beta_0)\left\{X_{-1}-\mathbb{E}[X_{-1}|X'\beta_0]\right\}^{\otimes 2}\right] \tag{3.4}$$

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

The following regularity conditions are adapted from Ichimura (1993), Klein and Spady (1993), Klein and Sherman (2002), and Groeneboom and Hendrickx (2018).

Condition 5. The true β_{0-} belongs to the interior of $\mathcal{B}$ where $\mathcal{B}$ is a compact set in $\mathbb{R}^{K-1}$. 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 6. The function $F_0(u;\beta)$ is twice continuously differentiable on the interior of the support for all β. The function $F_0(\cdot)$ has a strictly positive continuous derivative, which stays away from zero.

Condition 7. The random variable $X'\beta$ admits a continuous density function denoted by $g_0(u;\beta)$ for all β. For $\beta = \beta_0$, the random variable $[\Delta_1 - F_0(U_i)]g_0(U_i-\alpha_0)/g_0(U_i)$ has a finite second moment.

Condition 8. The density $g_0(u;\beta)$ and conditional expectations $\mathbb{E}[X_{-1}|X'\beta = u]$ and $E[X_{-1}X_{-1}'|X'\beta = u]$ are twice continuously differentiable with respect to u. The functions $\beta \mapsto g_0(u;\beta)$, $\beta \mapsto \mathbb{E}[X_{-1}|X'\beta = u]$, and $\beta \mapsto \mathbb{E}[X_{-1}X_{-1}'|X'\beta = u]$ are continuous functions for u in the definition domain and all β.

Condition 9. The matrix H_{β_0} is of full rank. The scalar $V_{\alpha_0} \neq 0$, where $V_{\alpha_0} = \int f_0(u+\alpha_0)g_0(u)du$.

The asymptotic property of $\hat{\beta}_n$ is stated in Theorem 4.1 on page 1426 of Groeneboom and Hendrickx (2018), and more generally in Theorem 3 on page 532 of Balabdaoui, Groeneboom, and Hendrickx (2019). Specifically, $\hat{\beta}_n$ is root-n consistent and asymptotically normal. Regarding the latent error distribution, one gets a cubic-root rate (modulo the logarithm factor) convergence in the L_2-norm. Note that the statement in Groeneboom and Hendrickx (2018) applied trimming on the distribution function (to be coherent with the efficient estimators they developed). Here, we do not need any trimming, in the same spirit of Proposition 2 and Theorem 3 in Balabdaoui, Groeneboom, and Hendrickx (2019). We refer readers to Lemma S6 and its discussion in the Supplementary Material for more details.

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

$$\Psi_n(\alpha) = \frac{1}{n}\sum_{i=1}^{n}\left[1-\Delta_{3i}-\hat{F}_n(X_i'\hat{\beta}_n+\alpha;\hat{\beta}_n)\right] \tag{3.5}$$

and its probability limit $\Psi(\alpha) = \int\left[1-\Delta_3-F_0(X'\beta_0+\alpha)\right]dP$. Because our estimation procedure belongs to the general Z-estimation with bundled parameter and nuisance functional components, we follow the route in Nan and Wellner (2013). Unlike the examples in Nan and Wellner (2013), which have nuisance nonparametric components that are either estimable with root-n rate or subject to certain smoothness restriction, the nonparametric part is estimated utilizing shape restriction with a cubic-root rate in our model.

THEOREM 3.1. *Under Conditions 1–9, for all large n, the unique zero-crossing point $\hat{\alpha}_n$ of $\Psi_n(\alpha)$ exists with probability tending to one and it is a consistent estimator of α_0. Moreover, the following linear representations hold:*

$$\sqrt{n}(\hat{\beta}_{n-}-\beta_{0-}) = \mathbb{G}_n\left[\psi_{\beta_0}(Z_i)\right]+o_p(1), \text{and} \tag{3.6}$$

$$\sqrt{n}\left(\hat{\alpha}_n-\alpha_0\right) = V_{\alpha_0}^{-1}\mathbb{G}_n\left[(\psi_0+\psi_{F_0}+V_{\beta_0}\psi_{\beta_0})(Z_i)\right]+o_p(1), \tag{3.7}$$

where

$$\psi_{\beta_0}(Z_i) = H_{\beta_0}^{-1}(X_{i,-1}-\mathbb{E}[X_{i,-1}|U_i])(F_0(U_i)-\Delta_{1i}), \tag{3.8}$$

$$\psi_0(Z_i) = [1-F_0(U_i+\alpha_0)-\Delta_{3i}], \tag{3.9}$$

$$\psi_{F_0}(Z_i) = \frac{g_0(U_i-\alpha_0)[\Delta_1-F_0(U_i)]}{g_0(U_i)}. \tag{3.10}$$

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}. 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 3.1. *Under Conditions 1–9, we have*

$$\sqrt{n}(\hat{\theta}_n-\theta_0) \Rightarrow \mathbb{N}(0,\Sigma_0), \tag{3.11}$$

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

https://doi.org/10.1017/S0266466622000317 Published online by Cambridge University Press

Remark 3.1. It is a central theme in semiparametric econometrics to determine the influence of the first-stage nonparametric estimation. General theorems available (Newey, 1994; Chen et al., 2003; Ichimura and Lee, 2010) lead to

$$\sqrt{n}\left[S_n(\theta_0)+\Gamma(\theta_0)[\hat{F}_n-F_0]\right] \Rightarrow \mathbb{N}(0,\Sigma), \tag{3.12}$$

for some finite positive definite matrix Σ, given some generic nonparametric estimator $\hat{F}_n$. Herein, $S_n(\theta_0)$ stands for the (normalized) oracle score function for the parametric part, whereas the directional derivative $\Gamma(\theta_0)[\hat{F}_n-F_0]$ encodes the estimation effect of the nonparametric component. The latter one can further be shown to have a linear representation

$$\sqrt{n}\left[\Gamma(\theta_0)[\hat{F}_n-F_0]\right] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_i + o_p(1), \tag{3.13}$$

for some zero-mean and square integrable random variables ψ_i. The examples in Newey (1994), Chen et al. (2003), and Ichimura and Lee (2010) are about nonparametric components with sufficient smoothness restrictions and estimated by sieve- or kernel-type estimators. The essential part in our proof of Theorem 3.1 is to verify (3.13) by showing that the linear functional of the shape-restricted nonparametric estimator is asymptotically normal. The verification in our context is nontrivial due to the fact that the isotonic estimator or the NPMLE is neither smooth nor linear.

Remark 3.2. As the two-stage rank estimator described in Section 2.4.1 is another tuning-parameter-free estimation method for the ordered response model, we comment on the relative efficiency between these two approaches. Section S1.1 of the Supplementary Material presents the asymptotic variance of the two-stage rank estimator. We highlight the hybrid nature of our isotonic estimator and the one from Groeneboom and Hendrickx (2018): after the profiled NPMLE, the finite-dimensional parameters are estimated by moment conditions (or estimating equations) that correspond to inefficient scores. Thus, they do not satisfy the generalized information equality. The main reason for the lack of efficiency ranking between the isotonic estimator and the rank estimator lies in their different ways of deviating from the efficient estimator (Coppejans, 2007). Our estimator is not efficient, as it uses a simple moment condition rather than the complicated efficient score. The efficiency loss in the maximum rank estimator, on the other hand, lies in its ignorance of the information contained in the distribution function. In the numerical example presented in Section S1.1 of the Supplementary Material, the variances of the rank estimator for elements of β_{0-} are almost twice as large as those of the isotonic two-stage estimator over a wide range of values for β_{0-}. When the estimation of α_0 is concerned, on the other hand, the isotonic estimator has a smaller variance for small values of α_0, whereas the rank estimator enjoys a smaller variance for large values of α_0. Our Monte Carlo study finds consistent results that the isotonic two-stage estimator yields a smaller variance (and also a smaller MSE) than the rank approach for estimating β_{0-} and α_0, with the gap larger for the

https://doi.org/10.1017/S0266466622000317 Published online by Cambridge University Press

estimation of β_{0-} than for α_0. Overall, we conjecture that our isotonic two-stage estimator may have a smaller asymptotic variance in some circumstances, given that it is likelihood-based, and also uses the information of the error distribution and the moment conditions.[13] In a recent study on the monotone single-index model, Groeneboom and Hendrickx (2019) numerically compare the isotonic-regression-based approach and the rank estimation approach, and find that the former performs better than the latter in estimating the regression coefficients.

Despite the fact that we have a closed-form representation here, a plug-in estimation of the asymptotic variance involves estimating density functions such as $g_0(\cdot)$ or $f_0(\cdot)$. This motivates us to propose a simple bootstrap as follows. Note that bootstrap for regression coefficients in the binary choice model can be found in Groeneboom and Hendrickx (2017). Let $(M_{n1},\dots,M_{nn}) \sim \text{Multi}(n,(1/n,\dots,1/n))$ be the multinomial weights.

Stage 1(i)*. First of all, the bootstrap version $\hat{F}_n^*(\cdot,\beta)$ 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:

$$\frac{1}{n}\sum_{i=1}^{n} M_{ni}X_i\left[\Delta_{1i} - \hat{F}_n^*(X_i'\hat{\beta}_n^*;\hat{\beta}_n^*)\right] = 0. \qquad \textbf{(3.14)}$$

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

$$\frac{1}{n}\sum_{i=1}^{n} M_{ni}\left[1 - \Delta_{3i} - \hat{F}_n^*(X_i'\hat{\beta}_n^* + \hat{\alpha}_n^*;\hat{\beta}_n^*)\right] = 0. \qquad \textbf{(3.15)}$$

Since the bootstrap estimate $\hat{F}_n^*$ is a stepwise monotone function, the estimating equation for $\hat{\alpha}_n^*$ is also 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 $(Z_1,\dots,Z_n)$ 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, pp. 3464–3465). A

[13] We thank an anonymous referee for the comments on the asymptotic variance comparison.

https://doi.org/10.1017/S0266466622000317 Published online by Cambridge University Press

direct consequence of the following theorem is the validity of percentile bootstrap confidence intervals.

THEOREM 3.2 (Bootstrap validity for the two-stage estimator). *Suppose Conditions 1–9 hold. For the bootstrap estimators $\hat{\beta}^*_{n-}$ and $\hat{\alpha}^*_n$ with the multinomial weights $(M_{n1},\ldots,M_{nn})$, we have*

$$\sqrt{n}(\hat{\beta}^*_{n-} - \hat{\beta}_{n-}) \Rightarrow \mathbb{N}(0, \Omega_{\beta_0}), \tag{3.16}$$

$$\sqrt{n}\left(\hat{\alpha}^*_n - \hat{\alpha}_n\right) \Rightarrow V^{-1}_{\alpha_0} \times \mathbb{N}(0, \Omega_{\alpha_0}), \tag{3.17}$$

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

3.2. Asymptotic Properties of the Joint Estimator

The asymptotic analysis of our joint estimation method is more involved, due to the challenge to pin down the influence function capturing the effect of the NPMLE $\tilde{F}_n$. This is expressed via some linear functional of $\tilde{F}_n$ after eliminating smaller-order terms. Unlike the binary choice case where the isotonic estimator can be characterized as the left derivative of the greatest convex minorant of certain cumulative sum diagram (described in Section 2.2), such an explicit representation is lacking herein. Thus, we seek an alternative characterization that builds on a sequence of research by Van de Geer (1995) and Geskus and Groeneboom (1996, 1997, 1999) for the interval censored data (Case 2). This calls for a careful analysis of the related score operator and its adjoint (Van der Vaart, 1991). For that purpose, we define $c_1(u) = \int_{C_L}^{u} \mathbb{E}[X_{-1}|v]g_0(v)dv$, $c_3(u) = G_0(u-\alpha_0)$, and $c(u) = (c_1'(u), c_3(u))'$. Let $\dot{c}(u)$ be its derivative. Consider the linear functional $\kappa(F_0) = \int c(v)dF_0(v)$ and its canonical (with zero mean) gradient $\tilde{\kappa}_F(u) = c(u) - \int c(v)dF(v)$. 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. We denote $u = x'\beta_0$ and the support of it as $[C_L, C_U]$. Let δ_1 and δ_2 be the values of Δ_1 and Δ_2. It turns out that the influence function takes the following form (Van de Geer, 1995, Exam. 4.2):

$$\phi_{F_0}(u,\delta_1,\delta_2) = \delta_1 \frac{\varsigma_{F_0}(u)}{F_0(u)} + \delta_2 \frac{\varsigma_{F_0}(u+\alpha_0) - \varsigma_{F_0}(u)}{F_0(u+\alpha_0) - F_0(u)} - (1-\delta_1-\delta_2)\frac{\varsigma_{F_0}(u+\alpha_0)}{1-F_0(u+\alpha_0)}, \tag{3.18}$$

where

$$\varsigma_{F_0}(u) = \begin{cases} -F_0(u)\,[(1-F_0(u))\omega(u) + (1-F_0(u+\alpha_0))\omega(u+\alpha_0)], & \text{for } C_L \le u \le \alpha_0, \\ (1-F_0(u))\,[F_0(u)\omega(u) + F_0(u-\alpha_0)\omega(u-\alpha_0)], & \text{for } \alpha_0 \le u \le C_U, \end{cases}$$

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

We need one more set of assumptions to guarantee the asymptotic normality for the linear functional of NPMLE (see page 31 of Van de Geer (1995)).

https://doi.org/10.1017/S0266466622000317 Published online by Cambridge University Press

Condition 10. We assume that $\omega(u)$ is uniformly bounded for all u in the support. Moreover, the following ratios are all uniformly bounded:

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

for some universal finite constant c. Furthermore, for any α in the parameter space, we assume that $F_0(u+\alpha) - F_0(u)$ 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 $\theta_0 \equiv (\alpha_0, \beta'_{0-})'$. Let the Hessian matrix evaluated at the true parameter be

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

The following theorem presents large sample properties of our joint estimator. Note that the term ϕ_{F_0} captures the estimation effect of the distribution function and we sought to verify an expansion like (3.13) in the proof.

THEOREM 3.3. *Under Conditions 1–10, for large n, the zero-crossing point* $\tilde{\theta}_n$ *for* $\Psi_n(\tilde{\theta}_n)$ *exists with probability tending to one and is a consistent estimator of* θ_0. *We also have*

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

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} *is defined in equation (3.18), and*

$$\phi_0 = \begin{pmatrix} [\Delta_{1i} - F_0(X_i'\beta_0)]X_{i,-1} \\ 1 - \Delta_{3i} - F_0(X_i'\beta_0 + \alpha_0) \end{pmatrix}. \tag{3.19}$$

Remark 3.3. It is known that NPMLE is more efficient than the isotonic estimator only using the binary choice 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 isotonic estimator (Groeneboom and Jongbloed, 2014, Chap. 4, Exer. 4.27). Our simulation results also confirm the theory. Apropos of the asymptotic covariances of the estimators for β_0 and α_0, the comparison between our two-stage and joint estimation is not obvious analytically, as both influence functions are complicated. Since the joint approach simultaneously estimates α_0 and β_0, one may naturally expect that it works better. This is supported by our simulation results. Practitioners may decide to choose between the two-stage estimator and the joint estimator based on their tasks and computing resources.

https://doi.org/10.1017/S0266466622000317 Published online by Cambridge University Press

Now, we describe the bootstrap inference for our joint estimator. For any α and β, we can derive the bootstrap NPMLE for $F(\cdot;\alpha,\beta)$ based on the *ordered response* data:

$$\tilde{F}_n^*(\cdot;\alpha,\beta) = \\ \arg\max_{F\in\mathcal{F}} \prod_{i=1}^n \Big\{ F(X_i'\beta)^{M_{ni}\Delta_{1i}} \big[F(X_i'\beta+\alpha) - F(X_i'\beta)\big]^{M_{ni}\Delta_{2i}} \big[1 - F(X_i'\beta+\alpha)\big]^{M_{ni}\Delta_{3i}} \Big\}.$$

Given $\tilde{F}_n^*(\cdot;\alpha,\beta)$ from the previous step, the bootstrap estimator $\tilde{\theta}_n^* \equiv (\tilde{\alpha}_n^*, \tilde{\beta}_{n-}^{*\prime})'$ is the zero-crossing point of the estimating equations simultaneously:

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

$$\frac{1}{n}\sum_{i=1}^n M_{ni}\left[1 - \Delta_{3i} - \tilde{F}_n^*(X_i'\tilde{\beta}_n + \tilde{\alpha}_n;\tilde{\alpha}_n,\tilde{\beta}_n)\right] = 0. \tag{3.20}$$

The following theorem justifies the bootstrap confidence intervals of the regression coefficients and threshold parameter for the joint estimator.

THEOREM 3.4 (Bootstrap validity for the joint estimator). *Suppose Conditions 1–10 hold. For the bootstrap estimator $\tilde{\theta}_n^*$ with the multinomial weights $(M_{n1},\dots,M_{nn})$, we have*

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

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

4. NUMERICAL RESULTS

In this section, we conduct Monte Carlo simulations to evaluate the finite sample performances of the proposed isotonic two-stage estimator and NPMLE-based joint estimators, and then apply them to an empirical example that studies the retirement timing of married couples. We focus on an ordered response model with three categories, derived from the interdependent duration model of Honoré and de Paula (2010). The original model is a non-cooperative stopping game, in which two players (e.g., a married couple), respectively, decide (T_1,T_2) as the optimal timing of switching from an initial state (having a full-time job) to an alternative state (retirement). The model setup allows the utility flow in the alternative state for one player (the husband) to depend on whether the other player (the wife) has switched or not, which causes an endogenous interaction effect. Honoré and de Paula (2010) characterize the equilibrium of the switching (retirement) time (T_1,T_2) by

$$T_j = \inf\{t : t^a \exp(X_j'\beta_0)\exp[\alpha^*\mathbb{I}\{T_{-j}\le t\}] \ge \epsilon_j\},\ \ j=1,2, \tag{4.1}$$

https://doi.org/10.1017/S0266466622000317 Published online by Cambridge University Press

where a is a power parameter on the time scale, and latent variables (ϵ_1, ϵ_2) denote random utility flows from the initial state (full-time job) and are assumed to have a joint cdf $G(\cdot,\cdot)$. Covariates X_j denote the vector of characteristics of the player j. The subscript $-j$ denotes the player j's opponent. In this way, the parameter α^* captures the endogenous interaction effect. When $\alpha^* = 0$, the model boils down to the generalized accelerated failure time model (Horowitz, 2009).

Despite the sophisticated game structure and presence of multiple equilibria, Honoré and de Paula (2010) prove the identifiability of parameters and show that this interdependent duration model induces a simple ordered response model. The ordered response variable Y contains three categories depending on whether the switching decisions for two players are sequential or simultaneous. That is, $Y = 1$ if $T_1 < T_2$, $Y = 2$ if $T_1 = T_2$, and $Y = 3$ if $T_1 > T_2$. The resulting conditional choice probabilities are

$$\begin{aligned} &\Pr\{Y=1|X_1,X_2\} = H\left((X_1-X_2)'\beta_0-\alpha^*\right), \\ &\Pr\{Y=2|X_1,X_2\} = H\left((X_1-X_2)'\beta_0+\alpha^*\right) - H\left((X_1-X_2)'\beta_0-\alpha^*\right), \end{aligned} \tag{4.2}$$

where $H(w) = \Pr\{\log(\epsilon_1) - \log(\epsilon_2) \le w\}$. This falls into the setup of ordered response model (1.1) with the normalization $\alpha^* = \alpha_0/2$ and $\varepsilon = \log(\epsilon_1) - \log(\epsilon_2)$. The covariate effect of the original interdependent duration model is captured by the coefficients β_0 and the endogenous interaction by α^*. The semiparametric estimation does not impose parametric assumptions on the joint distribution G and thus maintains the flexibility to leave the functional form of H unspecified.

4.1. Monte Carlo Simulations

In our simulation experiments, covariates X_j contain five variables $\left(X_{j1},\dots,X_{j5}\right)$ and they are independent between $j = 1,2$. The first element X_{j1} is a standard normal variable truncated over $[-5,5]$; X_{j2} is a $\chi^2(1)$ variable standardized to mean zero, variance one, and truncated from above at 3; the remaining components (X_{j3}, X_{j4}, X_{j5}) are multivariate standard normal with the pairwise correlation coefficient between X_{jk_1} and X_{jk_2} equal to $0.5^{|k_1-k_2|}$, all truncated over the interval $[-5,5]$. We consider two types of distributions of the error terms (ϵ_1,ϵ_2): (I) Normal errors: $\log(\epsilon_1)$ and $\log(\epsilon_2)$ have the truncated standard normal distribution over the interval $[-5,5]$, and (II) Exponential errors: $\log(\epsilon_1)$ and $\log(\epsilon_2)$ have the unit exponential distribution truncated from above at 5. We write the corresponding regression coefficients as $\beta_0 = (1,\beta_{02},\beta_{03},\beta_{04},\beta_{05})'$ and set the last four elements $(1,1,0,\sqrt{2})'$. The true interaction effect is $\alpha^* = 1$. All simulation results are based on $1,000$ replications. The sample size $n = 250, 500, 750$, and $1,000$. Because our empirical application only contains bounded covariates, we choose to work with bounded covariates and errors herein. The Supplementary Material reports simulation results for the unbounded cases. The findings are similar to the bounded cases considered here.

https://doi.org/10.1017/S0266466622000317 Published online by Cambridge University Press

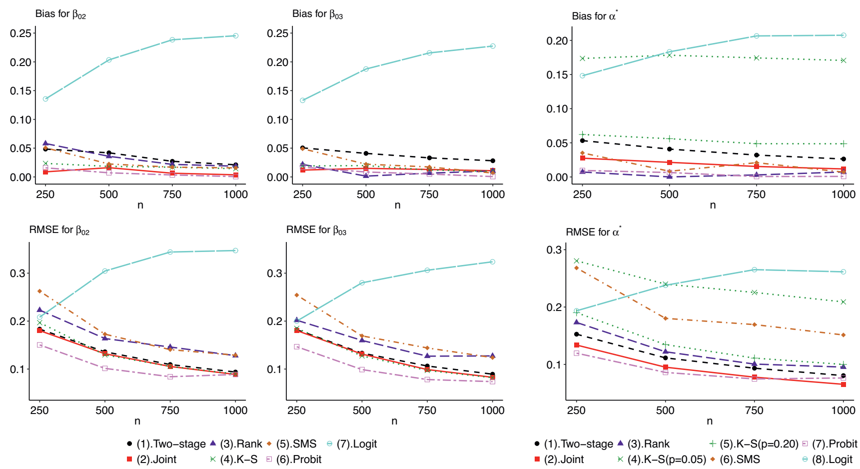

FIGURE 3a. (Color online) Finite sample performances of estimators for $(\beta_{02}, \beta_{03}, \alpha^*)$, normal error: two-stage (black, dashed, ●), joint (red, solid, ■), rank (blue, long-dashed, ▲), K–S (green, dotted, $\times$ for β_{02}, β_{03}; $\times$ and $+$ for α^* depending on p), SMS (brown, dot-dashed, ♦), ordered probit (violet, two-dashed, □), and ordered logit (cyan, very long-dashed, ∘).

Figures 3a and 3b present the finite sample bias and root mean square error (RMSE) of several estimators for $(\beta_{02}, \beta_{03}, \alpha^*)$:[14] the two-stage estimator and the joint estimator for $(\beta_{02}, \beta_{03}, \alpha^*)$ are tuning-parameter-free semiparametric estimators proposed by this paper. As a comparison, we also include three alternative semiparametric estimators described in Section 2.4, i.e., Klein and Sherman's (2002; hereafter K–S) kernel-based approach, Melenberg and Van Soest's (1996) SMS, and the rank estimator that combines Cavanagh and Sherman (1998) and Chen (2002), and the parametric ordered probit and logit estimators. 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).[15] The trimming parameter $\hat{\xi}$ for the quasi-likelihood function (see Section 2.4.2) is set to be 0.95th quantile of the euclidean norm of covariates. We experiment with two values for the trimming proportion in constructing the target set: $p = 0.05$ and $p = 0.20$. For the SMS approach, we use the kernel function in footnote 12 and select the bandwidth by cross-validation based on maximizing the sample criterion function in Section 2.4.3.

---

[14]For expository convenience, we focus on the coefficient estimators for (β_{02}, β_{03}). The performances of the estimators for (β_{04}, β_{05}) are similar.

[15]The value δ in Klein and Sherman (2002, p. 669) 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 page 670). The rate of the pilot bandwidth follows Lemma 5A in 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 page 670).

https://doi.org/10.1017/S0266466622000317 Published online by Cambridge University Press

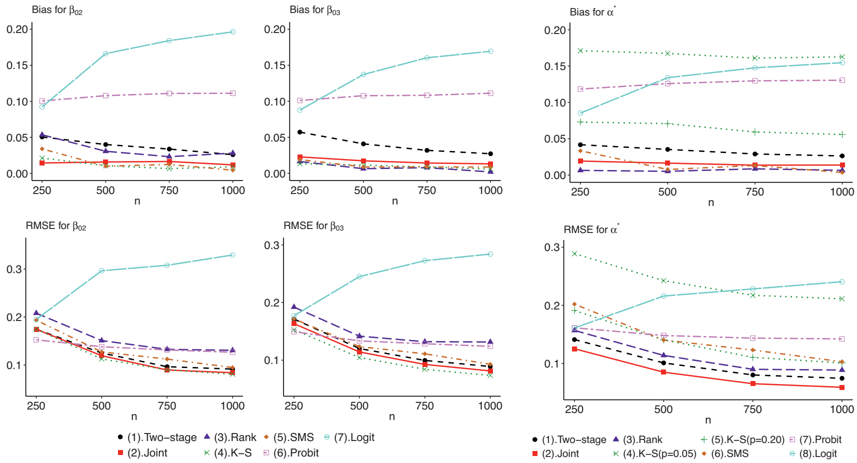

FIGURE 3b. (Color online) Finite sample performances of estimators for $(\beta_{02}, \beta_{03}, \alpha^*)$, exponential errors.

In terms of the computation time, the two-stage estimator is fastest among all semiparametric estimators under consideration mainly because the isotonic estimator $\hat{F}_n(\cdot,;\beta)$ can be computed efficiently by the PAVA. For example, one replication of the two-stage estimator for a sample with $n = 1,000$ in our Monte Carlo experiment takes about 1.4 seconds on an Intel Core i7 processor with 32 GB of RAM. The computation of the joint estimator is more time-consuming because the iterative convex minorant algorithm used to compute the NPMLE $\tilde{F}_n(\cdot,;\alpha,\beta)$ is more complicated. As a result, in the same environment, a replication of the joint estimator costs about 9.5 seconds, which is slower than the two-stage estimator but is still reasonable for practical use. In comparison, the computation time per replication for the rank, K–S, and SMS estimators are 2.0, 17.5, and 24.8 seconds, respectively.

We make the following observations regarding the finite sample performances of estimators presented in Figures 3a and 3b. First, the two-stage estimator and especially the joint estimator yield smaller RMSE than other semiparametric methods in almost all cases. For the coefficients β_{0-}, the RMSE of the two-stage estimator and the joint estimator are close to each other. For the threshold parameter α^*, the joint estimator exhibits smaller RMSE than the two-stage estimator. In comparison, the K–S estimator produces similarly small RMSE for β_{0-} but generates considerably larger RMSE for α^*. The RMSEs of the rank and SMS estimators are consistently larger than those of the two-stage estimator and the joint estimator. Second, although the bias of the two-stage estimator for β_{0-} is bigger than that of other semiparametric estimators, it is below 5% for $n = 500$ or larger. The joint estimator yields smaller bias than the two-stage estimator, reflecting a benefit of using the information across three categories. The bias of the joint estimator is similar to that of the K–S estimator for coefficients β_{0-} and smaller than that of the latter for the threshold parameter α^*. The rank estimator produces smaller bias than other semiparametric methods in most cases. Third, the ordered probit estimator performs very well in Figure 3a, as the distribution function H is (truncated) normal. However, its performance severely deteriorates in the misspecified scenario (Figure 3b). The ordered logit estimator performs badly in both Figures 3a and 3b, as the true distribution functions substantially differ from logistic. Therefore, semiparametric estimators are recommended when the H is unknown. Fourth, among the semiparametric approaches that rely on tuning parameters, the performance of the K–S estimator for α^* is sensitive to the trimming parameter p, which determines the subset used for averaging individual estimators (see Section 2.4.2). It is clear that $p = 0.05$ does not sufficiently exclude individual estimators that perform poorly and thus leads to an unreliable estimate for α^*. More trimming ($p = 0.20$) greatly improves the performance of the K–S estimator for α^*; however, its bias and RMSE are still larger than those of the two-stage estimator and the joint estimator. The SMS estimator also produces larger RMSEs than our proposals, especially for the threshold parameter α^*. In addition, the performance of SMS depends on the choice of the bandwidth. In other experiments where we shift from the cross-validation bandwidth selector to a simple rule $c \times n^{-1/9}$ for $c = 0.25, 0.5,$ and 1,[16] there are cases in which the bias increases by 50% and the RMSE by 15%.

We also examine the nonparametric estimators of the distribution H obtained by different approaches. Figures 4a and 4b present the bias and RMSE of several pointwise estimators of H evaluated at $-2, -1$, and 1.[17] The NPMLE-based joint estimator performs better than the isotonic two-stage estimator, which once again confirms the benefit of utilizing additional information that differentiates the category with $Y = 2$ and the one with $Y = 3$. We also notice that the kernel-based K–S approach outperforms the tuning-parameter-free approaches for estimating the distribution function, as the former (which uses a smoothing device) achieves faster convergence rate than the latter. Similar to the estimators of finite-dimensional parameters, the parametric probit and logit models produce substantially larger bias than the semiparametric approaches for the pointwise estimation of H, when the distribution is misspecified.

Section S2.2 of the Supplementary Material investigates the performance of the nonparametric bootstrap confidence intervals of β_{0-} and α^* obtained by the two-stage, joint, and alternative semiparametric estimators mentioned above. We find that the coverage proportions of the two-stage estimator and the joint estimator are reasonably accurate. In sum, our simulation studies demonstrate encouraging performances of the isotonic two-stage estimator and the NPMLE-based joint

[16]Note that $n^{-1/9}$ is the MSE-optimal rate for the fourth-order kernel function used here.

[17]As the rank and SMS estimators themselves do not specify an estimator for the error distribution H, we plug their estimates for the finite-dimensional parameters into the isotonic estimator described in Section 2.2, Stage 1(i) to compute the corresponding pointwise estimate for H.

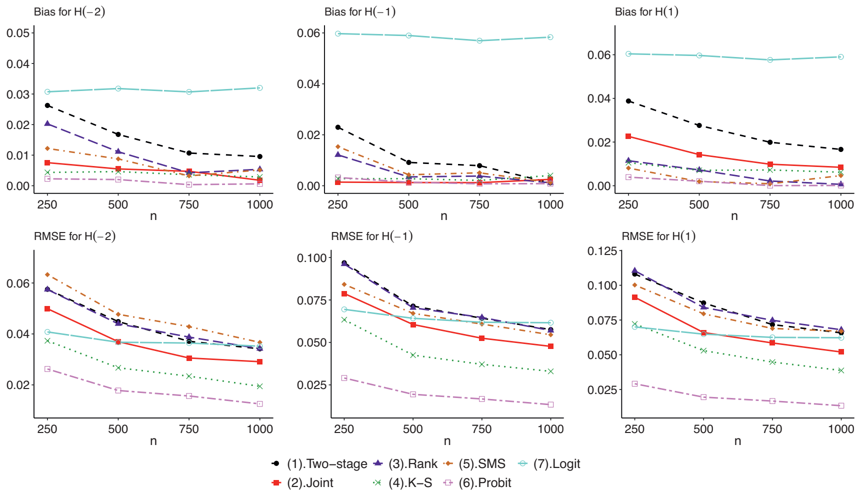

FIGURE 4a. (Color online) Pointwsie estimators for the function $H(w)$ at $w = -2, -1$, and 1, normal errors: two-stage (black, dashed, ●), joint (red, solid, ■), rank (blue, long-dashed, ▲), K–S (green, dotted, ×), SMS (brown, dot-dashed, ◆), ordered probit (violet, two-dashed, □), and ordered logit (cyan, very long-dashed, ○).

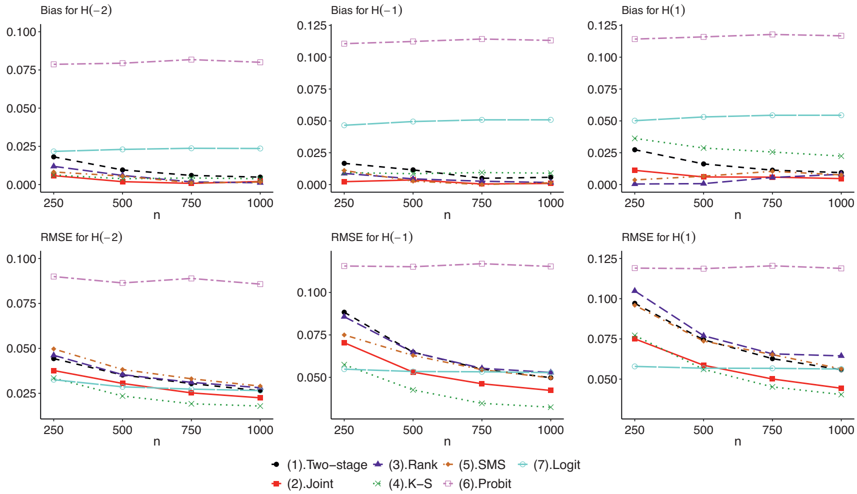

FIGURE 4b. (Color online) Pointwsie estimators for the function $H(w)$ at $w = -2, -1$, and 1, exponential errors.

estimator, especially for the threshold parameter α^*. They free the practitioners from choosing any tuning parameters and meanwhile can have smaller RMSE and shorter bootstrap confidence intervals than other competing semiparametric approaches under consideration.

https://doi.org/10.1017/S0266466622000317 Published online by Cambridge University Press

Table 1. Couples' retirement timing decisions. Sample size = 813, the coefficient on *Age* is normalized to 1, α^* = interaction parameter, 95% bootstrap confidence interval in the bracket.

	Oprobit	Two-stage	Joint	Rank	K–S	SMS
α^*	1.17	0.93	1.11	0.93	1.44	1.50
	[0.88, 1.59]	[0.59, 1.39]	[0.74, 1.45]	[0.49, 1.34]	[1.00, 3.54]	[0.02, 3.26]
College	−2.32	−1.78	−1.67	−1.41	−1.76	−1.30
	[−4.01, −0.03]	[−3.39, −0.08]	[−3.14, −0.32]	[−2.26, 0.31]	[−8.36, 1.82]	[−3.34, 2.11]
Health	−5.08	−3.79	−3.95	−4.39	−3.94	−3.31
	[−6.70, −2.23]	[−5.57, −1.82]	[−5.88, −2.19]	[−5.29, −2.24]	[−10.50, 1.40]	[−5.45, −1.50]
Black	−1.82	−0.02	−0.03	0.24	0.46	1.73
	[−10.03, 12.52]	[−3.36, 3.16]	[−0.50, 1.05]	[−0.99, 1.83]	[−5.10, 8.95]	[−4.00, 10.26]
Pension	0.37	0.17	−0.05	−0.05	−0.33	−1.28
	[−1.67, 1.57]	[−1.38, 1.44]	[−1.55, 1.44]	[−1.30, 1.29]	[−3.28, 5.39]	[−4.00, 2.50]

4.2. Empirical Application: Joint Retirement Decisions of Couples

We revisit the joint retirement decisions of married couples in the United States using the Health and Retirement Study (HRS) data, which were studied by Honoré and de Paula (2018). The HRS data provide information on the retirement time of the husbands and wives (denoted as T_1 and T_2) at a monthly frequency. We focus on the households where both partners were in the labor market in the initial period (year 1992), and at least one partner was above 60 years old. We define the three-categorical ordered outcome as follows: $Y = 1$ if $T_1 < T_2 - 1$ (husband retired first), $Y = 2$ if $|T_1 - T_2| \leq 1$ (joint retirement), and $Y = 3$ if $T_1 > T_2 + 1$. Note that if T_1 (or T_2) is right-censored due to death or the end of interview, as long as the partner retired earlier, we can still assign $Y = 3$ (or $Y = 1$) to that couple. The resulting sample contains 813 couples, among which 58.7% have $Y = 1$ and 7.3% have $Y = 2$. We estimate the ordered response model (4.2) in which the covariates X_j ($j = 1,2$ denotes the husbands and wives) contain *Age* (in years), *College* (= 1 if the person had some college or above), *Health* (= 1 if self-reported health status is good or very good), *Black* (= 1 if non-Hispanic black), and *Pension* (= 1 if had a defined benefit plan), and parameter α^* captures the interaction effect between the couple. For identification purposes, the coefficient on *Age* is normalized to 1.

Table 1 presents the estimates of the interaction parameter α^* and the coefficients on covariates obtained from various estimation methods: ordered probit (Oprobit),[18] the proposed two-stage and joint estimators, the rank estimator as an alternative tuning-parameter-free method, and kernel-based K–S and SMS estimators that rely on smoothing parameters (chosen in the same way as in Section 4.1). The estimated coefficients on *College* and *Health* are both significantly negative for the ordered probit model and our two-stage and joint estimators. The minus sign suggests that college education and a good health status reduce the utility flow of retirement relative to working and thus postpone the retirement decision. The rank estimator yields a significantly negative coefficient estimate for *Health* but an insignificantly negative estimate for *College*. K–S and SMS give insignificantly negative estimates for both of these coefficients. When it comes to coefficients on *Black* and *Pension*, all methods give insignificant results.[19] We then focus on the magnitude of the interaction parameter α^* that captures the complementarity within a household. According to the two-stage estimator and the joint estimator, this complementarity amounts to about 25%–28% of the health effect. Consider a household where the husband has a good health status but the wife does not; this health effect is a driving force toward a sequential retirement decision (the wife retires earlier). However, the estimate of the interaction parameter implies that over one quarter of the heath effect leaning toward sequential retirement will be

[18]For comparison purposes, the estimates of the parametric Oprobit are divided by the coefficient on *Age*.

[19]The K–S and SMS estimators produce longer confidence intervals for some parameters, which is also observed in our simulation results reported in Section S2.2 of the Supplementary Material. A different empirical study by (Bellemare, Melenberg, and Van Soest, 2002, pp. 194–195) finds that the SMS estimate deviates from other methods including the ordered probit, the partial linear, and the semiparametric least squares.

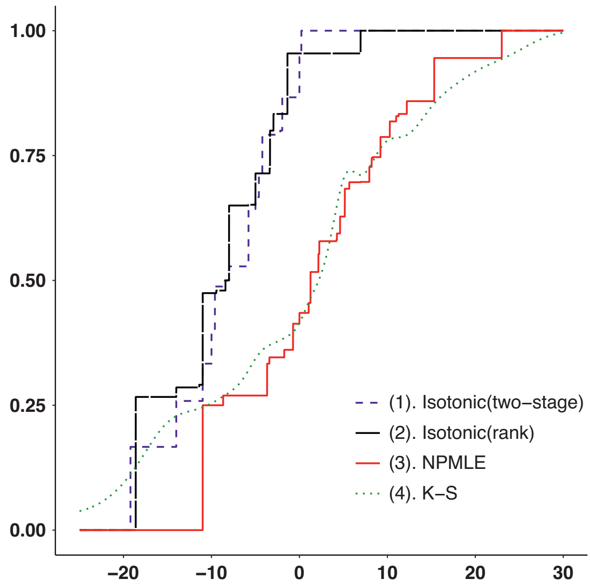

FIGURE 5. (Color online) The estimated cdf of $\log(\epsilon_1/\epsilon_2)$: the distribution of the log ratio of husband–wife unobservables, the coefficient on *Age* is normalized to 1, isotonic estimator for the cdf with the two-stage estimate plugged in (blue, dashed), NPMLE for the cdf with the joint estimate plugged in (red, solid), isotonic estimator for the cdf with the rank estimate plugged in (black, long-dashed), and K–S (green, dotted).

countered by the complementary effect that synchronizes the retirement decision of the couple.

Figure 5 plots the estimated cdf of the log ratio of the error terms (ϵ_1, ϵ_2) in the interdependent duration model. By construction, both the isotonic estimates (with either the two-stage or the rank estimate for (α^*, β_0') plugged in) and the NPMLE (with the joint estimate plugged in) are nondecreasing step functions. In contrast, the K–S estimate slightly fluctuates up and down in some parts. In addition, the NPMLE in Figure 5 is quite close to the K–S estimate. Given that in theory the smoothing estimator (e.g., the K–S estimator) of the nonparametric component converges at a faster rate, this suggests a better performance of NPMLE over the isotonic approach in estimating the nonparametric component.

5. CONCLUSION

In this paper, we have proposed two simple semiparametric estimation methods for ordered response models with an unknown error distribution. We establish the asymptotic properties of finite-dimensional parameters, tackling the challenging issues related to the nonparametric components based on NPMLE. Our methods

https://doi.org/10.1017/S0266466622000317 Published online by Cambridge University Press

are easy to implement and free of any tuning parameter, complementing several important contributions as in Lee (1992), Klein and Sherman (2002), Lewbel (2002), and Coppejans (2007). Moreover, the methodology is applicable to estimate the social interaction effect in the interdependent durations model by Honoré and de Paula (2010). Both the Monte Carlo simulation and a real data application demonstrate the usefulness of our approach.

APPENDIX

A. PROOFS OF MAIN RESULT

This appendix provides proofs of Theorems 2.1, 3.1, and 3.3. The proofs of Theorems 3.2 and 3.4 and all the lemmas are collected in the Supplementary Material. Some more tedious proofs, such as the existence of the unique zero-crossing point for Ψ_n and Φ_n, are also relegated to the Supplementary Material. In terms of notations, we denote some positive constants by c or C whose value might change line by line. For a function $f(\cdot)$ of a random vector $Z=(Y,X)$ that follows distribution P, we use the standard empirical process notations: $Pf=\int f(z)dP(z), \mathbb{P}_nf=n^{-1}\sum_{i=1}^{n}f(Z_i)$, and $\mathbb{G}_nf=n^{1/2}(\mathbb{P}_n-P)f$.

Proof of Theorem 2.1. Using the binary choice data $(\Delta_{1i},X_i')'$ and Condition 1, we have

$$\mathbb{E}[\Delta_{1i}\mid X=x]=F_0(x'\beta_0). \tag{A.1}$$

According to Theorem 4.1 in Ichimura (1993), the coefficient β_{0-} can be identified from the single index model (A.1) under Conditions 1–3 (see also Theorem 2.1 in Horowitz (2009)). Let $\mathcal{U}$ be the support of $X'\beta$, and we have

$$F_0(u)=[\Delta_{1i}\mid X'\beta_0=u],\ \text{for all } u\in\mathcal{U}.$$

By Condition 4, the entire distribution function F_0 is identified. Therefore, $F_0(X'\beta_0+\alpha)$ is knowable for all the values of X and $\alpha\in\mathcal{A}$.

We then use another set of binary choice data $(\Delta_{3i},X_i')'$ to identify α_0. For $u\in\mathcal{U}$, define

$$\Psi(\alpha,u)\equiv 1-\mathbb{E}[\Delta_{3i}\mid X'\beta_0=u]-F_0(u+\alpha).$$

We show that $\mathbb{E}[\Psi(\alpha,U)]=0$ only at $\alpha=\alpha_0$. First, it is obvious that $\mathbb{E}[\Psi(\alpha_0,U)]=0$, as $\Psi(\alpha_0,u)=0$ for all u. We then focus on the uniqueness. Suppose to the contrary that there exists an $\alpha'\neq\alpha_0$ such that $\mathbb{E}[\Psi(\alpha',U)]=0$. By the monotonicity of F_0, $\Psi(\alpha,u)$ is nonincreasing in α for all u. As a result, the only way to make $\mathbb{E}[\Psi(\alpha',U)]=0$ is $\Psi(\alpha',u)=\Psi(\alpha_0,u)$ for all u, which is equivalent to $F_0(u+\alpha')=F_0(u+\alpha_0)$ for all u. In other words, it requires that $F_0(u)=F_0(u+\alpha'-\alpha_0)$ for all $u\in\mathcal{U}+\alpha_0$, which is only possible if F_0 is periodic over $\mathcal{U}+\alpha_0$. Together with the monotonicity, it enforces the function F_0 to be constant over $\mathcal{U}+\alpha_0$, which violates Condition 2(ii). □

Proof of Theorem 3.1. Recall that the estimating equation in Stage 2 is

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

https://doi.org/10.1017/S0266466622000317 Published online by Cambridge University Press

with its probability limit $\Psi(\alpha) = \mathbb{E}\left[1 - \Delta_3 - F_0(X'\beta_0 + \alpha)\right]$. Observe that

$$\sup_\alpha |\Psi_n(\alpha) - \Psi(\alpha)| \leq \sup_\alpha \left|(\mathbb{P}_n - P)\left[\Delta_3 + \hat{F}_n(X'\hat{\beta}_n + \alpha; \hat{\beta}_n)\right]\right| + \sup_\alpha \left|P\left[\hat{F}_n(X'\hat{\beta}_n + \alpha; \hat{\beta}_n) - F_0(X'\beta_0 + \alpha)\right]\right|.$$

Hence, the uniform convergence $\sup_\alpha |\Psi_n(\alpha) - \Psi(\alpha)| \to_p 0$ holds by the Glivenko–Cantelli property in Lemma S3 in the Supplementary Material and the fact that $\hat{F}_n(u; \hat{\beta}_n)$ converges to $F_0(u)$ in the L_∞-norm w.p.1. Considering the monotone estimating equation $\Psi_n(\alpha)$ given the monotonic estimate $\hat{F}_n$, the consistency result is a direct consequence of Lemma 5.10 in Van Der Vaart (1998).

The linear representation of the coefficient estimator $\hat{\beta}_n$ and its asymptotic normality follow Theorem 3 and equation (34) on page 533 of Balabdaoui, Groeneboom, and Hendrickx (2019) with the general monotone link function ψ_0 being the cdf F_0, the response $Y = \Delta_1$, $\mathbb{S}$ being the identity function, and $J_\mathbb{S}$ being the identity matrix.[20] Here, we focus on the threshold estimator $\hat{\alpha}_n$. Note that $\Psi_n(\hat{\alpha}_n)$ is the convex combination of the left and right limits at $\hat{\alpha}_n$:

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

where we can choose $\lambda \in [0,1]$ such that the above equation holds, following Groeneboom and Hendrickx (2018). We start with the estimating equation $\Psi_n(\hat{\alpha}_n) = 0$ and decompose the left-hand side as

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

where

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

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

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

Here, $\hat{F}_n(X_i'\beta_0 + \alpha_0; \beta_0)$ is the (oracle) 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 Theorem 3.1. Recall that $U = X'\beta_0$. Referring to I_{2n}, we get $I_{2n} = I_{2n}^a + I_{2n}^b$, where

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

[20] The terms $\mathbb{S}$ and $J_\mathbb{S}$ in Balabdaoui, Groeneboom, and Hendrickx (2019) stem from their normalization scheme, which sets the euclidean norm of the coefficient vector equal to one. Recall that we instead normalize the first component to one.

https://doi.org/10.1017/S0266466622000317 Published online by Cambridge University Press

We utilize the P-Donsker property (Van Der Vaart and Wellner, 1996) to show $I_{2n}^b = o_p(n^{-1/2})$ in Lemmas S3 and S4 in the Supplementary Material. By Lemma S8 in the Supplementary Material, we obtain the linear representation for I_{2n}^a as follows:

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

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

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

Lemmas S4 and S10 in the Supplementary Material prove that $I_{3n}^b = o_p(n^{-1/2})$ and $I_{3n}^c = o_p(n^{-1/2})$ using the P-Donsker property of related functional classes. Furthermore, we have the following expansion:

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

In sum, the desired linear representation for $\hat{\alpha}_n$ follows by collecting the leading terms in I_{1n}, I_{2n}^a, and I_{3n}^a and substituting the linear representation for $\hat{\beta}_{n-}$. □

Proof of Theorem 3.3. Given the compactness of the parameter space, any subsequence of $\tilde{\theta}_n$ has a further subsequence $\tilde{\theta}_{n_k}$ converging to some element $\theta^* = (\alpha^*, \beta_-^{*\prime})'$. In the proof of Lemma S13 in the Supplementary Material, we apply Theorem 7.4 in Van de Geer (2000) to show the following convergence in terms of the Hellinger distance:

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

where the underlying density function is

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

and $\tilde{q}_{n,\theta}$ is the corresponding maximum likelihood estimator given θ. Moreover, by Lemma S13 in the Supplementary Material,

$$\sup_\theta \| \tilde{F}_n(\alpha + x'\beta; \theta) - F_0(\alpha + x'\beta; \theta) \|_2 = O_p(n^{-1/3} \log^2 n),$$

in the L_2-norm. 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:

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

Given that the true θ_0 is the unique root of the limiting function $\Phi(\cdot)$ (see Section S4 of the Supplementary Material), we must have $\theta^* = \theta_0$, which leads to the consistency of $\tilde{\theta}_n$.

As for the linear representation, denote the stacked moment conditions by

$$\zeta(Z_i; \alpha, \beta, F(;\alpha,\beta)) = \begin{pmatrix} [\Delta_{1i} - F(X_i'\beta; \alpha, \beta)]X_{i,-1} \\ 1 - \Delta_{3i} - F(\alpha + X_i'\beta; \alpha, \beta) \end{pmatrix}. \quad \textbf{(A.2)}$$

https://doi.org/10.1017/S0266466622000317 Published online by Cambridge University Press

Let $\Phi_{n,j}$ be the jth component of Φ_n in (2.7), and we represent it as a convex combination of the left and right limits at $\tilde{\theta}_n$:

$$\Phi_{n,j}(\tilde{\theta}_n) = \lambda_j \Phi_{n,j}(\tilde{\theta}_n -) + (1-\lambda_j)\Phi_{n,j}(\tilde{\theta}_n +) = 0, \tag{A.3}$$

where we can choose λ_j from the unit interval in such a way that (A.3) holds since we have a crossing of zero for each component without changing the location of the zero-crossing point. Then, we proceed with

$$\begin{aligned} 0 = {} & (\mathbb{P}_n - P)\zeta(Z;\tilde{\alpha}_n,\tilde{\beta}_n,\tilde{F}_n(\cdot;\tilde{\alpha}_n,\tilde{\beta}_n)) \\ & + P[\zeta(Z;\alpha_0,\beta_0,\tilde{F}_n(\cdot;\alpha_0,\beta_0)) - \zeta(Z;\alpha_0,\beta_0,F_0(\cdot))] \\ & + P[\zeta(Z;\tilde{\alpha}_n,\tilde{\beta}_n,\tilde{F}_n(\cdot;\tilde{\alpha}_n,\tilde{\beta}_n)) - \zeta(Z;\alpha_0,\beta_0,\tilde{F}_n(\cdot;\alpha_0,\beta_0))]. \end{aligned} \tag{A.4}$$

For the first term in (A.4), Lemma S14 in the Supplementary Material gives

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

For the third term in (A.4), applying Lemma S14 in the Supplementary Material and a Taylor expansion yields

$$\begin{aligned} & P[\zeta(Z;\tilde{\alpha}_n,\tilde{\beta}_n,\tilde{F}_n(\cdot;\tilde{\alpha}_n,\tilde{\beta}_n)) - \zeta(Z;\alpha_0,\beta_0,\tilde{F}_n(\cdot;\alpha_0,\beta_0))] \\ & \quad = P[\zeta(Z;\tilde{\alpha}_n,\tilde{\beta}_n,F_0(\cdot;\tilde{\alpha}_n,\tilde{\beta}_n)) - \zeta(Z;\alpha_0,\beta_0,F_0(\cdot))] + o_p(n^{-1/2}) \\ & \quad = H_0\begin{pmatrix} \tilde{\alpha}_n - \alpha_0 \\ \tilde{\beta}_{n-} - \beta_{0-} \end{pmatrix} + o_p(n^{-1/2} + (\tilde{\alpha}_n - \alpha_0) + |\tilde{\beta}_n - \beta_0|), \end{aligned}$$

where the Hessian matrix H_0 is calculated in Lemma S21 in the Supplementary Material.

When it comes to the second term in (A.4), we have

$$P[\zeta(Z;\alpha_0,\beta_0,\tilde{F}_n(\cdot;\alpha_0,\beta_0)) - \zeta(Z;\alpha_0,\beta_0,F_0(\cdot))] = \int c(u)d(\tilde{F}_n - F_0)(u) = \kappa(\tilde{F}_n) - \kappa(F_0), \tag{A.5}$$

where $c(u) = (c_1'(u), c_3(u))'$ is defined at the beginning of Section 3.2. To see the first equality of (A.5), note that, by (A.2), the first component of the left-hand side of (A.5) becomes

$$\int (F_0(u) - \tilde{F}_n(u))\mathbb{E}[X_{-1}|u]g_0(u)du,$$

by the law of iterated expectation. Integration by parts yields

$$\int (F_0(u) - \tilde{F}_n(u))\mathbb{E}[X_{-1}|u]g_0(u)du = \int c_1(u)d(\tilde{F}_n - F_0)(u),$$

where $c_1(u) = \int_{C_L}^{u} \mathbb{E}[X_{-1}|v]g_0(v)dv$. Similarly, the second component of the left-hand side of (A.5) becomes

$$\int (F_0(u+\alpha_0) - \tilde{F}_n(u+\alpha_0))dG_0(u) = \int G_0(u-\alpha_0)d(\tilde{F}_n - F_0)(u),$$

where $c_3(u) = G_0(u-\alpha_0)$. The second equality of (A.5) follows from the definition of κ.

Equation (A.5) reduces the problem to characterizing the asymptotic property of the linear functional for the NPMLE. Then, we apply Lemma S17 in the Supplementary

https://doi.org/10.1017/S0266466622000317 Published online by Cambridge University Press

Material to get

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

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 stated conclusion follows given the shorthand notation $\phi_0\equiv\zeta(Z;\alpha_0,\beta_0,F_0)$ in (3.19). □

SUPPLEMENTARY MATERIAL

To view the online supplementary material for this article, please visit: https://doi.org/10.1017/S0266466622000317

REFERENCES

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

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

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

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

Beare, B.K. & Z. Fang (2017) Weak convergence of the least concave majorant of estimators for a concave distribution function. *Electronic Journal of Statistics* 11(2), 3841–3870.

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

Cavanagh, C. & 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. & S. Khan (2003) Rates of convergence for estimating regression coefficients in heteroskedastic discrete response models. *Journal of Econometrics* 117, 245–278.

Chen, X., Q. Lin, & B. Sen (2020) On degrees of freedom of projection estimators with applications to multivariate nonparametric regression. *Journal of the American Statistical Association* 115, 173–186.

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

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.

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

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

Geskus, R. & P. Groeneboom (1997) Asymptotically optimal estimation of smooth functionals for interval censoring, II. *Statistica Neerlandica* 51, 201–219.

Geskus, R. & P. Groeneboom (1999) Asymptotically optimal estimation of smooth functionals for interval censoring, case 2. *Annals of Statistics* 27, 627–674.

Greene, W.H. & D.A. Hensher (2010) *Modeling Ordered Choices: A Primer*. Cambridge University Press.

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

Groeneboom, P. & K. Hendrickx (2018) Current status linear regression. *Annals of Statistics* 46, 1415–1444.

Groeneboom, P. & K. Hendrickx (2019) Estimation in monotone single-index models. *Statistica Neerlandica* 73(1), 78–99.

Groeneboom, P. & G. Jongbloed (2014) *Nonparametric Estimation Under Shape Constraints*. Cambridge University Press.

Groeneboom, P. & J.A. Wellner (1992) *Information Bounds and Nonparametric Maximum Likelihood Estimation*. Birkhauser.

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

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

Honoré, B.E. & Á. de Paula (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*, 505–531.

Horowitz, J.L. (2009) *Semiparametric and Nonparametric Methods in Econometrics*. Springer.

Ichimura, H. (1993) Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics* 58, 71–120.

Ichimura, H. & S. Lee (2010) Characterization of the asymptotic distribution of semiparametric M-estimators. *Journal of Econometrics* 159, 252–266.

Ichimura, H. & P.E. Todd (2007) Implementing nonparametric and semiparametric estimators. In J.J. Heckman, E. Leamer (eds.), *Handbook of Econometrics*, vol. 6B, Elsevier. pp. 5369–5468.

Kiefer, J. & 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. & R.P. Sherman (2002) Shift restrictions and semiparametric estimation in ordered response models. *Econometrica* 70(2), 663–691.

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

La Cruz, W., J. Martínez, & M. Raydan (2006) Spectral residual method without gradient information for solving large-scale nonlinear systems of equations. *Mathematics of Computation* 75(255), 1429–1448.

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

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

Lewbel, A. (2002) Ordered Response Threshold Estimation. Working paper, Boston College.

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

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

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

Mukherjee, R. & B. Sen (2019) On efficiency of the plug-in principle for estimating smooth integrated functionals of a nonincreasing density. *Electronic Journal of Statistics* 13, 4416–4448.

Nan, B. & J.A. Wellner (2013) A general semiparametric Z-estimation approach for case-cohort studies. *Statistica Sinica* 23(3), 1155–1180.

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

Robertson, T., F. Wright, & R. Dykstra (1988) *Order Restricted Statistical Inference*. Wiley.

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

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

Van de Geer, S. (1995) Asymptotic normality in mixture models. *ESAIM: Probability and Statistics* 1, 17–33.

Van de Geer, S. (2000) *Empirical Processes in M-Estimation*. Cambridge University Press.

Van Der Vaart, A. (1991) On differentiable functionals. *Annals of Statistics* 19, 178–204.

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

Van Der Vaart, A. & J.A. Wellner (1996) *Weak Convergence and Empirical Processes*. Springer.

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

Wellner, J.A. & Y. Zhan (1997) A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data. *Journal of the American Statistical Association* 92, 945–959.