慶應義塾大学学術情報リポジトリ
Keio Associated Repository of Academic resouces

Title	Empirical likelihood inference for monotone index model
Sub Title	
Author	大津, 泰介(Ōtsu, Taisuke) 高畑, 圭佑(Takahata, Keisuke) Xu, Mengshan
Publisher	Keio Economic Observatory Sangyo Kenkyujo
Publication year	2019
Jtitle	KEO discussion paper No.149 (2019. 7)
JaLC DOI	10.14991/004.00000149-0001
Abstract	This paper proposes an empirical likelihood inference method for monotone index models. We construct the empirical likelihood function based on a modified score function developed by Balabdaoui, Groeneboom and Hendrickx (2019), where the monotone link function is estimated by isotonic regression. It is shown that the empirical likelihood ratio statistic converges to a weighted chi-squared distribution. We suggest inference procedures based on an adjusted empirical likelihood statistic that is asymptotically pivotal, and a bootstrap calibration with recentering. A simulation study illustrates usefulness of the proposed inference methods.
Notes	
Genre	Technical Report
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AA10715861-00000149-0001

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the KeiO Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

Empirical likelihood inference for monotone index model

Taisuke Otsu, Keisuke Takahata and Mengshan Xu

July 23, 2019

KEO Discussion Paper No. 149

KEO Discussion Paper No. 149

EMPIRICAL LIKELIHOOD INFERENCE FOR MONOTONE INDEX MODEL

TAISUKE OTSU, KEISUKE TAKAHATA, AND MENGSHAN XU

Abstract. This paper proposes an empirical likelihood inference method for monotone index models. We construct the empirical likelihood function based on a modified score function developed by Balabdaoui, Groeneboom and Hendrickx (2019), where the monotone link function is estimated by isotonic regression. It is shown that the empirical likelihood ratio statistic converges to a weighted chi-squared distribution. We suggest inference procedures based on an adjusted empirical likelihood statistic that is asymptotically pivotal, and a bootstrap calibration with recentering. A simulation study illustrates usefulness of the proposed inference methods.

1. Introduction

Single index models are widely used in statistics since they compromise interpretability of index coefficients in the parametric part and flexibility of regression modeling in the nonparametric part (see, ch. 8 of Li and Racine, 2007, for a review). Many estimation methods have been proposed for single index models, such as the semiparametric least squares estimator (Härdle, Hall and Ichimura, 1993; Ichimura, 1993), M-estimator (Klein and Spady, 1993), and average derivative estimator (Powell, Stock and Stoker, 1989). Although these estimation methods have desirable theoretical properties under certain regularity conditions, they typically require some nonparametric smoothing method to evaluate the unknown link function, which involves tuning parameters, such as bandwidth and series length parameters, and the optimal choices of them are substantial (theoretical and practical) problems.

The monotone single index model, in which monotonicity is imposed on the link function, has been studied in recent years. Balabdaoui, Durot and Jankowski (2016) showed that the least square estimator of a monotone single index model generally converges at the cube root rate, but its asymptotic distribution is still unknown. The main difficulty for deriving the asymptotic distribution of the least square estimator arises from the non-differentiability of the objective function; in a monotone single index model, the link function, which is an infinite-dimensional nuisance parameter, is generally estimated by a nonparametric approach such as isotonic regression, while the index part is parametrically modeled as a linear combination of the covariates. Then the derivative of the objective function with respect to the index coefficients is intractable due to the non-smoothness of the estimated nuisance parameter.

To overcome this issue, Groeneboom and Hendrickx (2018) developed a score-type estimator for the current status model, which is a special case of monotone single index models. Their approach is based on the estimating equation which is the same as the first-order condition of the least square estimator except that it ignores the derivative of the estimated link function.

They proved $\sqrt{n}$-consistency and asymptotic normality of their estimator without any tuning parameter. Their result was extended to general monotone single index models by Balabdaoui, Groeneboom and Hendrickx (2019), where they derived $\sqrt{n}$-consistency and asymptotic normality for the parametric component and an $n^{1/3}/\log n$ convergence rate for the nonparametric estimator of the link function.

Although the score estimation approach is remarkable, the main drawback is that it requires smoothing parameters to estimate the asymptotic variance to implement hypothesis testing and interval estimation. Because the estimating function in the score-type approach is dependent on the estimated link function, some conditional expectation is involved in the asymptotic variance. Besides, the partial derivative of the link function is also included in the asymptotic variance even though the estimated link function is not smooth. Therefore, smoothing methods, such as the kernel smoothing, are employed to estimate such quantities, which require us to select multiple smoothing parameters and make statistical inference cumbersome.

To address this problem, we propose an empirical likelihood inference method based on the score-type approach for monotone index models. We show that the empirical likelihood statistic based on the estimating equation of Balabdaoui, Groeneboom and Hendrickx (2019) converges in distribution to the weighted chi-squared distribution. Even in our empirical likelihood approach, the conditional expectation as mentioned above appears in the asymptotic distribution. To circumvent selection of smoothing parameters, we adapt the bootstrap calibration method proposed by Hjort, McKeague and van Keilegom (2009) to our context. Because of the estimating equation with the estimated nuisance parameter plugged-in, a classical naive bootstrap method is not asymptotically valid. Hjort, McKeague and van Keilegom (2009) provided a modified bootstrap method by recentering and reweighting to deal with such a situation. Combining the empirical likelihood and modified bootstrap methods, our approach provides a simple and theoretically justified method for statistical inference in monotone single index models.

The remainder of this paper is organized as follows. Section 2 presents our basic setup, methodology, and theoretical results. In Section 3, we conduct a small simulation study to illustrate the proposed method. All proofs are contained in the appendix.

2. Main result

We closely follow the setup and notation of Balabdaoui, Groeneboom and Hendrickx (2019) (hereafter BGH). Consider the monotone index model

$$Y = \psi_0(X'\alpha_0) + \epsilon, \qquad E[\epsilon|X] = 0, \tag{1}$$

where Y is a scalar response variable, X is a d-dimensional vector of covariates, ϵ is an error term, α_0 is a k-dimensional vector of parameters, and $\psi_0 : \mathbb{R} \to \mathbb{R}$ is an unknown monotone increasing function. For identification, we assume that α_0 belongs to the d-dimensional unit sphere $\mathcal{S}_{d-1} = \{\alpha \in \mathbb{R}^d : ||\alpha|| = 1\}$. We are interested in conducting statistical inference (i.e., interval estimation and hypothesis testing) on α_0 based on the empirical likelihood approach.

Let $\mathbb{S} : \mathbb{R}^{d-1} \to \mathcal{S}_{d-1}$ be a parameterization such that for each α in a neighborhood of α_0 on $\mathcal{S}_{d-1}$, there exists a unique $\beta \in \mathbb{R}^{d-1}$ which satisfies $\alpha = \mathbb{S}(\beta)$. To motivate the score-type

approach of BGH, we tentatively assume that ψ_0 is known. The population score equation for the least square estimation of β_0 is

$$E\left[\mathbb{J}(\beta_0)'X\psi_0^{(1)}(X'\mathbb{S}(\beta_0))\{Y-\psi_0(X'\mathbb{S}(\beta_0))\}\right]=0, \tag{2}$$

where $\psi_0^{(1)}$ is the derivative of ψ_0 and $\mathbb{J}(\beta)$ is the Jacobian of $\mathbb{S}(\beta)$. Thus, it is natural to construct an estimator of β_0 by taking an empirical counterpart of (2) and inserting estimators for $\psi_0^{(1)}$ and ψ_0. However, when we estimate ψ_0 by the isotonic regression method, the resulting estimator of ψ_0 is typically discontinuous and it is not clear how to evaluate the derivative $\psi_0^{(1)}$ without introducing smoothing parameters. To address this issue, BGH and Groeneboom and Hendrickx (2018) considered the modified population score equation

$$E\left[\mathbb{J}(\beta_0)'X\{Y-\psi_0(X'\mathbb{S}(\beta_0))\}\right]=0. \tag{3}$$

In particular, for point estimation of α_0, BGH proposed to solve the following score-type equation

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{J}(\hat{\beta})'X_i\{Y_i-\hat{\psi}_{\hat{\beta}}(X_i'\mathbb{S}(\hat{\beta}))\}=0, \tag{4}$$

with respect to $\hat{\beta}$, and estimate α_0 by $\hat{\alpha}=\mathbb{S}(\hat{\beta})$, where for given β, $\hat{\psi}_\beta$ is obtained by the isotonic regression

$$\hat{\psi}_\beta=\arg\min_{\psi\in\mathcal{M}}\sum_{i=1}^{n}\{Y_i-\psi(X_i'\mathbb{S}(\beta))\}^2, \tag{5}$$

and $\mathcal{M}$ is the set of monotone increasing functions defined on $\mathbb{R}$.

In this paper, we employ the score-type equation in (3) as a moment function and propose the following empirical likelihood statistic

$$\ell(\beta_0)=-2\max_{\{p_i\}_{i=1}^n}\sum_{i=1}^{n}\log(np_i)\qquad\text{s.t.}\ \sum_{i=1}^{n}p_i=1,\quad\sum_{i=1}^{n}p_i\hat{g}_i(\beta_0)=0, \tag{6}$$

where

$$\hat{g}_i(\beta)=\mathbb{J}(\beta)'X_i\{Y_i-\hat{\psi}_\beta(X_i'\mathbb{S}(\beta))\}.$$

By the Lagrange multiplier argument, its dual form is obtained as

$$\ell(\beta_0)=2\sum_{i=1}^{n}\log(1+\hat{\lambda}'\hat{g}_i(\beta_0)), \tag{7}$$

where the Lagrange multiplier $\hat{\lambda}$ solves

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{g}_i(\beta_0)}{1+\hat{\lambda}'\hat{g}_i(\beta_0)}=0. \tag{8}$$

In practice, we use the dual representation in (7) to implement statistical inference. To study the asymptotic properties of the empirical likelihood statistic $\ell(\beta_0)$, we impose the following assumptions. Let $\|\cdot\|$ be the Euclidean norm and $\mathcal{B}(a_0,A)=\{a:\|a-a_0\|\le A\}$ be a ball around a_0 of radius A.

Assumption.

A1: *$\{Y_i, X_i\}_{i=1}^n$ is an iid sample generated by (1). The support $\mathcal{X}$ of X is convex with a nonempty interior, and $\mathcal{X} \subset \mathcal{B}(0, R)$ for some $R > 0$. The Lebesgue density of X has a bounded derivative on $\mathcal{X}$. There exist positive constants c and C such that $E[|Y|^m|X = x] \leq cm!C^{m-2}$ for all integers $m \geq 2$ and almost every $x \in \mathcal{X}$.*

A2: *ψ_0 is monotone increasing and there exists $K_0 > 0$ such that $|\psi_0(u)| \leq K_0$ for all $u \in \{x'\alpha_0 : x \in \mathcal{X}\}$.*

These assumptions are adaptations of Assumptions A1-A6 in BGH. Compared to BGH, our assumptions are simpler because we do not need to control the behavior of the score function outside the true parameter $\alpha_0 = \mathbb{S}(\beta_0)$. Assumption A1 is on the distribution form of the data. The support condition in A1 may be relaxed by assuming X to follow a sub-Gaussian distribution. The moment condition in A1, which is analogous to BGH's A6, is required to guarantee $\max_{1\leq i\leq n} |Y_i| = O_p(\log n)$ to control the entropy of a class of score functions. Assumption A2 is on the true link function ψ_0. Compared to BGH which considers point estimation, we only need to impose boundedness, which is a mild requirement.

Under these assumptions, our main result is presented as follows.

Theorem 1. *Under Assumptions A1-A2, it holds*

$$\ell(\beta_0) \xrightarrow{d} Z'VZ,$$

where $Z \sim N(0, \Sigma)$ with $\Sigma = \mathbb{J}(\beta_0)'E[\epsilon^2(X - E[X|X'\mathbb{S}(\beta_0)])(X - E[X|X'\mathbb{S}(\beta_0)])']\mathbb{J}(\beta_0)$ and $V = \mathbb{J}(\beta_0)'E[\epsilon^2 XX']\mathbb{J}(\beta_0)$.

Remark 1. This theorem says that the empirical likelihood statistic $\ell(\beta_0)$ is not asymptotically pivotal and converges to a weighted chi-squared distribution $w_1\chi^2_{1,1} + \cdots + w_{d-1}\chi^2_{1,d-1}$, where $w_1, \ldots, w_{d-1}$ are the eigenvalues of $\Sigma^{-1}V$ and $\chi^2_{1,1}, \ldots, \chi^2_{1,d-1}$ are independent χ^2_1 random variables. This lack of asymptotic pivotalness is caused by the mismatch in the asymptotic variance Σ of the score function $\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{g}_i(\beta_0)$ and the limit V of the sample variance $\hat{V} = \frac{1}{n}\sum_{i=1}^n \hat{g}_i(\beta_0)\hat{g}_i(\beta_0)'$. In the literature of empirical likelihood, weighted chi-squared limiting distributions often emerge when the score (or moment) functions involve estimated nuisance parameters (e.g., Qin and Jing, 2001; Xue and Zhu, 2006; Hjort, McKeague, and van Keilegom, 2009).

Remark 2. One way to conduct statistical inference based on $\ell(\beta_0)$ is to estimate the critical values of $w_1\chi^2_{1,1} + \cdots + w_{d-1}\chi^2_{1,d-1}$ based on some estimators of Σ and V. Based on (13), V is consistently estimated by $\hat{V}$. On the other hand, Σ can be estimated by

$$\hat{\Sigma} = \mathbb{J}(\beta_0)'\frac{1}{n}\sum_{i=1}^n \hat{\epsilon}_i^2\{X_i - \hat{m}(X_i'\mathbb{S}(\beta_0))\}\{X_i - \hat{m}(X_i'\mathbb{S}(\beta_0))\}\mathbb{J}(\beta_0),$$

where $\hat{\epsilon}_i = Y_i - \hat{\psi}_{\beta_0}(X_i'\mathbb{S}(\beta_0))$ and $\hat{m}(\cdot)$ is a nonparametric estimator of $m(\cdot) = E[X|X'\mathbb{S}(\beta_0) = \cdot]$. An alternative way for statistical inference is to adjust the empirical likelihood statistic $\ell(\beta_0)$ to recover the asymptotic pivotalness. Based on Rao and Scott (1981) (see also Xue and Zhu,

2006), the above theorem implies

$$\ell_A(\beta_0) = \frac{d-1}{\text{trace}(\hat{\Sigma}^{-1}\hat{V})}\ell(\beta_0) \xrightarrow{d} \chi^2_{d-1}. \tag{9}$$

Then the confidence region of $\alpha_0 = \mathbb{S}(\beta_0)$ can be obtained by $\{\mathbb{S}(\beta) : \ell_A(\beta) \leq q_a\}$, where q_a is the $(1-a)$-th quantile of the χ^2_{d-1} distribution.

Remark 3. A drawback of the asymptotic inference method presented in the previous remark is that it requires a selection of a tuning parameter to implement the nonparametric estimator $\hat{m}(\cdot)$. In order to obtain an inference procedure which is free from tuning parameters, we adapt the bootstrap method of Hjort, McKeague, and van Keilegom (2009) as follows.

(1) Based on the original sample $\{Y_i, X_i\}_{i=1}^n$, compute $\hat{\beta}$ as in (4), and then compute

$$M_n(\hat{\beta}) = \frac{1}{n}\sum_{i=1}^n \hat{g}_i(\mathbb{S}(\hat{\beta})), \qquad \bar{V} = \frac{1}{n}\sum_{i=1}^n \hat{g}_i(\mathbb{S}(\hat{\beta}))\hat{g}_i(\mathbb{S}(\hat{\beta}))'.$$

(2) Draw $\{Y_i^*, X_i^*\}_{i=1}^n$ from the original sample $\{Y_i, X_i\}_{i=1}^n$ with equal weights. Then compute

$$M_n^*(\hat{\beta}) = \frac{1}{n}\sum_{i=1}^n \mathbb{J}(\hat{\beta})' X_i^* \{Y_i^* - \hat{\psi}^*_{\hat{\beta}}(X_i^{*\prime}\mathbb{S}(\hat{\beta}))\},$$

where $\hat{\psi}^*_{\hat{\beta}} = \arg\min_{\psi\in\mathcal{M}} \sum_{i=1}^n \{Y_i^* - \psi(X_i^{*\prime}\mathbb{S}(\hat{\beta}))\}^2$.

(3) The bootstrap counterpart of $\ell(\beta_0)$ is given by

$$\ell^* = n\{M_n^*(\hat{\beta}) - M_n(\hat{\beta})\}'\bar{V}^{-1}\{M_n^*(\hat{\beta}) - M_n(\hat{\beta})\}. \tag{10}$$

Under the additional assumptions A3-A5 in the appendix, the validity of this bootstrap approximation is obtained as follows.

Theorem 2. *Under Assumptions A1-A5, it holds*

$$\sup_{t\geq 0} |P^*\{\ell^* \leq t\} - P_0\{\ell(\beta_0) \leq t\}| \xrightarrow{p} 0,$$

where P^* is the bootstrap distribution conditional on the data.

3. Simulation

We conduct a simulation study to investigate the finite sample performance of the proposed inference methods. We consider the following data generation process:

$$\begin{aligned} Y &= \psi_0(X'\alpha_0) + \epsilon, \quad \psi_0(u) = u^3, \quad \alpha_0 = (1,1,1)'/\sqrt{3} \\ \epsilon &\sim N(0,1), \quad X \sim N(0, I_3), \end{aligned}$$

where I_3 is the 3×3 identity matrix. We consider sample sizes $n = 100, 500, 1000$. The number of Monte Carlo replications is 1000. We consider two testing methods discussed in Remarks 2 and 3. For the adjusted statistic in (9), we estimate $m(\cdot) = E[X|X'\mathbb{S}(\beta_0) = \cdot]$ by the Nadaraya-Watson estimator, and choose the bandwidths based on the expected Kullback-Leibler cross-validation (Hurvich, Simonoff and Tsai, 1998). To test the null hypothesis $H_0 : \alpha_0 = (1,1,1)'/\sqrt{3}$, we calculate the test statistic (9) and compare it with the 95 percentile of the χ^2_{d-1} distribution.

TABLE 1. Rejection frequencies (%)

n	Adjusted				Bootstrap				$\hat{\alpha}_1$	
	N	A1	A2	A3	N	A1	A2	A3	mean	s.d.
100	4.7	4.9	6.1	8.7	8.1	8.3	9.0	13.9	0.577	0.0528
500	4.2	7.5	15.9	51.1	6.6	10.0	18.1	53.3	0.576	0.0166
1000	7.4	14.8	31.5	86.1	5.6	18.2	34.9	87.8	0.577	0.0113

For the bootstrap-calibrated test statistic (10), we compute $\hat{\beta}$ as in BGH (the computer code is available at Groeneboom's website), and generate 499 bootstrap samples, and calculate the bootstrap counterpart ℓ^* in (10).

Table 1 presents the rejection frequencies of the above empirical likelihood tests for the null $H_0 : \alpha_0 = (1,1,1)'/\sqrt{3}$ when the true values of α_0 are (N) $\alpha_0 = (1,1,1)'/\sqrt{3}$, (A1) $\alpha_0 = (1.03,1,1)'/\sqrt{1.03^2+2}$, (A2) $\alpha_0 = (1.05,1,1)'/\sqrt{1.05^2+2}$, and (A3) $\alpha_0 = (1.10,1,1)'/\sqrt{1.10^2+2}$. (N) is for the size properties, and (A1)-(A3) are to evaluate power properties.

The column "$\hat{\alpha}_1$" reports the Monte Carlos averages and standard deviations of the first element of the BGH estimator $\hat{\alpha}$. It shows that the mean is close to the truth, $\alpha_{01} = 1/\sqrt{3} \simeq 0.577$, while the standard deviation becomes smaller with the sample size. From the columns (N), we can see that both the adjusted and bootstrap empirical likelihood tests have reasonable size properties. Both tests become powerful as the sample size increases and the true values of α_0 are more distinct from the null values (i.e., from A1 to A3). Also, we find that overall the bootstrap test rejects slightly more often than the adjusted test.

Overall, our simulation results are encouraging.

Appendix A. Mathematical appendix

A.1. **Proof of theorem 1.** Here we denote $\hat{g}_{0i} = \hat{g}_i(\beta_0)$, $\mathbb{S}_0 = \mathbb{S}(\beta_0)$, and $\mathbb{J}_0 = \mathbb{J}(\beta_0)$.

Note that (i) X has bounded support (by Assumption A1), (ii) $\max|Y_i| = O_p(\log n)$ (by Assumption A2 and Lemma 7.1 of Balabdaoui, Durot and Jankowski, 2016), and (iii) $\sup_{x\in\mathcal{X}} |\hat{\psi}_{\beta_0}(x'\mathbb{S}_0)| = O_p(\log n)$ by Lemma 8 of the supplementary material of BGH (hereafter BGH-supp). Combining these results, it holds

$$\max_{1\le i\le n} |\hat{g}_{0i}| = O_p(\log n). \tag{11}$$

Thus, an expansion of (8) around $\hat{\lambda} = 0$ using the same argument in Owen (1991, proof of Theorem 2) based on (11) implies

$$\hat{\lambda} = \left[\frac{1}{n}\sum_{i=1}^n \hat{g}_{0i}\hat{g}_{0i}{}'\right]^{-1} \frac{1}{n}\sum_{i=1}^n \hat{g}_{0i} + o_p(n^{-1/2}). \tag{12}$$

A second-order expansion of (7) around $\hat{\lambda} = 0$ using (12) yields

$$\begin{aligned}
\ell(\beta_0) &= 2\hat{\lambda}'\sum_{i=1}^n \hat{g}_{0i} - \hat{\lambda}'\left[\sum_{i=1}^n \hat{g}_{0i}\hat{g}_{0i}'\right]\hat{\lambda} + o_p(1) \\
&= \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{g}_{0i}\right)'\left[\frac{1}{n}\sum_{i=1}^n \hat{g}_{0i}\hat{g}_{0i}'\right]^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{g}_{0i}\right) + o_p(1).
\end{aligned}$$

Then it is enough for the conclusion to show that

$$\frac{1}{n}\sum_{i=1}^n \hat{g}_{0i}\hat{g}_{0i}' \xrightarrow{p} V = \mathbb{J}_0' E[\epsilon^2 XX']\mathbb{J}_0, \tag{13}$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{g}_{0i} \xrightarrow{d} N(0,\Sigma). \tag{14}$$

We first show (13). Decompose

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n \hat{g}_{0i}\hat{g}_{0i}' &= \mathbb{J}_0'\left[\frac{1}{n}\sum_{i=1}^n \epsilon_i^2 X_iX_i'\right]\mathbb{J}_0 + \mathbb{J}_0'\left[\frac{1}{n}\sum_{i=1}^n \{\psi_0(X_i'\mathbb{S}_0) - \hat{\psi}_{\beta_0}(X_i'\mathbb{S}_0)\}^2 X_iX_i'\right]\mathbb{J}_0 \\
&\quad + \mathbb{J}_0'\left[\frac{2}{n}\sum_{i=1}^n \epsilon_i\{\psi_0(X_i'\mathbb{S}_0) - \hat{\psi}_{\beta_0}(X_i'\mathbb{S}_0)\}X_iX_i'\right]\mathbb{J}_0.
\end{aligned} \tag{15}$$

By the law of large numbers, the first term of (15) converges to V. By Mukerjee (1988, Theorem 3.3), we have the uniform consistency

$$\sup_{x\in\mathcal{X}} |\psi_0(x'\mathbb{S}_0) - \hat{\psi}_{\beta_0}(x'\mathbb{S}_0)| \xrightarrow{p} 0. \tag{16}$$

Therefore, both the second and third terms of (15) converge to zero. Combining these results, we obtain (13).

We now show (14). Let $\mathbb{P}_n$ be the empirical measure of $\{X_i, Y_i\}_{i=1}^n$, P_0 be the true measure of (X, Y), and

$$E[X|x'\mathbb{S}_0] = E[X|X'\mathbb{S}_0 = u] \text{ evaluated at } u = x'\mathbb{S}_0.$$

Decompose

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\hat{g}_{0i} &= \mathbb{J}_0'\int x\{y-\hat{\psi}_{\beta_0}(x'\mathbb{S}_0)\}d\mathbb{P}_n(y,x)\\ &= \mathbb{J}_0'\int\{x-E[X|x'\mathbb{S}_0]\}\{y-\hat{\psi}_{\beta_0}(x'\mathbb{S}_0)\}d\mathbb{P}_n(y,x)\\ &\quad+\mathbb{J}_0'\int\{E[X|x'\mathbb{S}_0]-\bar{E}_n(x'\mathbb{S}_0)\}\{y-\hat{\psi}_{\beta_0}(x'\mathbb{S}_0)\}d\mathbb{P}_n(y,x)\\ &\quad+\mathbb{J}_0'\int\bar{E}_n(x'\mathbb{S}_0)\{y-\hat{\psi}_{\beta_0}(x'\mathbb{S}_0)\}d\mathbb{P}_n(y,x)\\ &:= \mathbb{J}_0'(I+II+III),\end{aligned}$$

where

$$\bar{E}_n(u)=\begin{cases}E[X|x'\mathbb{S}_0=\tau_{i,\mathbb{S}_0}] & \text{if } \psi_0(u)>\hat{\psi}_{\beta_0}(u) \text{ for all } u\in(\tau_i,\tau_{i+1}),\\ E[X|x'\mathbb{S}_0=s] & \text{if } \psi_0(s)=\hat{\psi}_{\beta_0}(s) \text{ for some } s\in(\tau_i,\tau_{i+1}),\\ E[X|x'\mathbb{S}_0=\tau_{i+1,\mathbb{S}_0}] & \text{if } \psi_0(u)<\hat{\psi}_{\beta_0}(u) \text{ for all } u\in(\tau_i,\tau_{i+1}),\end{cases} \tag{17}$$

and $\tau_{i,\mathbb{S}_0}$ is the sequence of jump points of $\hat{\psi}_{\beta_0}$. By the definition of $\bar{E}_n(x'\mathbb{S}_0)$, it holds $III=0$ (see, (C.10) in BGH-supp).

For II, decompose

$$\begin{aligned}II &= \int\{E[X|x'\mathbb{S}_0]-\bar{E}_n(x'\mathbb{S}_0)\}\{y-\hat{\psi}_{\beta_0}(x'\mathbb{S}_0)\}d(\mathbb{P}_n-P_0)(y,x)\\ &\quad+\int\{E[X|x'\mathbb{S}_0]-\bar{E}_n(x'\mathbb{S}_0)\}\{y-\psi_{\beta_0}(x'\mathbb{S}_0)\}dP_0(y,x)\\ &\quad+\int\{E[X|x'\mathbb{S}_0]-\bar{E}_n(x'\mathbb{S}_0)\}\{\hat{\psi}_{\beta_0}(x'\mathbb{S}_0)-\psi_0(x'\mathbb{S}_0)\}dP_0(y,x)\\ &:= II_a+II_b+II_c.\end{aligned} \tag{18}$$

The same argument as in pp. 19-20 of BGH-supp guarantees $II_a=o_p(n^{-1/2})$ and $II_b=o_p(n^{-1/2})$. For II_c, using (C.11) of BGH-supp and Proposition 4 of BGH, we have

$$\begin{aligned}\|II_c\| &\le C\int\{\hat{\psi}_{\beta_0}(x'\mathbb{S}_0)-\psi_0(x'\mathbb{S}_0)\}^2dP_0(y,x)\\ &= O_p((\log n)^2n^{-2/3})=o_p(n^{-1/2}),\end{aligned}$$

for some $C>0$. Therefore, we obtain

$$II=o_p(n^{-1/2}). \tag{19}$$

For I, decompose

$$\begin{aligned}I &= \int\{x-E[X|x'\mathbb{S}_0]\}\{y-\psi_0(x'\mathbb{S}_0)\}d\mathbb{P}_n(y,x)\\ &\quad+\int\{x-E[X|x'\mathbb{S}_0]\}\{\psi_0(x'\mathbb{S}_0)-\hat{\psi}_{\beta_0}(x'\mathbb{S}_0)\}d\mathbb{P}_n(y,x)\\ &:= I_a+I_b.\end{aligned}$$

From pp. 21-22 of BGH-supp, we can show that $I_b = o_p(n^{-1/2})$. Therefore,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\hat{g}_{0i} &= \mathbb{J}_0'\int\{x - E[X|x'\mathbb{S}_0]\}\{y - \psi_0(x'\mathbb{S}_0)\}d\mathbb{P}_n(y,x) + o_p(n^{-1/2}) \\ &= \mathbb{J}_0'\frac{1}{n}\sum_{i=1}^{n}\{X_i - E[X_i|X_i'\mathbb{S}_0]\}\epsilon_i + o_p(n^{-1/2}), \end{aligned} \tag{20}$$

and the central limit theorem implies (14). Therefore, the conclusion is obtained.

A.2. **Proof of Theorem 2.** Based on Hjort, McKeague and van Keilegom (2009), it is sufficient for the conclusion to show that

$$\bar{V} \xrightarrow{P_0} \mathbb{J}_0'E[\epsilon^2 XX']\mathbb{J}_0, \tag{21}$$

$$\sqrt{n}\{M_n^*(\hat{\beta}) - M_n(\hat{\beta})\} \xrightarrow{d} N(0,\Sigma), \tag{22}$$

where $\hat{\beta}$ is obtained by solving (4). For the validity of bootstrap, we add the following assumptions.

A3: There exists $\delta_0 > 0$ such that the mapping $u \mapsto E[Y|X'\alpha = u]$ is monotone increasing on $I_\alpha = \{z'\alpha, z \in \mathcal{Z}\}$ for each $\alpha \in \mathcal{B}(\alpha_0, \delta_0)$.

A4: For all $\beta \neq \beta_0$ with $\mathbb{S}(\beta) \in \mathcal{B}(\alpha_0, \delta_0)$, $Cov\big[(\beta_0 - \beta)'\mathbb{J}(\beta)'X, \psi_0(\mathbb{S}(\beta_0)'X)|\mathbb{S}(\beta)'X\big] \neq 0$ almost surely.

A5: $\mathbb{J}_0'E[\psi_0^{(1)}(X'\alpha_0)Var(X|X'\alpha_0)]\mathbb{J}_0$ is non-singular.

By BGH, it can be shown that under A1-A5, $\hat{\beta}$ is consistent and $\sqrt{n}(\hat{\beta} - \beta_0)$ is asymptotically normal. Let $\psi_\beta(u) = E[Y|X'\mathbb{S}(\beta) = u]$. For (21), note that

$$\begin{aligned}\bar{V} &= \mathbb{J}(\hat{\beta})'\left[\frac{1}{n}\sum_{i=1}^{n}X_i\{\epsilon_i + \psi_{\hat{\beta}}(X_i'\mathbb{S}(\hat{\beta})) - \hat{\psi}_{\hat{\beta}}(X_i'\mathbb{S}(\hat{\beta}))\}^2X_i'\right]\mathbb{J}(\hat{\beta}) + o_p(1) \\ &= \{\mathbb{J}_0 + o_p(1)\}'\left\{\frac{1}{n}\sum_{i=1}^{n}\epsilon_i^2X_iX_i' + o_p(1)\right\}\{\mathbb{J}_0 + o_p(1)\},\end{aligned}$$

where the first equality follows from $\psi_0(x'\mathbb{S}(\beta_0)) - \psi_{\hat{\beta}}(x'\mathbb{S}(\hat{\beta})) = O_p(\hat{\beta} - \beta_0)$ for almost every x (by p. 26 and Lemma 17 of BGH-supp) and the consistency of $\hat{\beta}$, and the second equality follows from (16), A3, and the consistency of $\hat{\beta}$. Thus, by the law of large numbers, we obtain (21).

We now prove (22). Note that $M_n^*(\hat{\beta}) - M_n(\hat{\beta}) = M_n^*(\hat{\beta})$ by (4). Let $\hat{\mathbb{P}}_n$ be the empirical measure of the bootstrap resample. Decompose

$$\begin{aligned}M_n^*(\hat{\beta}) &= \mathbb{J}(\hat{\beta})'\int\{x - E(X|x'\mathbb{S}(\hat{\beta}))\}\{y - \hat{\psi}_{\hat{\beta}}^*(x'\mathbb{S}(\hat{\beta}))\}d\hat{\mathbb{P}}_n \\ &\quad + \mathbb{J}(\hat{\beta})'\int\{E(X|x'\mathbb{S}(\hat{\beta})) - \bar{E}_n^*(x'\mathbb{S}(\hat{\beta}))\}\{y - \hat{\psi}_{\hat{\beta}}^*(x'\mathbb{S}(\hat{\beta}))\}d\hat{\mathbb{P}}_n \\ &\quad + \mathbb{J}(\hat{\beta})'\int\bar{E}_n^*(x'\mathbb{S}(\hat{\beta}))\{y - \hat{\psi}_{\hat{\beta}}^*(x'\mathbb{S}(\hat{\beta}))\}d\hat{\mathbb{P}}_n \\ &:= I^* + II^* + III^*,\end{aligned} \tag{23}$$

where $\bar{E}_n^*(\cdot)$ is defined similarly to (17) with respect to $\hat{\psi}_{\hat{\beta}}^*$. Again, we have $III^* = 0$ by the definition of $\bar{E}_n^*(\cdot)$. For II^*, similar to (19) and p. 3481 of Groeneboom and Hendrickx (2017) (GH hereafter), we have $II^* = o_{P_M}(n^{-1/2})$, where P_M is defined in p. 3450 of GH.

For I^*, decompose

$$\begin{aligned} I^* &= \mathbb{J}(\hat{\beta})' \int \{x - E(X|x'\mathbb{S}(\hat{\beta}))\}\{y - \hat{\psi}_{\hat{\beta}}^*(x'\mathbb{S}(\hat{\beta}))\} d(\hat{\mathbb{P}}_n - \mathbb{P}_n) \\ &\quad + \mathbb{J}(\hat{\beta})' \int \{x - E(X|x'\mathbb{S}(\hat{\beta}))\}\{y - \hat{\psi}_{\hat{\beta}}(x'\mathbb{S}(\hat{\beta}))\} d\mathbb{P}_n \\ &\quad + \mathbb{J}(\hat{\beta})' \int \{x - E(X|x'\mathbb{S}(\hat{\beta}))\}\{\hat{\psi}_{\hat{\beta}}(x'\mathbb{S}(\hat{\beta})) - \hat{\psi}_{\hat{\beta}}^*(x'\mathbb{S}(\hat{\beta}))\} d\mathbb{P}_n \\ &:= I_a^* + I_b^* + I_c^*. \end{aligned}$$

For I_b^*, (4) and pp. 19-20 of BGH-supp combined with $\hat{\beta} - \beta_0 = O_p(n^{-1/2})$ imply

$$\begin{aligned} I_b^* &= \mathbb{J}(\hat{\beta})' \int x\{y - \hat{\psi}_{\hat{\beta}}(x'\mathbb{S}(\hat{\beta}))\} d\mathbb{P}_n \\ &\quad - \mathbb{J}(\hat{\beta})' \int \{E(X|x'\mathbb{S}(\hat{\beta})) - \bar{E}_n(x'\mathbb{S}(\hat{\beta}))\}\{y - \hat{\psi}_{\hat{\beta}}(x'\mathbb{S}(\hat{\beta}))\} d\mathbb{P}_n \\ &= o_p(n^{-1/2}). \end{aligned}$$

For I_c^*, (6.21) in GH and pp. 21-22 of BGH-supp yield

$$\begin{aligned} I_c^* &= \mathbb{J}(\hat{\beta})' \int \{x - E(X|x'\mathbb{S}(\hat{\beta}))\}\{\psi_{\hat{\beta}}(x'\mathbb{S}(\hat{\beta})) - \hat{\psi}_{\hat{\beta}}^*(x'\mathbb{S}(\hat{\beta}))\} d\mathbb{P}_n \\ &\quad + \mathbb{J}(\hat{\beta})' \int \{x - E(X|x'\mathbb{S}(\hat{\beta}))\}\{\hat{\psi}_{\hat{\beta}}(x'\mathbb{S}(\hat{\beta})) - \psi_{\hat{\beta}}(x'\mathbb{S}(\hat{\beta}))\} d\mathbb{P}_n \\ &= o_p(n^{-1/2}). \end{aligned}$$

Finally, for I_a^*, we have

$$\begin{aligned} I_a^* &= \mathbb{J}(\hat{\beta})' \int \{x - E(X|x'\mathbb{S}_0)\}\{y - \psi_0(x'\mathbb{S}_0)\} d(\hat{\mathbb{P}}_n - \mathbb{P}_n) + o_{P_M}(n^{-1/2} + (\hat{\beta} - \beta_0)) \\ &= \mathbb{J}(\hat{\beta})' \int \{x - E(X|x'\mathbb{S}_0)\}\epsilon d(\hat{\mathbb{P}}_n - \mathbb{P}_n) + o_{P_M}(n^{-1/2}), \end{aligned}$$

where the first equality follows from a similar argument to (6.25) in GH, and the second equality follows from a rearrangement and $\hat{\beta} - \beta_0 = O_p(n^{-1/2})$.

Combining these results, we have

$$M_n^*(\hat{\beta}) - M_n(\hat{\beta}) = \mathbb{J}(\hat{\beta})' \int \{x - E(X|x'\mathbb{S}_0)\}\epsilon d(\hat{\mathbb{P}}_n - \mathbb{P}_n) + o_{P_M}(n^{-1/2}).$$

Comparing this and (20), the central limit theorem yields (22). Therefore, the conclusion follows.

References

[1] Balabdaoui, F., Durot, C. and H. Jankowski (2016) Least squares estimation in the monotone single index model, arXiv:1610.06026.

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

[3] Groeneboom, P. and K. Hendrickx (2017) The nonparametric bootstrap for the current status model, *Electronic Journal of Statistics*, 11, 3446-3484.

[4] Groeneboom, P. and K. Hendrickx (2018) Current status linear regression, *Annals of Statistics*, 46, 1415-1444.

[5] Härdle, W., Hall, P. and H. Ichimura (1993) Optimal smoothing in single-index models, *Annals of Statistics*, 21, 157-178.

[6] Hjort, N. L., McKeague, I. W. and I. van Keilegom (2009) Extending the scope of empirical likelihood, *Annals of Statistics*, 37, 1079-1111.

[7] Hurvich, C. M., Simonoff, J. S. and C.-L. Tsai (1998) Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion, *Journal of the Royal Statistical Society*, B, 60, 271-293.

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

[9] Klein, R. W. and R. H. Spady (1993) An efficient semiparametric estimator for binary response models, *Econometrica*, 61, 387-421.

[10] Li, Q. and J. S. Racine (2007) *Nonparametric Econometrics*, Princeton University Press.

[11] Mukerjee, H. (1988) Monotone nonparametric regression, *Annals of Statistics*, 16, 741-750.

[12] Owen, A. (1991) Empirical likelihood for linear models, *Annals of Statistics*, 19, 1725-1747.

[13] Powell, J. L., Stock, J. H. and T. M. Stoker (1989) Semiparametric estimation of index coefficients, *Econometrica*, 57, 1403-1430.

[14] Qin, G. S. and B. Y. Jing (2001) Censored partial linear models and empirical likelihood, *Journal of Multivariate Analysis*, 78, 37-61.

[15] Rao, J. N. K. and A. J. Scott (1981) The analysis of categorical data from complex sample surveys: chi-squared tests for goodness of fit and independence in two-way tables, *Journal of the American Statistical Association*, 76, 221-230.

[16] Xue, L.-G. and L. Zhu (2006) Empirical likelihood for single-index models, *Journal of Multivariate Analysis*, 97, 1295-1312.

Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK, and Keio Economic Observatory (KEO), 2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan.

E-mail address: t.otsu@lse.ac.uk

Faculty of Economics, Keio University, 2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan.

E-mail address: kei_t_12@keio.jp

Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK.

E-mail address: M.Xu8@lse.ac.uk