Score estimation of monotone partially linear index model

Mengshan Xu[*]and Taisuke Otsu[†]

London School of Economics

February 25, 2020

Abstract

This paper studies semiparametric estimation of a partially linear single index model with a monotone link function. Our estimator is an extension of the score-type estimator developed by Balabdaoui, Groeneboom and Hendrickx (2019) for the monotone single index model, which profiles out the unknown link function by isotonic regression. An attractive feature of the proposed estimator is that it is free from tuning parameters for nonparametric smoothing. We show that our estimator for the finite-dimensional components is $\sqrt{n}$-consistent and asymptotically normal. By introducing an additional smoothing to obtain the efficient score, we propose an asymptotically efficient estimator for the finite-dimensional components. Furthermore, we establish the asymptotic validity of a bootstrap inference method based the score-type estimator, which is also free from tuning parameters. A simulation study illustrates usefulness of the proposed method.

1 Introduction

This paper is concerned with the monotone partially linear single index (PLSI) model

$$Y = X'\beta_0 + \psi_0(Z'\alpha_0) + \epsilon, \qquad E[\epsilon|X, Z] = 0, \tag{1}$$

where $Y \in \mathbb{R}$ is a response variable, $X \in \mathcal{X} \subseteq \mathbb{R}^k$ and $Z \in \mathcal{Z} \subseteq \mathbb{R}^d$ are covariates, $\epsilon \in \mathbb{R}$ is an error term, α_0 and β_0 are finite dimensional parameters, and $\psi_0 : \mathbb{R} \to \mathbb{R}$ is an unknown monotone increasing function. For identification, we assume that Z does not contain a constant and α_0 belongs to the d-dimensional unit sphere $\mathcal{S}_{d-1} = \{\alpha \in \mathbb{R}^d : ||\alpha|| = 1\}$.

Since a seminal work by Carroll *et al.* (1997), the model (1) (without the monotonicity assumption about ψ_0) has been studied by many authors, including Xia, Tong and Li (1999),

Yu and Ruppert (2002), Xia and Härdle (2006), Wang *et al.* (2010), and Ma and Zhu (2013), among others. The model (1) is very flexible. If α_0 is known, it becomes a partially linear model. If $\beta_0 = 0$, it becomes a single index model. See, e.g., Wang *et al.* (2010) for a review on these models. Estimation of the model (1) typically requires some nonparametric smoothing method to evaluate the unknown function ψ_0, which involves tuning parameters, such as bandwidth and series length parameters.

In this paper, we consider the situation where ψ_0 is known to be monotone. Instead of assuming certain degree of smoothness as in the above cited papers, we impose a shape restriction on ψ_0, and propose a $\sqrt{n}$-consistent estimator for the parameters (α_0, β_0) that is free from tuning parameters. Furthermore, we establish the asymptotic validity of a bootstrap inference method based the proposed estimator, which is also free from tuning parameters.

A natural approach to incorporate monotonicity into nonparametric estimation is to employ the isotonic regression technique (see, e.g., Groeneboom and Jongbloed, 2014, for a review). For example, one may consider the least square estimation for the model (1), say $\min_{\alpha,\beta}[\min_{\psi \in \mathcal{M}} \sum_{i=1}^{n} \{Y_i - X_i'\beta + \psi(Z_i'\alpha)\}^2]$, where $\mathcal{M}$ the set of monotone increasing functions. In this case, we can apply the isotonic regression technique for each (α, β), and then minimize the concentrated criterion function with respect to (α, β). However, because of lack of smoothness of the isotonic regression estimator for ψ_0, it is not clear whether such a profile least square estimator for (α_0, β_0) will be $\sqrt{n}$-consistent or asymptotically normal. This point was clarified by Balabdaoui, Groeneboom and Hendrickx (2019) (BGH hereafter) and Groeneboom and Hendrickx (2018) for single index (and current status) models.

For this problem, BGH and Groeneboom and Hendrickx (2018) developed a novel score estimation approach for single index models, say $Y = \psi_0(Z'\alpha_0) + \epsilon$. Their basic idea is to construct a feasible score equation $\sum_{i=1}^{n} Z_i\{Y_i - \psi_\alpha(Z_i'\alpha)\} = 0$ where ψ_α is estimated by isotonic regression for given α. Then the estimator for α_0 is obtained by the solution of the feasible score equation. BGH showed that their score estimator for α_0 is $\sqrt{n}$-consistent and asymptotically normal. Furthermore, BGH proposed an asymptotically efficient estimator for α_0 by evaluating an optimal score equation. Groeneboom and Hendrickx (2018) and Groeneboom and Hendrickx (2017) studied the score-type estimator for current status models and its bootstrap validity, respectively.

In this paper, we extend the score estimation approach developed by BGH and Groeneboom and Hendrickx (2018) to the monotone PLSI model in (1). We show that the proposed score-type estimator for (α_0, β_0) is $\sqrt{n}$-consistent and asymptotically normal. Also, by estimating nonparametrically the efficient score function, we derive an asymptotically efficient estimator for (α_0, β_0) whose asymptotic variance coincides with the efficient variance matrix in Carroll *et al.* (1997). Finally, we establish the validity of a bootstrap inference method based on the score-type estimator. Similar to the existing papers on (not necessarily monotone) PLSI models cited above, the extension from single index or current status models to the PLSI model is not a trivial task. In particular, the presence of linear indices both inside and outside the nonparametric monotone

function complicates the theoretical development.

This paper complements the literature on score-type estimation for semiparametric models with isotonic nuisance parameter estimates. Groeneboom and Hendrickx (2018) and BGH argued that score-type estimation and monotone least square estimation are not equivalent methods; they showed theoretically and numerically that the score-type estimator behaves at least as good as (or even better than) the monotone least square in single index models. The present paper shows analogous advantages continue to hold in PLSI models. Huang (2002), Cheng (2009), and Yu (2014) studied asymptotic properties of the monotone least square estimator, but it was unclear whether the score-type estimator could also achieve the $\sqrt{n}$-convergence rate and semiparametric efficiency. Our paper fills this gap.

Furthermore, the results in this paper can be considered as extensions of the ones for monotone partially linear models (Huang, 2002, and Cheng, 2009). However, since the partially linear model does not involve unknown parameters (i.e., α_0) in the argument of the unknown function ψ_0, the theoretical development is very different from ours.

This paper is organized as follows. In Section 2, we introduce our score-type estimator for the model (1) and present its asymptotic properties. We also propose an asymptotically efficient estimator for (α_0, β_0) and bootstrap inference method. Section 3 presents some simulation evidence to illustrate the finite sample performance of our estimators and bootstrap method.

2 Main results

2.1 Estimation method

Let us first introduce our estimator for the PLSI model in (1). In particular, we extend the score estimation approach by BGH to estimate the parameters (α_0, β_0) in (1). Consider a parameterization $\mathbb{S}$ from a subset of $\mathbb{R}^{d-1}$ to $\mathcal{S}_{d-1}$ such that for each α in a neighborhood of α_0 on $\mathcal{S}_{d-1}$, there exists a unique $\gamma \in \mathbb{R}^{d-1}$ satisfying $\alpha = \mathbb{S}(\gamma)$.[1] Then the reparameterized model (1) is written as

$$Y = X'\beta_0 + \psi_0(Z'\mathbb{S}(\gamma_0)) + \epsilon, \qquad E[\epsilon|X, Z] = 0. \tag{2}$$

To motivate our estimation approach, we tentatively assume that ψ_0 is known. In this case, the population score equation for $\theta_0 = (\beta_0', \gamma_0')'$ is

$$E\left[\begin{pmatrix} X \\ \mathbb{J}(\gamma_0)'Z\psi_0'(Z'\mathbb{S}(\gamma_0)) \end{pmatrix}\{Y - X'\beta_0 - \psi_0(Z'\mathbb{S}(\gamma_0))\}\right] = 0, \tag{3}$$

---

[1]Examples of such parametrization are the spherical coordinate system $\mathbb{S} : [0, \pi]^{d-2} \times [0, 2\pi] \to \mathcal{S}_{d-1}$ with

$$\mathbb{S}(\gamma) = (\cos(\gamma_1), \sin(\gamma_1)\cos(\gamma_2), \sin(\gamma_1)\sin(\gamma_2)\cos(\gamma_3), \dots, \sin(\gamma_1)\cdots\sin(\gamma_{d-2})\cos(\gamma_{d-1}), \sin(\gamma_1)\cdots\sin(\gamma_{d-2})\sin(\gamma_{1-2}))',$$

and the half sphere $\mathbb{S} : \{\gamma \in [0, 1]^{d-1} : ||\gamma|| \le 1\} \to \mathcal{S}_{d-1}$ with

$$\mathbb{S}(\gamma) = (\gamma_1, \dots, \gamma_{d-1}, \sqrt{1 - \gamma_1^2 - \cdots - \gamma_{d-1}^2})'.$$

where ψ_0' is the derivative of ψ_0 and $\mathbb{J}(\gamma)$ is the Jacobian of $\mathbb{S}(\gamma)$. Thus, it is natural to construct an estimator of θ_0 by taking an empirical counterpart of (3) and inserting estimators for ψ_0' 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 ψ_0' without introducing smoothing parameters. To address this issue, we follow the idea in BGH and Groeneboom and Hendrickx (2018) and focus on the following modified population score equation

$$E\left[\begin{pmatrix} X \\ \mathbb{J}(\gamma_0)'Z \end{pmatrix}\{Y - X'\beta_0 - \psi_0(Z'\mathbb{S}(\gamma_0))\}\right] = 0. \qquad (4)$$

Since the error term ϵ is orthogonal to any function of (X, Z) under $E[\epsilon|X, Z] = 0$, (4) is also a valid score equation, and we construct an estimator for θ_0 based on this equation.

In particular, for each $\theta = (\beta', \gamma')'$, we estimate the monotone function ψ_0 by the least squares

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

where $\mathcal{M}$ is the set of monotone increasing functions defined on $\mathbb{R}$. The function $\hat{\psi}_{n\theta}$ can be obtained by isotonic regression (see, e.g., Groeneboom and Jongbloed, 2014, for a review). Then our estimator $\hat{\theta} = (\hat{\beta}', \hat{\gamma}')'$ of θ_0 is given by the zero-crossing root of the score function[2]

$$\phi_n(\theta) = \frac{1}{n} \sum_{i=1}^n \begin{pmatrix} X_i \\ \mathbb{J}(\gamma)'Z_i \end{pmatrix}\{Y_i - X_i'\beta - \hat{\psi}_{n\theta}(Z_i'\mathbb{S}(\gamma))\}, \qquad (6)$$

and α_0 is estimated by $\hat{\alpha} = \mathbb{S}(\hat{\gamma})$. The reason for the definition based on the zero-crossing is due to the fact that $\hat{\psi}_{n\theta}$ is a discrete function taking finite different values. Thus, we might be unable to solve $\phi_n(\theta) = 0$ exactly.[3] As $n \to \infty$, the zero-crossing solution should become an exact solution. In practice, we can minimize the square sum of the right hand side of (6) to obtain a good approximation of the zero-crossing.

Remark 1. [Technical intuition for the difference between the score estimation and least square approaches] Our discussion is based on Groeneboom and Hendrickx (2018, pp. 1419-1420). Let $\Gamma_n(\theta)$ be some objective function for θ and $\Gamma(\theta)$ is its population counterpart. The M-estimator is defined as a maximizer of $\Gamma_n(\theta)$. The $\sqrt{n}$-consistency of the estimator is typically derived from a quadratic expansion $\Gamma(\theta) - \Gamma(\theta_0) \le -c\|\theta - \theta_0\|^2$ for some $c > 0$ in a neighborhood of θ_0 combined with the approximation to the objective function

$$\Gamma_n(\theta) - \Gamma_n(\theta_0) = \Gamma(\theta) - \Gamma(\theta_0) + O_p(n^{-1/2}\|\theta - \theta_0\|) + o_p(\|\theta - \theta_0\|^2) + O_p(n^{-1}), \qquad (7)$$

---

[2]We say that θ^* is a zero-crossing of a real-valued function $\zeta : \Theta \to \mathbb{R}$ if each open neighborhood of θ^* contains points $\theta_1, \theta_2 \in \Theta$ such that $\zeta(\theta_1)\zeta(\theta_2) \le 0$. This definition can be extended to a vector of functions, where a zero-crossing vector has each of its component to be a zero-crossing in the corresponding dimension.

[3]Similar to other estimators by BGH or Groeneboom and Hendrickx (2018), our zero-crossing estimator $\hat{\theta}$ may not be unique. Indeed there are many flat parts in $\phi_n(\theta)$, and the intersection of $\phi_n(\theta)$ and zero could be an interval. In this case, any point on this interval will satisfy the results in Theorems 1 and 3 below.

uniformly over a shrinking neighborhood of θ_0. However, when we apply this argument to the (profile) least square objective function $\frac{1}{n} \sum_{i=1}^{n} \{Y_i - X_i'\beta - \hat{\psi}_{n\theta}(Z_i'\mathbb{S}(\gamma))\}^2$, it seems to have an extra term of order $O_p(n^{-2/3})$ in (7) due to discontinuity of $\hat{\psi}_{n\theta}$ in θ (although there is no rigorous proof). If there is such an extra term, we expect that the least square estimator for θ will not achieve $\sqrt{n}$-consistency.[4] On the other hand, it turns out that our score (or Z-) estimating equation $\phi_n(\theta)$ can be approximated by $\phi_n(\theta) = \phi'(\theta_0)(\theta - \theta_0) + O_p(n^{-1/2})$ uniformly over a shrinking neighborhood of θ_0, where $\phi'(\theta_0)$ is the derivative of the population counterpart of $\phi_n(\theta)$ displayed in (4). In short, the difference between the score estimation and least square approaches is due to different orders of the remainders in the Z- and M-estimation approaches in this setup.

Remark 2. [Comparison with smoothing approach] Let us take Xia and Härdle (2006) as an example for the conventional smoothing approach to estimate the PLSI model (without monotonicity on ψ_0) and compare with our estimation approach. A common feature is that both methods estimate the nonparametric function ψ_0 with fixed θ, and then optimize or solve for $\hat{\theta}$ in a two step or recursive strategy. The main difference is that we use the isotonic regression to estimate the monotone function ψ_0, but Xia and Härdle (2006) employ a weighted local linear regression to estimate ψ_0 for each fixed θ. Our score-type estimation method does not require any tuning parameter to estimate ψ_0, while a smoothing parameter is innate in Xia and Härdle (2006). The technical arguments are very different as well. Our consistency and asymptotic normality proofs below heavily rely on properties of the monotone function class and associated empirical processes. On the other hand, the argument in Xia and Härdle (2006) is to show how the linear regression for θ_0 averages out the estimation errors from the local linear regression for ψ_0 based on the U-statistic theory to achieve the $\sqrt{n}$-consistency of their estimator for θ_0.

2.2 Asymptotic properties of estimator

We now investigate asymptotic properties of the estimator $\hat{\theta}$. Let $\mathbb{I}_k$ be the $k \times k$ identity matrix, $\|\cdot\|$ be the Euclidean norm, $\mathcal{B}(a_0, A) = \{a : \|a - a_0\| \leq A\}$ be a ball around a_0 of radius A, and

$$
T_0 = \begin{bmatrix} \mathbb{I}_k & 0 \\ 0 & \mathbb{J}(\gamma_0)' \end{bmatrix}, \qquad V_{x,z} = \begin{pmatrix} x - E[X|z'\mathbb{S}(\gamma_0)] \\ z - E[Z|z'\mathbb{S}(\gamma_0)] \end{pmatrix},
$$

$$
V_{x,z,\psi'} = \begin{pmatrix} x - E[X|z'\mathbb{S}(\gamma_0)] \\ \{z - E[Z|z'\mathbb{S}(\gamma_0)]\}\psi_0'(z'\mathbb{S}(\gamma_0)) \end{pmatrix}.
$$

We impose the following assumptions.

Assumption.

A1 *The spaces $\mathcal{X}$ and $\mathcal{Z}$ are convex with non-empty interiors, and satisfy $\mathcal{X} \subset \mathcal{B}(0, R)$ and $\mathcal{Z} \subset \mathcal{B}(0, R)$ for some $R > 0$.*

---

[4]We note that even for single index models, the convergence rate and asymptotic distribution of the least square estimator, $\arg\min_{\gamma} \{\min_{\psi \in \mathcal{M}} \frac{1}{n} \sum_{i=1}^{n} \{Y_i - \psi(Z_i'\mathbb{S}(\gamma))\}^2\}$, is an open problem.

A2 *There exists $K_0 > 0$ such that $|\psi_0(u)| < K_0$ for all $u \in \{z'\alpha : z \in \mathcal{Z}, \alpha \in \mathcal{S}_{d-1}\}$.*

A3 *There exists $\delta_0 > 0$ such that the function $\psi_\theta(u) = \psi_{\alpha,\beta}(u) = E[Y - X'\beta|Z'\alpha = u]$ is monotone increasing on $I_\alpha = \{z'\alpha, z \in \mathcal{Z}\}$ for each $\theta \in \mathcal{B}(\theta_0, \delta_0)$.*

A4 *For $W = X$ or Z, the mapping $u \mapsto E[W|Z'\alpha = u]$ defined on I_α is bounded and has a finite total variation.*

A5 *There exist $c_0 > 0$ and $M_0 > 0$ such that $E[|Y - X'\beta|^m|Z = z] \leq m!M_0^{m-2}c_0$ for all integers $m \geq 2$, each $\theta \in \mathcal{B}(\theta_0, \delta_0)$ and almost every $z \in \mathcal{Z}$ (according to the true distribution).*

A6 $\mathrm{Cov}[(\beta_0 - \beta)'X + Z'(\mathbb{S}(\gamma_0) - \mathbb{S}(\gamma)), \ (\beta_0 - \beta)'X + \psi_0(Z'\mathbb{S}(\gamma_0))|Z'\mathbb{S}(\gamma)] \neq 0$ *almost surely for each $\theta \neq \theta_0$.*

A7 $B = T_0 \int V_{x,z}V'_{x,z,\psi'}dP_0(x,z)T'_0$ *and* $B_E = T_0 \int V_{x,z,\psi'}V'_{x,z,\psi'}dP_0(x,z)T'_0$ *are non-singular.*

A1 and A2, which are similar to the assumptions A1 and A2 in BGH, impose boundedness on the support of covariates and the monotone function ψ_0. These conditions are used to control the entropy of the function classes that characterize (6). We note that Xia and Härdle (2006) and Wang *et al.* (2010) imposed similar conditions. A3, which is an adaptation of BGH's A3, requires monotonicity of ψ_θ in a neighborhood of θ_0. This assumption is used to establish the consistency of the estimator $\hat{\psi}_{n\theta}(z'\alpha)$ for each $\theta \in \mathcal{B}(\theta_0, \delta_0)$. For example, A3 is satisfied with $\psi_0(2^{-1/2}z_1 + 2^{-1/2}z_2) = (2^{-1/2}z_1 + 2^{-1/2}z_2)^3$ and $Z_1, Z_2 \sim U[0,1]$ as in p. 12 of BGH. A4 is imposed to control the entropy of function classes to achieve the $\sqrt{n}$-convergence rate. This assumption can be derived from BGH's A4 and A5. A5 is a modified version of BGH's A6. This assumption is introduced to show that $\max_{\theta \in \mathcal{B}(\theta_0,\delta_0)} \sup_{z \in \mathcal{Z}} \hat{\psi}_{n\theta}(z'\alpha) = O_p(\log n)$, which is used to obtain an entropy result associated with the $\sqrt{n}$-convergence rate. A5 is satisfied if the conditional distribution $Y - X'\beta|Z$ belongs to some exponential family.[5] A6 and A7 are to ensure the consistency and existence of limiting variances of the simple score and efficient score estimators, respectively. A6 is related to BGH's A7 after taking expansion of $\mathbb{S}(\gamma_0) - \mathbb{S}(\gamma)$ around $\gamma = \gamma_0$.

Under these assumptions, the asymptotic properties of the simple estimator $\hat{\theta}$ are presented as follows.

Theorem 1. *Suppose Assumptions A1-A7 hold true. Then $\hat{\theta}$ exists with probability approaching one, $\hat{\theta} \xrightarrow{p} \theta_0$, and*

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Pi),$$

where $\Pi = B^{-1}T_0\Sigma T'_0(B^{-1})'$, $\Sigma = \mathrm{Var}(V_{X,Z}\epsilon)$, and $V_{X,Z}$ is $V_{x,z}$ evaluated at $(x,z) = (X,Z)$.

---

[5] Let W be an integrable random variable with the density $w \mapsto h(w, \vartheta_2) \exp\{\vartheta_2^{-1}w\ell(\vartheta_1) - \vartheta_2^{-1}B(\ell(\vartheta_1))\}$, where ϑ_1 is the mean, ϑ_2 is a dispersion parameter, ℓ is a real valued function with a strictly positive first derivative on an open interval, B is a real valued function, and h is a normalizing function. Balabdaoui, Durot and Jankowski (2019, Proposition 9.2) showed that there exist $c > 0$ and $M > 0$ such that $E[|W|^m] \leq m!M^{m-2}c$ for all integers $m \geq 2$. This proposition can be adapted to provide primitive conditions for A5 on the conditional distribution $Y - X'\beta|Z = z$, where the parameters ϑ_1 and ϑ_2 may vary with z.

This theorem says that our score-type estimator $\hat{\theta}$ for the monotone PLSI model is $\sqrt{n}$-consistent and asymptotically normal without any tuning parameter.[6] The asymptotic variance Π can be estimated by (i) replacing P_0 with the empirical measure $\mathbb{P}_n$, (ii) replacing γ_0 with its estimator $\hat{\gamma}$, (iii) replacing ψ'_0 with $\hat{\psi}'_{nh,\theta}$ in (9) below, (iv) replacing ϵ with the residuals based on our estimator, and (v) replacing the conditional expectations with kernel estimators.[7] Our result can be considered as an extension of BGH for the monotone PLSI model. Technically a major difference from BGH is the treatment on the mapping $\psi_\theta(\cdot)$, which involves an additional term from the linear component $X'\beta$ (i.e., the second term of (10) in Appendix). Most entropy results in our proof are modified to accommodate this additional term.

We note that the estimator $\hat{\theta}$ is derived from the modified population score equation in (4) instead of the original one in (3). Consequently, the asymptotic variance Π of $\hat{\theta}$ is not the efficient variance for the PLSI model. If we allow one tuning parameter, we can evaluate the efficient score function in (3) as

$$\xi_{nh}(\theta) = \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} X_i \\ \mathbb{J}(\gamma)'Z_i\hat{\psi}'_{nh,\theta}(Z_i'\mathbb{S}(\gamma)) \end{pmatrix} \{Y_i - X_i'\beta - \hat{\psi}_{n\theta}(Z_i'\mathbb{S}(\gamma))\}, \tag{8}$$

where

$$\hat{\psi}'_{nh,\theta}(u) = \frac{1}{h} \int K\left(\frac{u-x}{h}\right) d\hat{\psi}_{n\theta}(x), \tag{9}$$

is an estimator for the derivative of ψ_θ (defined in A3) with a kernel function K and bandwidth h. Let $\tilde{\theta} = (\tilde{\beta}', \tilde{\gamma}')'$ be the zero-crossing of (8).[8] For this estimator, we add the following assumptions.

Assumption.

A8 $\psi_\theta(z'\alpha)$ *is twice continuously differentiable on* $I_\alpha = \{z'\alpha, z \in \mathcal{Z}\}$ *for each* $\theta \in \mathcal{B}(\theta_0, \delta_0)$.

A9 $K(\cdot)$ *is a symmetric twice differentiable kernel function with compact support* $[-1, 1]$. *Furthermore,* $h \asymp n^{-1/7}$.

A8 is an additional condition to control the entropy for classes of functions to achieve the $\sqrt{n}$-consistency of $\tilde{\theta}$. A9 contains assumptions for the kernel function K and bandwidth h to evaluate $\hat{\psi}'_{nh,\theta}$ in (9). The condition $h \asymp n^{-1/7}$ is also imposed in BGH.

The asymptotic properties of the estimator $\tilde{\theta}$ are presented as follows.

---

[6]Due to discontinuity in $\hat{\psi}_{n\theta}$, we can only guarantee the existence of $\hat{\theta}$ with probability approaching one. Similar to other zero-crossing estimators using isotonic regression, its existence for a given sample size is an open question.

[7]For example, the conditional expectation $\mu(z) = E[X|z'\mathbb{S}(\gamma_0)]$ in $V_{x,z}$ and $V_{x,z,\psi'}$ can be estimated by

$$\hat{\mu}(z) = \frac{\sum_{i=1}^{n} K\left(\frac{Z_i'\mathbb{S}(\hat{\gamma}) - z'\mathbb{S}(\hat{\gamma})}{b}\right) X_i}{\sum_{i=1}^{n} K\left(\frac{Z_i'\mathbb{S}(\hat{\gamma}) - z'\mathbb{S}(\hat{\gamma})}{b}\right)},$$

where K is a kernel function (e.g., Gaussian and Epanechnikov) and b is a bandwidth.

[8]Similar to $\hat{\theta}$, the zero-crossing estimator $\tilde{\theta}$ may not be unique. If the intersection of $\xi_{nh}(\theta)$ and zero is an interval, any point on this interval satisfies the result in Theorem 2.

Theorem 2. *Suppose Assumptions A1-A9 hold true. Then $\tilde{\theta}$ exists with probability approaching one, $\tilde{\theta} \xrightarrow{p} \theta_0$, and*

$$\sqrt{n}(\tilde{\theta} - \theta_0) \xrightarrow{d} N(0, \Pi_E),$$

where $\Pi_E = B_E^{-1} T_0 \Sigma T_0' (B_E^{-1})'$, $\Sigma = \text{Var}(V_{X,Z,\psi'} \epsilon)$, and $V_{X,Z,\psi'}$ is $V_{x,z,\psi'}$ evaluated at $(x, z) = (X, Z)$.

If we additionally assume $\text{Var}(\epsilon | X, Z) = \text{Var}(\epsilon) = \sigma^2$ (i.e., the error term ϵ is homoskedastic), then Σ can be written as $\Sigma = \sigma^2 \int V_{x,z,\psi'} V_{x,z,\psi'}' dP_0(x, z)$. Therefore, the asymptotic variance becomes $\Pi_E = B_E^{-1}$, which coincides with the efficient variance matrix derived in Carroll *et al.* (1997) and Xia and Härdle (2006). The asymptotic variance Π_E can be estimated in the same manner as Π.

2.3   Bootstrap inference

One advantage of the proposed estimator $\hat{\theta}$ is that it is free from tuning parameters, such as bandwidths and series lengths. On the other hand, since its asymptotic variance Π involves conditional means, inference using estimation of Π requires some smoothing method. To obtain an inference procedure which is free from tuning parameters, we propose a bootstrap method to approximate the distribution of the score-type estimator $\hat{\theta}$. Groeneboom and Hendrickx (2017) established the bootstrap validity of their score estimator for the parametric part in a current status model. We extend their result to the monotone PLSI model.

Let $\hat{\theta}^*$ be the bootstrap counterpart of $\hat{\theta}$ defined in Section 2.1 based on resamples from the empirical distribution of $\{Y_i, X_i, Z_i\}_{i=1}^n$. The validity of the bootstrap approximation is obtained as follows.

Theorem 3. *Suppose Assumptions A1-A7 hold true. Then*

$$\sup_{t \in \mathbb{R}^{k+d-1}} |P^*\{\sqrt{n}(\hat{\theta}^* - \hat{\theta}) \le t\} - P_0\{\sqrt{n}(\hat{\theta} - \theta_0) \le t\}| \xrightarrow{p} 0,$$

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

The bootstrap confidence interval and standard error can be obtained by this result. Note that computation of $\hat{\theta}^*$ and the resulting bootstrap inference are free from tuning parameters.

3   Simulation

In this section, we conduct a simulation study to illustrate the finite sample performance of the proposed estimators.

3.1 Simple score and efficient score estimators

We consider the following partial linear model:

$$Y = X\beta_0 + \psi_0(Z'\alpha_0) + \epsilon,$$
$$\psi_0(u) = u^3, \qquad \beta_0 = 1, \qquad \alpha_0' = (1,1)/\sqrt{2} \approx (0.7071, 0.7071),$$

where $X \sim N(0,1)$ and $\epsilon \sim N(0,1)$. For Z, we consider two data generating processes: (i) $Z \sim U[1,2]^2$ (in Table 1) and (ii) $Z \sim N(0, \mathbb{I}_2)$ with the 2×2 identity matrix (in Table 2). The sample sizes are $n = 100, 500$, and 1000. The number of Monte Carlo replications is 1000. Tables 1 and 2 present the Monte Carlo averages $(\hat{\mu}_\beta, \hat{\mu}_{\alpha_1}, \hat{\mu}_{\alpha_2})$ and variances $(\hat{\sigma}_\beta^2, \hat{\sigma}_{\alpha_1}^2, \hat{\sigma}_{\alpha_2}^2)$ (multiplied by n) of the estimates $(\hat{\beta}, \hat{\alpha}_1, \hat{\alpha}_2)$ and $(\tilde{\beta}, \tilde{\alpha}_1, \tilde{\alpha}_2)$ for Cases (i) and (ii), respectively.

In the tables, SSE is the simple score estimator obtained by solving the zero-crossing of (6), and ESE is the efficient score estimator obtained by solving the zero-crossing of (8). SSE_L and ESE_L are the Lagrange versions of SSE and ESE suggested by BGH and Groeneboom (2018).[9] All these methods are implemented by the Hooke-Jeeves algorithm to search a minimizer of the sum of squared score components. In the reported simulation results, we follow BGH and use the true values as starting values. Preliminary simulation suggests that the results are not sensitive to local changes for the starting values. For comparison, we include monotone least square methods (LSE in the tables). We also include the smoothing method by Xia and Härdle (2006) into our comparison (S_LSE in the tables). Xia and Härdle (2006) showed that the optimal bandwidth for their methods is of order $n^{-1/5}$. BGH showed that the optimal bandwidth for their efficient estimator is of order $n^{-1/7}$, and suggested to use $h = \hat{r}n^{-1/7}$, where $\hat{r}$ is the range of $Z'\alpha$, as bandwidth. Here we follow BGH's practice. We choose $\hat{r}n^{-1/7}$ as bandwidth for ESE and $\hat{r}n^{-1/5}$ for S_LSE.

The theoretical asymptotic variances are calculated for SSE, ESE, and S_LSE. Both ESE and S_LSE achieve semiparametric efficiency and therefore they should have the same limit. The asymptotic variance of LSE is unknown in the literature (see, Balabdaoui, Durot and Jankowski, 2019, for a detail). It can be shown that for both settings, $Z \sim U[1,2]^2$ and $Z \sim N(0, \mathbb{I}_2)$, we have $E[X|z'\alpha] = 0$ and $E[Z|z'\alpha] = \frac{\sqrt{2}}{2}z'\alpha(1,1)'$. The asymptotic variances of $(\hat{\beta}, \hat{\alpha})$ and $(\tilde{\beta}, \tilde{\alpha})$ (the estimators without reparameterization) can be obtained with Lemma 7 in BGH and numerical integral. In particular, we have $\sqrt{n}\{(\hat{\beta}', \hat{\alpha}')' - (\beta_0', \alpha_0')'\} \xrightarrow{d} N(0, V)$ and $\sqrt{n}\{(\tilde{\beta}', \tilde{\alpha}')' -$

---

[9]More precisely, the estimator SSE_L is obtained by a zero-crossing of

$$\phi_n^L(\theta) = \left[\begin{array}{c} \frac{1}{n}\sum_{i=1}^n X_i\{Y_i - X_i'\beta - \hat{\psi}_{n\alpha}(Z_i'\alpha)\} \\ \frac{1}{n}(1 - \alpha'\alpha)\sum_{i=1}^n X_i\{Y_i - X_i'\beta - \hat{\psi}_{n\alpha}(Z_i'\alpha)\} \end{array} \right],$$

and ESE_L is defined analogously.

$(\beta'_0, \alpha'_0)'\} \xrightarrow{d} N(0, V_E)$, where

$$
\text{Case (i)} \quad : \quad V = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.0324 & -0.0324 \\ 0 & -0.0324 & 0.0324 \end{pmatrix}, \quad V_E = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.0315 & -0.0315 \\ 0 & -0.0315 & 0.0315 \end{pmatrix}.
$$

$$
\text{Case (ii)} \quad : \quad V = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.0555 & -0.0555 \\ 0 & -0.0555 & 0.0555 \end{pmatrix}, \quad V_E = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.0185 & -0.0185 \\ 0 & -0.0185 & 0.0185 \end{pmatrix}.
$$

Tables 1 and 2 show that the estimation biases are reasonably small for the both estimators even for $n = 100$. For the single index part ($\hat{\alpha}_1$ and $\hat{\alpha}_2$), ESE performs better than SSE in terms of efficiency, which is in accordance with the implication of Theorems 1 and 2. As the sample size increases, SSE_L and ESE_L become almost identical to SSE and ESE, respectively. LSE performs differently in two cases. In Table 2, LSE performs better than SSE but worse than ESE. In Table 1, LSE performs worse than SSE.

In general, all the variances of SSE and ESE are approaching to their theoretical limits. It seems that the approaching rates are faster in Case (i) than those in Case (ii). S_LSE is approaching the limit in Case (i), but stays away from the limit in Case (ii). Note that Case (ii) violates the assumption that the support of Z is compact required in both Xia and Härdle (2006) and our estimators. Therefore, some irregular behaviors of those estimators might be expected in Case (ii). Nevertheless, SSE and ESE seem to be more stable even if the support of Z is not compact.

Overall, the simulation results are encouraging to support our estimation strategy.

Table 1: Simulation results for Case (i) $Z \sim U[1, 2]^2$

Methods	n	$\hat{\mu}_\beta$	$\hat{\mu}_{\alpha_1}$	$\hat{\mu}_{\alpha_2}$	$\hat{\sigma}^2_\beta$	$\hat{\sigma}^2_{\alpha_1}$	$\hat{\sigma}^2_{\alpha_2}$
SSE	100	0.9982	0.7068	0.7068	1.3401	0.0415	0.0416
	500	0.9982	0.7068	0.7073	1.0277	0.0364	0.0364
	1000	1.0002	0.7069	0.7073	1.1306	0.0322	0.0322
	∞	1	0.7071	0.7071	1	0.0324	0.0324
ESE	100	0.9984	0.7067	0.7069	1.3743	0.0404	0.0404
	500	0.9983	0.7068	0.7073	1.0252	0.0360	0.0359
	1000	1.0001	0.7069	0.7073	1.1310	0.0319	0.0319
	∞	1	0.7071	0.7071	1	0.0315	0.0315
SSE_L	100	0.9982	0.7072	0.7064	1.3425	0.0420	0.0421
	500	0.9982	0.7068	0.7073	1.0296	0.0363	0.0363
	1000	1.0002	0.7069	0.7073	1.1288	0.0323	0.0323
	∞	1	0.7071	0.7071	1	0.0324	0.0324
ESE_L	100	0.9982	0.7070	0.7066	1.3502	0.0408	0.0410
	500	0.9982	0.7069	0.7072	1.0262	0.0361	0.0360
	1000	1.0001	0.7069	0.7073	1.1336	0.0318	0.0318
	∞	1	0.7071	0.7071	1	0.0315	0.0315
LSE	100	0.9972	0.7074	0.7058	1.3967	0.0703	0.0699
	500	0.9984	0.7067	0.7073	1.0330	0.0754	0.0752
	1000	1.0002	0.7069	0.7072	1.1253	0.0740	0.0739
	∞	1	0.7071	0.7071	n/a	n/a	n/a
S_LSE	100	1.0022	0.7071	0.7065	1.2891	0.0441	0.0443
	500	1.0005	0.7069	0.7072	1.2213	0.0362	0.0361
	1000	1.0023	0.7069	0.7072	1.2053	0.0348	0.0348
	∞	1	0.7071	0.7071	1	0.0315	0.0315

Table 2: Simulation results for Case (ii) $Z \sim N(0, \mathbb{I}_2)$

Methods	n	$\hat{\mu}_\beta$	$\hat{\mu}_{\alpha_1}$	$\hat{\mu}_{\alpha_2}$	$\hat{\sigma}^2_\beta$	$\hat{\sigma}^2_{\alpha_1}$	$\hat{\sigma}^2_{\alpha_2}$
SSE	100	0.9981	0.7035	0.7075	1.3620	0.2310	0.2301
	500	1.0001	0.7065	0.7074	1.1481	0.1087	0.1086
	1000	0.9998	0.7079	0.7062	1.0532	0.0932	0.0937
	∞	1	0.7071	0.7071	1	0.0555	0.0555
ESE	100	0.9988	0.7049	0.7080	1.4422	0.0943	0.0940
	500	1.0000	0.7069	0.7072	1.1333	0.0356	0.0355
	1000	0.9999	0.7075	0.7067	1.0531	0.0309	0.0310
	∞	1	0.7071	0.7071	1	0.0185	0.0185
SSE_L	100	0.9981	0.7037	0.7072	1.3625	0.2352	0.2347
	500	1.0000	0.7065	0.7074	1.1467	0.1090	0.1091
	1000	0.9998	0.7079	0.7062	1.0548	0.0936	0.0941
	∞	1	0.7071	0.7071	1	0.0555	0.0555
ESE_L	100	0.9974	0.7054	0.7074	1.4086	0.0967	0.0973
	500	1.0000	0.7070	0.7071	1.1357	0.0355	0.0355
	1000	0.9999	0.7075	0.7066	1.0589	0.0310	0.0311
	∞	1	0.7071	0.7071	1	0.0185	0.0185
LSE	100	0.9978	0.7063	0.7061	1.3306	0.1269	0.1281
	500	1.0001	0.7071	0.7069	1.1441	0.0815	0.0815
	1000	0.9998	0.7077	0.7064	1.0595	0.0726	0.0729
	∞	1	0.7071	0.7071	n/a	n/a	n/a
S_LSE	100	1.0052	0.7058	0.7034	6.2528	0.3584	0.3599
	500	0.9972	0.7067	0.7065	7.0103	0.3560	0.3589
	1000	1.0022	0.7069	0.7068	6.9869	0.3878	0.3878
	∞	1	0.7071	0.7071	1	0.0185	0.0185

3.2 Bootstrap

As mentioned in Section 2.3, the purpose of our bootstrap method is to obtain an inference method that is free of tuning parameters. Therefore, we focus on SSE here, since ESE requires at least one tuning parameter. Since the results are analogous, we only consider Case (ii) above. Most notations in Table 3 are as defined in the previous subsection. Results for SSE are replicated from Table 2. SSE_b is the bootstrap counterpart of the estimator by SSE, and the number of the bootstrap replications is 500.

Table 3 shows that as the sample size increases, the distribution of SSE_b approaches to that of SSE, which is in accordance with the implication of Theorem 3.

Table 3: Simulation results for bootstrap counterparts

Methods	n	$\hat{\mu}_\beta$	$\hat{\mu}_{\alpha_1}$	$\hat{\mu}_{\alpha_2}$	$\hat{\sigma}^2_\beta$	$\hat{\sigma}^2_{\alpha_1}$	$\hat{\sigma}^2_{\alpha_2}$
SSE	100	0.9982	0.7068	0.7068	1.3401	0.0415	0.0416
	500	0.9982	0.7068	0.7073	1.0277	0.0364	0.0364
	1000	1.0002	0.7069	0.7073	1.1306	0.0322	0.0322
SSE_b	100	1.0599	0.7078	0.7059	1.2614	0.0488	0.0507
	500	0.9729	0.6970	0.7170	1.0354	0.0286	0.0270
	1000	0.9952	0.7092	0.7049	1.1236	0.0359	0.0364

A Proof of Theorem 1

Notation: We use the following notation. Let $||\mathbb{G}_n||_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |\sqrt{n}(\mathbb{P}_n - P_0)f|$, $|| \cdot ||_{B,P_0}$ be the Bernstein norm under a measure P_0,

$$H_B(\varepsilon, \mathcal{F}, || \cdot ||_{B,P_0}) = \log N_{[]}(\varepsilon, \mathcal{F}, || \cdot ||_{B,P_0}),$$

be the entropy of the ε-bracketing number of the function class $\mathcal{F}$ under $|| \cdot ||_{B,P_0}$, and

$$J_n(\delta) = J_n(\delta, \mathcal{F}, || \cdot ||_{B,P_0}) = \int_0^{\delta} \sqrt{1 + H_B(\varepsilon, \mathcal{F}, || \cdot ||_{B,P_0})} d\varepsilon.$$

A.1 Proof of existence and consistency

For fixed α and β (γ is also fixed by the uniqueness of reparameterization $\mathbb{S}(\cdot)$, so is θ). Let $\psi_\theta(u) = E[Y - X'\beta | Z'\alpha = u]$, which can be written as (by $E[\epsilon|Z] = 0$)

$$\psi_\theta(u) = E[\psi_0(Z'\alpha_0)|Z'\alpha = u] + (\beta_0 - \beta)'E[X|Z'\alpha = u]. \tag{10}$$

A similar argument to Theorem 5 of BGH implies that $\hat{\theta}$ exists with probability approaching one. We now show the consistency of $\hat{\theta}$. Since $\hat{\theta} = \hat{\theta}_n$ is estimated in a compact set, there exists a subsequence $\{\hat{\theta}_{n_k}\}_{k \in \mathbb{N}}$ of $\{\hat{\theta}_n\}_{n \in \mathbb{N}}$ almost surely converging to some point $\theta^* = (\beta^{*\prime}, \gamma^{*\prime})'$. By Proposition 4 in BGH combined with $\hat{\theta}_{n_k} \overset{as}{\to} \theta^*$, we have

$$\int \left\{ \hat{\psi}_{n_k \hat{\theta}_{n_k}}(z'\mathbb{S}(\hat{\gamma}_{n_k})) - \psi_{\theta^*}(z'\mathbb{S}(\gamma_*)) \right\}^2 dP_0(z) \overset{p}{\to} 0.$$

Also by Proposition 9 in supplementary material of BGH (hereafter BGH-supp), the zero-crossing $\hat{\theta}$ becomes a root of the continuous limiting function, i.e.,

$$\phi_{n_k}(\hat{\theta}_{n_k}) \overset{p}{\to} \phi(\theta^*) = 0,$$

as $k \to \infty$, where $\phi(\theta) = \int \begin{pmatrix} x \\ J(\gamma)'z \end{pmatrix} \{y - x'\beta - \psi_\theta(z'\mathbb{S}(\gamma))\} dP_0(x, y, z)$, and the equality follows from the definition of zero-crossing and the continuity of $\psi_\theta(\cdot)$. Then we have

$$
\begin{aligned}
0 &= (\theta_0 - \theta^*)'\phi(\theta^*) \\
&= (\theta_0 - \theta^*)' \int \begin{pmatrix} x \\ \mathbb{J}(\gamma^*)'z \end{pmatrix} \left\{ \begin{array}{c} x'\beta_0 + \psi_0(z'\mathbb{S}(\gamma_0)) - x'\beta^* \\ -\{E[\psi_0(Z'\mathbb{S}(\gamma_0))|z'\mathbb{S}(\gamma^*)] + (\beta_0 - \beta^*)'E[X|z'\mathbb{S}(\gamma^*)]\} \end{array} \right\} dP_0(x,z) \\
&= \begin{pmatrix} \beta_0 - \beta^* \\ \gamma_0 - \gamma^* \end{pmatrix}' \int \begin{pmatrix} x - E[X|z'\mathbb{S}(\gamma^*)] \\ \mathbb{J}(\gamma^*)'\{z - E[Z|z'\mathbb{S}(\gamma^*)]\} \end{pmatrix} \left\{ \begin{array}{c} (\beta_0 - \beta^*)'\{x - E[X|z'\mathbb{S}(\gamma^*)]\} \\ +\psi_0(z'\mathbb{S}(\gamma_0)) - E[\psi_0(Z'\alpha_0)|z'\mathbb{S}(\gamma^*)] \end{array} \right\} dP_0(x,z) \\
&= E\left[\mathrm{Cov}[(\beta_0 - \beta^*)'X + (\gamma_0 - \gamma^*)'\mathbb{J}(\gamma^*)'Z, (\beta_0 - \beta^*)'X + \psi_0(Z'\mathbb{S}(\gamma_0))|Z'\mathbb{S}(\gamma^*)] \right] \\
&= E\left[\mathrm{Cov}[(\beta_0 - \beta^*)'X + Z'(\mathbb{S}(\gamma_0) - \mathbb{S}(\gamma^*)) + o(\gamma_0 - \gamma^*), (\beta_0 - \beta^*)'X + \psi_0(Z'\mathbb{S}(\gamma_0))|Z'\mathbb{S}(\gamma^*)] \right] \\
&= E\left[\mathrm{Cov}[(\beta_0 - \beta^*)'X + Z'(\mathbb{S}(\gamma_0) - \mathbb{S}(\gamma^*)), (\beta_0 - \beta^*)'X + \psi_0(Z'\mathbb{S}(\gamma_0))|Z'\mathbb{S}(\gamma^*)] \right] + o(\gamma_0 - \gamma^*),
\end{aligned}
$$

where the second equality follows from (10), the third equality follows from the law of iterated expectations, the fifth equality follows from an expansion of $\mathbb{S}(\gamma_0)$ around $\gamma_0 = \gamma_*$, and the last equality follows from A1. Therefore, by A6, $0 = (\theta_0 - \theta^*)'\phi(\theta^*)$ holds true only if $\theta^* = \theta_0$, and the consistency of $\hat{\theta}$ follows.

A.2 Proof of asymptotic normality

The proof is split into several steps.

Step 1: Derive a decomposition of $\phi_n(\hat{\theta})$

For each $\theta = (\beta', \gamma')'$, let $u_i = z_i'\mathbb{S}(\gamma)$ and $\{u_{n_j,\theta}\}_{j=1}^k$ be the subsequence of $\{u_i\}_{i=1}^n$ representing all the jump points of $\hat{\psi}_{n\theta}(\cdot)$. By the construction of $\hat{\psi}_{n\theta}(\cdot)$ (see, Lemmas 2.1 and 2.3 in Groeneboom and Jongbloed, 2014), we have $\sum_{i=n_j}^{n_{j+1}-1}\{y_i - x_i'\beta - \hat{\psi}_{n\theta}(u_i)\} = 0$ for each $j = 1,\ldots,k$, which means

$$
\sum_{j=1}^k m_j \sum_{i=n_j}^{n_{j+1}-1} \{y_i - x_i'\beta - \hat{\psi}_{n\theta}(u_i)\} = 0, \tag{11}
$$

for any weights $\{m_j\}_{j=1}^k$. As in BGH, we define for $W = X$ or Z,

$$
\bar{E}_{n,\theta}[W|u] = \bar{E}_{n,\theta}[W|z'\mathbb{S}(\gamma)] = \begin{cases} E[W|Z'\mathbb{S}(\gamma) = u_{n_j}] & \text{if } \psi_\theta(u) > \hat{\psi}_{n\theta}(u_{n_j}) \text{ for all } u \in (u_{n_j}, u_{n_{j+1}}) \\ E[W|Z'\mathbb{S}(\gamma) = s] & \text{if } \psi_\theta(u) = \hat{\psi}_{n\theta}(s) \text{ for some } s \in (u_{n_j}, u_{n_{j+1}}) \\ E[W|Z'\mathbb{S}(\gamma) = u_{n_{j+1}}] & \text{if } \psi_\theta(u) < \hat{\psi}_{n\theta}(u_{n_j}) \text{ for all } u \in (u_{n_j}, u_{n_{j+1}}) \end{cases} \tag{12}
$$

for $u \in [u_{n_j}, u_{n_{j+1}})$ with $j = 1,\ldots,k$ (if $j = k$, set $u_{n_{j+1}} = \max_i u_{n_i}$). By (11), it holds

$$
\int \bar{E}_{n,\hat{\theta}}[W|z'\mathbb{S}(\gamma)]\{y - x'\hat{\beta} - \hat{\psi}_{n\hat{\theta}}(z'\mathbb{S}(\hat{\gamma}))\}d\mathbb{P}_n(x,y,z) = 0, \tag{13}
$$

for $W = X$ and Z. Thus, $\phi_n(\hat{\theta})$ can be decomposed as

$$
\begin{aligned}
\phi_n(\hat{\theta}) &= T_n \int V_{I,n}^{x,z}\{y - x'\hat{\beta} - \hat{\psi}_{n\hat{\theta}}(z'\mathbb{S}(\hat{\gamma})\}d\mathbb{P}_n(x,y,z) + T_n \int V_{II,n}^{x,z}\{y - x'\hat{\beta} - \hat{\psi}_{n\hat{\theta}}(z'\mathbb{S}(\hat{\gamma})\}d\mathbb{P}_n(x,y,z) \\
&:= T_n(I + II),
\end{aligned}
\tag{14}
$$

where $T_n = \begin{bmatrix} \mathbb{I}_k & 0 \\ 0 & \mathbb{J}(\hat{\gamma})' \end{bmatrix}$,

$$
V_{I,n}^{x,z} = \begin{pmatrix} x - E[X|z'\mathbb{S}(\hat{\gamma})] \\ z - E[Z|z'\mathbb{S}(\hat{\gamma})] \end{pmatrix}, \quad V_{II,n}^{x,z} = \begin{pmatrix} E[X|z'\mathbb{S}(\hat{\gamma})] - \bar{E}_{n,\hat{\theta}}[X|z'\mathbb{S}(\hat{\gamma})] \\ E[Z|z'\mathbb{S}(\hat{\gamma})] - \bar{E}_{n,\hat{\theta}}[Z|z'\mathbb{S}(\hat{\gamma})] \end{pmatrix}.
$$

Step 2: Show $II = o_p(n^{-1/2}) + o_p(\hat{\theta} - \theta_0)$

Note that the term II can be decomposed as

$$
\begin{aligned}
II &= \int V_{II,n}^{x,z}\{y - x'\hat{\beta} - \hat{\psi}_{n\hat{\theta}}(z'\mathbb{S}(\hat{\gamma}))\}d(\mathbb{P}_n - P_0)(x,y,z) \\
&\quad + \int V_{II,n}^{x,z}\{y - x'\hat{\beta} - \psi_{\hat{\theta}}(z'\mathbb{S}(\hat{\gamma}))\}dP_0(x,y,z) + \int V_{II,n}^{x,z}\{\psi_{\hat{\theta}}(\mathbb{S}(\hat{\gamma})) - \hat{\psi}_{n\hat{\theta}}(z'\mathbb{S}(\hat{\gamma}))\}dP_0(x,y,z) \\
&:= II_a + II_b + II_c.
\end{aligned}
$$

First, we consider II_a. Note that Lemma 13 of BGH-supp and Lemma 1 imply the following (15) and (16), with probability approaching one:

$$
H_B(\varepsilon, \tilde{\mathcal{F}}_a, ||\cdot||_{B,P_0}) \le \frac{C_1}{\varepsilon},
\tag{15}
$$

for some $C_1 > 0$, where $\tilde{\mathcal{F}}_a = (C_2 \log n)^{-1} \mathcal{F}_a$ with some $C_2 > 0$ and $\mathcal{F}_a$ is defined in (41) below. Also, there exists a constant $C_3 > 0$ such that

$$
||\tilde{f}||_{B,P_0} \le C_3 (\log n) n^{-1/3},
\tag{16}
$$

for all $\tilde{f} \in \tilde{\mathcal{F}}_a$. Let $\delta_n = C_3(\log n)n^{-1/3}$ and $II_{a,j}$ be the j-th component of II_a. For any positive constants A and ν, there exist positive constants K_1, B_1, and B_2, such that $K = K_1 \log n$ and

$$
\begin{aligned}
P\{|II_{a,j}| > An^{-1/2}\} &= P\left\{|II_{a,j}| > An^{-1/2}, \sup_{\theta \in \mathcal{B}(\theta_0, \delta_0)} \sup_{z \in \mathcal{Z}} |\hat{\psi}_{n\theta}(z)| \leq K\right\} + \frac{\nu}{2} \\
&\leq P\left\{||\mathbb{G}_n||_{\mathcal{F}_a} > A, \sup_{\theta \in \mathcal{B}(\theta_0, \delta_0)} \sup_{z \in \mathcal{Z}} |\hat{\psi}_{n\theta}(z)| \leq K\right\} + \frac{\nu}{2} \\
&\leq \frac{E[||\mathbb{G}_n||_{\mathcal{F}_a} \mid \sup_{\theta \in \mathcal{B}(\theta_0, \delta_0)} \sup_{z \in \mathcal{Z}} |\hat{\psi}_{n\theta}(z)| \leq K]}{A} + \frac{\nu}{2} \\
&= \frac{1}{AC_2 \log n} E[||\mathbb{G}_n||_{\tilde{\mathcal{F}}_a} \mid \sup_{\theta \in \mathcal{B}(\theta_0, \delta_0)} \sup_{z \in \mathcal{Z}} |\hat{\psi}_{n\theta}(z)| \leq K] + \frac{\nu}{2} \\
&\lesssim \frac{1}{AC_2 \log n} J_n(\delta_n)\left(1 + \frac{J_n(\delta_n)}{\sqrt{n}\delta_n^2}\right) + \frac{\nu}{2} \\
&\lesssim \frac{\log n}{A}(\delta_n + 2B_1^{1/2}\delta_n^{1/2})\left(1 + \frac{\delta_n + 2B_1^{1/2}\delta_n^{1/2}}{\sqrt{n}\delta_n^2}\right) + \frac{\nu}{2} \\
&\lesssim \frac{1}{A}(\log n)^{3/2}n^{-1/6}\left(1 + \frac{B_2}{(\log n)^{3/2}}\right) + \frac{\nu}{2} \\
&\lesssim \nu,
\end{aligned}
\tag{17}
$$

for all n large enough, where the first equality follows from Lemma 8 in BGH-supp, the first inequality follows from the definition of $\mathcal{F}_a$ (in (41)), the second inequality follows from the Markov inequality, the second equality follows from the definition of $\tilde{\mathcal{F}}_a$, the first wave inequality ($\lesssim$) follows from van der Vaart and Wellner (1996, Lemma 3.4.3) and the definition of δ_n, the second wave inequality follows from (15) and Equation (.2) in BGH-supp, the third wave inequality follows from $\delta_n \lesssim \delta_n^{1/2}$ and the definition of δ_n. Therefore,

$$
II_a = o_p(n^{-1/2}).
\tag{18}
$$

Next, we consider II_b. Note that (see Lemma 17 in BGH-supp)

$$
\frac{\partial}{\partial \alpha_j} E[\psi_0(Z'\alpha_0)|Z'\alpha = z'\alpha]\bigg|_{\alpha = \alpha_0} = \{z_j - E[Z_j|Z'\alpha = z'\alpha_0]\}\psi_0'(z'\alpha_0),
\tag{19}
$$

for $j = 1, \ldots, d$. Using an expansion around $\hat{\gamma} = \gamma_0$ with (19) and $E[\psi_0(Z'\mathbb{S}(\gamma_0))|z'\mathbb{S}(\gamma_0)] = \psi_0(z'\mathbb{S}(\gamma_0))$, we have

$$
E[\psi_0(Z'\mathbb{S}(\gamma_0))|z'\mathbb{S}(\hat{\gamma})] = \psi_0(z'\mathbb{S}(\gamma_0)) + (\hat{\gamma} - \gamma_0)'\mathbb{J}(\hat{\gamma})'\{z - E[Z|z'\mathbb{S}(\gamma_0)]\}\psi_0'(z'\mathbb{S}(\gamma_0)) + o_p(\hat{\gamma} - \gamma_0).
\tag{20}
$$

16

Then we have

$$
\begin{aligned}
II_b &= \int V_{II,n}^{x,z} \left\{ \begin{array}{c} (\beta_0 - \hat{\beta})'\{x - E[X|z'\mathbb{S}(\hat{\gamma})]\} \\ +\psi_0(z'\mathbb{S}(\gamma_0)) - E[\psi_0(Z'\alpha_0)|z'\mathbb{S}(\hat{\gamma})] \end{array} \right\} dP_0(x,z) \\
&= \int V_{II,n}^{x,z} \left\{ \begin{array}{c} (\beta_0 - \hat{\beta})'\{x - E[X|z'\mathbb{S}(\hat{\gamma})]\} \\ -(\hat{\gamma} - \gamma_0)'\mathbb{J}(\gamma_0)'\{z - E[Z|z'\mathbb{S}(\gamma_0)]\}\psi_0'(z'\mathbb{S}(\gamma_0)) + o_p(\hat{\gamma} - \gamma_0) \end{array} \right\} dP_0(x,z) \\
&= -\int V_{II,n}^{x,z} \left(\begin{array}{c} x - E[X|z'\mathbb{S}(\gamma_0)] \\ \mathbb{J}(\gamma_0)'\{z - E[Z|z'\mathbb{S}(\gamma_0)]\}\psi_0'(z'\mathbb{S}(\gamma_0)) \end{array} \right)' dP_0(x,z) \left(\begin{array}{c} \hat{\beta} - \beta_0 \\ \hat{\gamma} - \gamma_0 \end{array} \right) + o_p(\hat{\gamma} - \gamma_0) \\
&= o_p(\hat{\theta} - \theta_0),
\end{aligned}
\tag{21}
$$

where the first equality follows from $E[\epsilon|X,Z] = 0$ and (10), the second equality follows from (20), and the last equality comes from $\int V_{II,n}^{x,z} dP_0(x,z) = o_p(1)$ and boundedness of the functions $x - E[X|z'\mathbb{S}(\gamma_0)]$ and $\mathbb{J}(\gamma_0)'\{\{z - E[Z|z'\mathbb{S}(\gamma_0)]\}\psi_0'(z'\mathbb{S}(\gamma_0))\}$.

Finally, we consider II_c. Since $E[W|z'\mathbb{S}(\gamma)]$ has totally bounded derivative for $W = X$ and Z by A4, there exists $C_0 > 0$ such that

$$
|E[W|Z'\mathbb{S}(\gamma) = u] - \bar{E}_{n,\theta}[W|Z'\mathbb{S}(\gamma) = u] \le C_0|\psi_\theta(u) - \hat{\psi}_{n\theta}(u)|,
\tag{22}
$$

for each $\theta \in \mathcal{B}(\theta_0, \delta_0)$ and $u \in I_\alpha$. By this, we obtain

$$
\begin{aligned}
||II_c|| &= ||\int V_{II,n}^{x,z}\{\psi_{\hat{\theta}}(z'\mathbb{S}(\hat{\gamma})) - \hat{\psi}_{n\hat{\theta}}(z'\mathbb{S}(\hat{\gamma}))\}dP_0(x,z)|| \\
&\lesssim \int \{\psi_{\hat{\theta}}(z'\mathbb{S}(\hat{\gamma})) - \hat{\psi}_{n\hat{\theta}}(z'\mathbb{S}(\hat{\gamma}))\}^2 dP_0(z) \\
&= O_p((\log)^2 n^{-2/3}) = o_p(n^{-1/2}),
\end{aligned}
\tag{23}
$$

uniformly in $\theta \in \mathcal{B}(\theta_0, \delta_0)$, where the second equality follows from Proposition 4 in BGH. Combining (18), (21), and (23), we conclude that

$$
II = o_p(n^{-1/2}) + o_p(\hat{\theta} - \theta_0).
\tag{24}
$$

Step 3: Decompose I

The term I can be decomposed as

$$
\begin{aligned}
I &= \int V_{I,n}^{x,z}\{y - x'\hat{\beta} - \psi_{\hat{\theta}}(z'\mathbb{S}(\hat{\gamma}))\}dP_0(x,y,z) + \int V_{I,n}^{x,z}\{y - x'\hat{\beta} - \psi_{\hat{\theta}}(z'\mathbb{S}(\hat{\gamma}))\}d(\mathbb{P}_n - P_0)(x,y,z) \\
&\quad + \int V_{I,n}^{x,z}\{\psi_{\hat{\theta}}(z'\mathbb{S}(\hat{\gamma})) - \hat{\psi}_{n\hat{\theta}}(z'\mathbb{S}(\hat{\gamma}))\}d\mathbb{P}_n(x,y,z) \\
&:= I_a + I_b + I_c.
\end{aligned}
\tag{25}
$$

In the following steps, we show that

$$T_n I_a = -T_0 \int V_{x,z} V'_{x,z\psi'} dP_0(x,z) T'_0 (\hat{\theta} - \theta_0) + o_p(\hat{\theta} - \theta_0), \tag{26}$$

$$T_n I_b = T_0 \int V_{x,z} \{y - x'\beta_0 - \psi_0(z'\mathbb{S}(\gamma_0)\} d(\mathbb{P}_n - P_0)(x,y,z)$$
$$+ o_p(\hat{\theta} - \theta_0) + o_p(n^{-1/2}), \tag{27}$$

$$I_c = o_p(n^{-1/2}). \tag{28}$$

Step 4: Show (26)

$$I_a = \int V_{I,n}^{x,z} \left\{ \begin{array}{c} (\beta_0 - \hat{\beta})'\{x - E[X|z'\mathbb{S}(\hat{\gamma})]\} \\ +\psi_0(z'\mathbb{S}(\gamma_0)) - E[\psi_0(Z'\alpha_0)|z'\mathbb{S}(\hat{\gamma})] \end{array} \right\} dP_0(x,z)$$

$$= \int V_{I,n}^{x,z} \left\{ \begin{array}{c} (\beta_0 - \hat{\beta})\{x - E[X|z'\mathbb{S}(\hat{\gamma})]\} \\ -(\hat{\gamma} - \gamma_0)'\mathbb{J}(\gamma_0)'\{z - E[Z|z'\mathbb{S}(\gamma_0)]\}\psi_0'(z'\mathbb{S}(\gamma_0)) + o_p(\hat{\gamma} - \gamma_0) \end{array} \right\} dP_0(x,z)$$

$$= -\int V_{x,z} V'_{x,z,\psi'} dP_0(x,z) T'_0 \left(\begin{array}{c} \hat{\beta} - \beta_0 \\ \hat{\gamma} - \gamma_0 \end{array} \right) + o_p(\hat{\gamma} - \gamma_0), \tag{29}$$

where the the first equality follows from $E[\epsilon|X,Z] = 0$ and (10), and some rearrangement, the second equality follows from (20), and the last equality follows from the definition of $V_{x,z,\psi'}$ and the fact that for $W = X$ or Z, we have $E[W|z'\mathbb{S}(\hat{\gamma})] - E[W|z'\mathbb{S}(\gamma_0)] = O_p(\hat{\gamma} - \gamma_0)$. Now, (26) follows by

$$T_n - T_0 = O_p(\hat{\gamma} - \gamma_0). \tag{30}$$

Step 5: Show (27)

Decompose

$$T_n I_b = T_n \int V_{I,n}^{x,z} \{y - x'\hat{\beta} - \psi_{\hat{\theta}}(z'\mathbb{S}(\hat{\gamma}))\} d(\mathbb{P}_n - P_0)(x,y,z)$$

$$= (T_n - T_0) \int V_{I,n}^{x,z} \{y - x'\hat{\beta} - \psi_{\hat{\theta}}(z'\mathbb{S}(\hat{\gamma}))\} d(\mathbb{P}_n - P_0)(x,y,z)$$

$$+ T_0 \int V_{I,n}^{x,z} \{x'\beta_0 - x'\hat{\beta} + \psi_0(z'\mathbb{S}(\gamma_0)) - \psi_{\hat{\theta}}(z'\mathbb{S}(\hat{\gamma}))\} d(\mathbb{P}_n - P_0)(x,y,z)$$

$$+ T_0 \int (V_{I,n}^{x,z} - V_{x,z})\{y - x'\beta_0 - \psi_0(z'\mathbb{S}(\gamma_0)\} d(\mathbb{P}_n - P_0)(x,y,z)$$

$$+ T_0 \int V_{x,z} \{y - x'\beta_0 - \psi_0(z'\mathbb{S}(\gamma_0)\} d(\mathbb{P}_n - P_0)(x,y,z)$$

$$:= (T_n - T_0) I_{b1} + T_0 I_{b2} + T_0 I_{b3}$$

$$+ T_0 \int V_{x,z} \{y - x'\beta_0 - \psi_0(z'\mathbb{S}(\gamma_0)\} d(\mathbb{P}_n - P_0)(x,y,z).$$

First, consider I_{b1}. Note that Lemma 13 BGH-supp and Lemma 2 imply the following (31) and (32):

$$H_B(\varepsilon, \mathcal{F}_{b1}, ||\cdot||_{B,P_0}) \le \frac{C_1}{\varepsilon}, \tag{31}$$

for some $C_1 > 0$, where $\mathcal{F}_{b1}$ is defined in (43). Also, there exists a constant $C_2 > 0$ such that

$$||f||_{B,P_0} \le C_2, \tag{32}$$

for all $f \in \mathcal{F}_{b1}$. Let $I_{b1,j}$ be the j-th component of I_{b1}. For any $A > 0$, there exists a positive constant C such that

$$P\{|I_{b1,j}| > An^{-1/2}\} \le \frac{1}{A}E[||\mathbb{G}_n||_{\mathcal{F}_{b1}}] \lesssim \frac{1}{A}J_n(C_2)\left(1 + \frac{J_n(C_2)}{\sqrt{n}C_2^2}\right) \lesssim \frac{C}{A},$$

for all n large enough, where the first inequality follows from the definition of $\mathcal{F}_{b1}$ and the Markov inequality, the first wave inequality follows from van der Vaart and Wellner (1996, Lemma 3.4.3), and the second wave inequality follows from (31), (32), and Equation (.2) in BGH. Thus, we have

$$I_{b1} = O_p(n^{-1/2}). \tag{33}$$

Next, consider I_{b2}. Let $I_{b2,j}$ be the j-th component of I_{b2}. For any positive constants A, ν, and η, there exist positive constants C', C_3, C_4, and C_5 such that

$$P\{|I_{b2,j}| > An^{-1/2}\} \le \frac{1}{A}E[||\mathbb{G}_n||_{\mathcal{F}_{b2}}|\mathfrak{B}_\eta] + \frac{\nu}{2} \lesssim \frac{1}{A}J_n(C'\eta)\left(1 + \frac{J_n(C'\eta)}{\sqrt{n}(C'\eta)^2}C_3\right) + \frac{\nu}{2}$$

$$\lesssim \frac{1}{A}C_4\eta^{1/2}\left(1 + \frac{C_5(1+\eta^{1/2})}{\sqrt{n}(C'\eta)^{3/2}}C_3\right) + \frac{\nu}{2}, \tag{34}$$

for all n large enough, where the event $\mathfrak{B}_\eta$ is defined in Lemma 3. The first inequality follows from Lemma 3, the definition of $\mathcal{F}_{b2}$ in (45), and the Markov inequality, the first wave inequality follows from van der Vaart and Wellner (1996, Lemma 3.4.2) and Lemma 3 (by choosing C' and η as therein), C_3 is a constant envelope of $\mathcal{F}_{b2}$, and the second wave inequality follows from Lemma 3 and Equation (.2) in BGH-supp. Since we can choose η arbitrarily small, it holds

$$I_{b2} = o_p(n^{-1/2}). \tag{35}$$

Finally, consider I_{b3}. This is similar to the case of I_{b1} but with one difference, $V_{I,n}^{x,z} - V_{x,z} = o_p(1)$. Therefore we can use the same methods as for I_{b2} to find a upper bound of the L_2-norm (as we did in the proof of Lemma 3 and (34).) Thus, we have

$$I_{b3} = o_p(n^{-1/2}). \tag{36}$$

Combining (33), (35), and (36) with (30), we obtain (27).

Step 6: Show (28)

Decompose

$$
\begin{aligned}
I_c &= \int V_{I,n}^{x,z} \{\psi_{\hat\theta}(z'\mathbb{S}(\hat\gamma)) - \hat\psi_{n\hat\theta}(z'\mathbb{S}(\hat\gamma))\} dP_0(x,y,z) \\
&\quad + \int V_{I,n}^{x,z} \{\psi_{\hat\theta}(z'\mathbb{S}(\hat\gamma)) - \hat\psi_{n\hat\theta}(z'\mathbb{S}(\hat\gamma))\} d(\mathbb{P}_n - P_0)(x,y,z) \\
&:= I_{c1} + I_{c2},
\end{aligned}
$$

For I_{c1}, the law of iterated expectation yields

$$
I_{c1} = E\left[E\left[\left. \begin{pmatrix} X - E[X|Z'\mathbb{S}(\hat\gamma)] \\ Z - E[Z|Z'\mathbb{S}(\hat\gamma)] \end{pmatrix} \right| Z'\mathbb{S}(\hat\gamma) \right] \{\psi_{\hat\theta}(Z'\mathbb{S}(\hat\gamma)) - \hat\psi_{n\hat\theta}(Z'\mathbb{S}(\hat\gamma))\} \right] = 0. \tag{37}
$$

Now consider I_{c2}. For any positive constants A and ν, there exist positive constants C_1, C_2, and C' such that

$$
\begin{aligned}
P\{|I_{c2}| > An^{-1/2}\} &\leq \frac{C_1}{A}(\log n)^{1/2}\eta_n^{1/2}\left(1 + \frac{C_1(\log n)^{3/2}\eta_n^{1/2}}{\sqrt{n}\eta_n^2}\right) + \frac{\nu}{2} \\
&\leq \frac{C_2}{A}(\log n)n^{-1/6} + \frac{\nu}{2} \leq \nu,
\end{aligned}
$$

for all n large enough and $\eta_n = C'(\log n)n^{-1/3}$, where the first inequality follows by Lemma 4 and a similar argument to (34), and the second inequality follows from the definition of η_n. Thus, we have $I_{c2} = o_p(n^{-1/2})$, and obtain (28).

Step 7: Conclusion

From Steps 1-6, we obtain

$$
\begin{aligned}
0 &= \phi_n(\hat\theta) \\
&= -T_0 \int V_{x,z} V'_{x,z,\psi'} dP_0(x,z) T'_0(\hat\theta - \theta_0) \\
&\quad + T_0 \int V_{x,z}\{y - x'\beta_0 - \psi_0(z'\mathbb{S}(\gamma_0))\} d(\mathbb{P}_n - P_0)(x,y,z) + o_p(n^{-1/2}) + o_p(\hat\theta - \theta_0).
\end{aligned}
$$

With B defined in A7, the central limit theorem implies

$$
\begin{aligned}
\sqrt{n}(\hat\theta - \theta_0) &= \sqrt{n}B^{-1}T_0 \int V_{x,z}\{y - x'\beta_0 - \psi_0(z'\mathbb{S}(\gamma_0))\} d(\mathbb{P}_n - P_0)(x,y,z) \\
&\quad + o_p(1 + \sqrt{n}(\hat\theta - \theta_0)) \\
&\xrightarrow{d} N(0,\Pi).
\end{aligned} \tag{38}
$$

A.3 Lemmas

In this subsection, we use the following notations:

$$
\begin{aligned}
\mathcal{M}_{RK} &= \{\text{monotone non-decreasing functions on } [-R, R] \text{ and bounded by } K\}, \\
\mathcal{G}_{RK} &= \{g : g(z) = \psi_\theta(\alpha' z), z \in \mathcal{Z}, (\psi, \theta) \in \mathcal{M}_{RK} \times \mathcal{B}(\theta_0, \delta_0)\}, \\
\mathcal{D}_{RKv} &= \{d : d(z) = g_1(z) - g_2(z), \ (g_1, g_2) \in \mathcal{G}_{RK}^2, \ \|d(z)\|_{P_0} \leq v\}, \\
\mathcal{H}_{RKv} &= \{h : h(\tilde{y}, z) = \tilde{y} d_1(z) - d_2(z), \ (d_1, d_2) \in \mathcal{D}_{RKv}^2, \ (\tilde{y}, z) \in \mathbb{R} \times \mathcal{Z}\}.
\end{aligned}
\tag{39}
$$

A.3.1 Lemma for II_a

Let W_j be the j-th component of X or Z. Then decompose

$$
\begin{aligned}
&\{E[W_j | z'\mathbb{S}(\hat{\gamma})] - \bar{E}_{n,\hat{\theta}}[W_j | z'\mathbb{S}(\hat{\gamma})]\}\{y - x\hat{\beta} - \hat{\psi}_{n\hat{\theta}}(z'\mathbb{S}(\hat{\gamma}))\} \\
=\ & \{E[W_j | z'\mathbb{S}(\hat{\gamma})] - \bar{E}_{n,\hat{\theta}}[W_j | z'\mathbb{S}(\hat{\gamma})]\}\{y - x\hat{\beta}\} \\
& - \{E[W_j | z'\mathbb{S}(\hat{\gamma})] - \bar{E}_{n,\hat{\theta}}[W_j | z'\mathbb{S}(\hat{\gamma})]\}\hat{\psi}_{n\hat{\theta}}(z'\mathbb{S}(\hat{\gamma})) \\
:=\ & d_1(z)\{y - x\hat{\beta}\} - d_2(z).
\end{aligned}
\tag{40}
$$

Let

$$
\mathcal{F}_a = \{f : f(x, y, z) = d_1(z)\{y - x\hat{\beta}\} - d_2(z), \ (x, y, z) \in \mathcal{X} \times \mathbb{R} \times \mathcal{Z}\},
\tag{41}
$$

be a function class of the integrand of II_a. To control the term II_a, we use the following lemma.

Lemma 1. *For some $K' \simeq \log n$ and positive constant v, it holds*

$$
\mathcal{F}_a \subset \mathcal{H}_{RK'v},
$$

with probability approaching one.

Proof. We use the following facts.

a) By A4, $E[W_j | z'\mathbb{S}(\hat{\gamma})]$ is a bounded function with a finite total variation.

b) $\bar{E}_{n,\hat{\theta}}[W_j | z'\mathbb{S}(\hat{\gamma})]$ is a discrete version of $E[W_j | z'\mathbb{S}(\hat{\gamma})]$ takes finite different values from it, so it is also bounded and has a finite total variation.

c) By Lemma 8 in BGH-supp, $\max_{\hat{\theta} \in \mathcal{B}(\theta_0, \delta_0)} \sup_{z \in \mathcal{Z}} |\hat{\psi}_{n\hat{\theta}}(z'\mathbb{S}(\hat{\gamma}))| = O_p(\log n)$. Thus, there exists $K = K_1 \log n$ such that $\hat{\psi}_{n\hat{\theta}} \in \mathcal{M}_{RK}$ with probability approaching to 1.

d) By Proposition 4 in BGH and (22), $\|E[W_j | z'\mathbb{S}(\hat{\gamma})] - \bar{E}_{n,\hat{\theta}}[W_j | z'\mathbb{S}(\hat{\gamma})]\|_2 \leq C_1 (\log n) n^{-1/3}$ for some $C_1 > 0$.

e) The addition or multiplication of two functions with finite total variations is a function with a finite total variation.

Then by Jordan's decomposition and a), b), d), and e), there exist a positive constant C_0 larger than twice the bound of $E[W_j|z'\mathbb{S}(\hat{\gamma})]$ and $v_1 = C_1(\log n)n^{-1/3}$ such that

$$d_1(\cdot) \in \mathcal{D}_{RC_0 v_1}, \tag{42}$$

with probability approaching 1. Additionally, c) and d) imply $d_2(\cdot) \in \mathcal{D}_{RK'v}$ with $K' = K_2 \log n$ for a large enough constant $K_2 > 0$ and $v = C_2(\log n)^2 n^{-1/3}$ for some $C_2 > 0$. Now, since $v_1 \lesssim v$ and $C_0 \lesssim K'$, setting $\tilde{y} = y - x\hat{\beta}$ in the definition of $\mathcal{H}_{RKv}$ in (39) yields the conclusion. $\square$

A.3.2 Lemma for I_{b1}

Let W_j (and w_j) be the j-th component of X or Z (x or z), $\tilde{y} = y - x\hat{\beta}$ as in Lemma 1, and

$$\mathcal{F}_{b1} = \big\{ f : f(w_j, y, z) = \{w_j - E[W_j|z'\mathbb{S}(\hat{\gamma})]\}\{\tilde{y} - \psi_{\hat{\theta}}(z'\mathbb{S}(\hat{\gamma})\}, \ (w_j, y, z) \in \mathcal{W}_j \times \mathbb{R} \times \mathcal{Z} \big\}, \tag{43}$$

be a function class of the j-th component of the integrand of I_{b1}. To control the term I_{b1}, we use the following lemma.

Lemma 2. *For some positive constants C and v, it holds*

$$\mathcal{F}_{b1} \subset \mathcal{H}_{RCv},$$

with probability approaching 1.

Proof. We use the following facts.

   a) w_j is bounded by $[-R, R]$.

   b) By A4, $E[W_j|z'\mathbb{S}(\hat{\gamma})]$ is a function bounded by $[-R, R]$ and has a finite total variation.

   c) By A1, A3, and (10), $\psi_{\hat{\theta}}$ is a bounded monotone function.

Let $d_1(z'\mathbb{S}(\hat{\gamma})) = E[W_j|z'\mathbb{S}(\hat{\gamma})]$ and $d_2(z'\mathbb{S}(\hat{\gamma})) = E[W_j|z'\mathbb{S}(\hat{\gamma})]\psi_{\hat{\theta}}(z'\mathbb{S}(\hat{\gamma}))$. Any function in $\mathcal{F}_{b1}$ can be expressed as

$$\{w_j - E[W_j|z'\mathbb{S}(\hat{\gamma})]\}\{y - x'\hat{\beta} - \psi_{\hat{\theta}}(z'\mathbb{S}(\hat{\gamma})\}$$
$$= w_j\{y - x'\hat{\beta} - \psi_{\hat{\theta}}(z'\mathbb{S}(\hat{\gamma})\} + d_1(z'\mathbb{S}(\hat{\gamma}))(y - x'\hat{\beta}) - d_2(z'\mathbb{S}(\hat{\gamma})). \tag{44}$$

By b) and c), we have

$$d_1(\cdot) \in \mathcal{D}_{RC_0 v_1},$$

for C_0 defined in (42), which is larger than twice the bound of $E[W_j|z'\mathbb{S}(\hat{\gamma})]$, and some v_1, which is larger than the L_2-norm of a constant function R (the upper bound in A1) on a compact support. Additionally, we have

$$d_2(\cdot) \in \mathcal{D}_{RC_1 v_2},$$

for some positive constants C_1 and v_2. Therefore, by setting $\tilde{y} = y - x\hat{\beta}$ in the definition of $\mathcal{H}_{RKv}$ in (39), the second and third terms in (44) satisfy

$$d_1(z'\mathbb{S}(\hat{\gamma}))(y - x'\hat{\beta}) - d_2(z'\mathbb{S}(\hat{\gamma})) \in \mathcal{H}_{RC_1 v_1}.$$

With similar steps we have:

$$w_j\{y - x'\hat{\beta} - \psi_{\hat{\theta}}(z'\mathbb{S}(\hat{\gamma})\} \in \mathcal{H}_{RC_1' v_1'},$$

for some positive constants C_1' and v_1'. By choosing $C \geq \max(C_1, C_1')$ and $v \geq \max(v_1, v_1')$, the conclusion follows. $\qquad\square$

A.3.3 Lemma for I_{b2}

Let

$$\mathcal{F}_{b2} = \big\{f : f(w_j, x, z) = \{w_j - E[W_j|z'\mathbb{S}(\hat{\gamma})]\}\{x'\beta_0 - x'\hat{\beta} + \psi_0(z'\mathbb{S}(\gamma_0)) - \psi_{\hat{\theta}}(z'\mathbb{S}(\hat{\gamma}))\}, \ (w_j, x, z) \in \mathcal{W}_j \times \mathcal{X} \times \mathcal{Z}\big\}, \tag{45}$$

be a function class of the integrand of $I_{b2,j}$, the j-th component of I_{b2}. To control the term I_{b2}, we use the following lemma.

Lemma 3.

For any positive constant η, we define the event $\mathfrak{B}_\eta$ as

$$\mathfrak{B}_\eta = \left\{ \sup_{x,z \in \mathcal{X} \times \mathcal{Z}, \hat{\theta} \in \mathcal{B}(\theta_0, \delta_0)} |x'\beta_0 - x'\hat{\beta} + \psi_0(z'\mathbb{S}(\gamma_0)) - \psi_{\hat{\theta}}(z'\mathbb{S}(\hat{\gamma}))| \leq \eta \right\}$$

1. For some $C > 0$, it holds $H_B(\varepsilon, \mathcal{F}_{b2}, ||\cdot||_{P_0}) \leq \frac{C}{\varepsilon}$.

2. For any positive constants ν and η, it holds $P(\mathfrak{B}_\eta) \geq 1 - \frac{\nu}{2}$ for all n large enough.

3. In case of the event $\mathfrak{B}_\eta$, there exists $C' > 0$ such that $||f||_2 \leq C'\eta$ for all $f \in \mathcal{F}_{b2}$.

Proof. Both $E[W_j|z'\mathbb{S}(\hat{\gamma})]$ and $\psi_0(z'\mathbb{S}(\gamma_0)) - \psi_{\hat{\theta}}(z'\mathbb{S}(\hat{\gamma}))$ are bounded functions with finite total variations. Thus, they should have entropy of order $\frac{C_1}{\varepsilon}$ for some $C_1 > 0$. Also, both w_j and $(x'\beta_0 - x'\hat{\beta})$ are bounded. Thus, they should have entropy of order $\frac{C_2}{\varepsilon}$ for some $C_2 > 0$ (see, Example 19.7 in van der Vaart, 2000). Combining these results, the statement (1) follows. The consistency of $\hat{\theta}$ and Lemma 19 of BGH-supp imply the statement (2). The statement (3) follows from the definition of $\mathcal{F}_{b2}$. $\qquad\square$

A.3.4 Lemma for I_{c2}

Let

$$\mathcal{F}_{c2} = \big\{f : f(w_j, z) = \{w_j - E[W_j|z'\mathbb{S}(\hat{\gamma})]\}\{\psi_{\hat{\theta}}(z'\mathbb{S}(\hat{\gamma})) - \hat{\psi}_{n\hat{\theta}}(z'\mathbb{S}(\hat{\gamma}))\}, \ (w_j, z) \in \mathcal{W}_j \times \mathcal{Z}\big\}, \tag{46}$$

be a function class of the integrand of $I_{c2,j}$, the j-th component of I_{c2}. To control the term I_{c2}, we use the following lemma.

Lemma 4.

1. *For some $C > 0$, it holds $H_B(\varepsilon, \mathcal{F}_{c2}, ||\cdot||_{P_0}) \leq \frac{C \log n}{\varepsilon}$ with probability approaching 1.*

2. *There exists a $C' > 0$ such that $||f||_{P_0} \leq C'(\log n)n^{-1/3}$ for all $f \in \mathcal{F}_{c2}$.*

Proof. We use the following facts.

a) w_j is bounded by $[-R, R]$.

b) By A4, $E[W_j | z'\mathbb{S}(\hat{\gamma})]$ is a function bounded by $[-R, R]$ and has a finite total variation.

c) By A1, A3, and (10), $\psi_{\hat{\theta}}$ is a bounded monotone function.

d) By Lemma 8 in BGH-supp, $\sup_{z \in \mathcal{Z}} |\hat{\psi}_{n\hat{\theta}}(z'\mathbb{S}(\hat{\gamma}))| = O_p(\log n)$. Therefore there exists $K = K_1 \log n$ such that $\hat{\psi}_{n\hat{\theta}} \in \mathcal{M}_{RK}$ with probability approaching to 1.

So, in the case that $\hat{\psi}_{n\hat{\theta}} \in \mathcal{M}_{RK}$:

1) $\{\psi_{\hat{\theta}}(z'\mathbb{S}(\hat{\gamma})) - \hat{\psi}_{n\hat{\theta}}(z'\mathbb{S}(\hat{\gamma}))\}$ is bounded by $K + R$ with a finite variation.

2) $E[W_j | z'\mathbb{S}(\hat{\gamma})]\{\psi_{\hat{\theta}}(z'\mathbb{S}(\hat{\gamma})) - \hat{\psi}_{n\hat{\theta}}(z'\mathbb{S}(\hat{\gamma}))\}$ is bounded by $R(K + R)$ with a finite variation, and the function class has an entropy of order $\frac{C_1 \log n}{\varepsilon}$ for some $C_1 > 0$.

3) From Lemma 10 of BGH-supp (by taking w_j as β in that lemma) and 1) above, the function class of $w_j\{\psi_{\hat{\theta}}(z'\mathbb{S}(\hat{\gamma})) - \hat{\psi}_{n\hat{\theta}}(z'\mathbb{S}(\hat{\gamma}))\}$ has an entropy of order $\frac{C_2 \log n}{\varepsilon}$ for some $C_2 > 0$.

From 2) and 3), the conclusion follows. $\square$

B   Proof of Theorem 2

Existence and consistency of $\tilde{\theta}$ can be shown similarly as in Appendix A.1. The rest of the proof is split into several steps.

Step 1: Derive a decomposition of $\xi_{nh}(\tilde{\theta})$

In the same spirit of Step 1 of Appendix A.2, we introduce a piecewise constant function $\bar{\rho}_{n,\theta}$. Let $\{u_{n_j}\}_{j=1}^k$ be all the jump points of the monotone LSE $\hat{\psi}_{n\theta}(u)$. We define for $u \in [u_{n_j}, u_{n_{j+1}})$

(if $j = k$, set $u_{n_{j+1}} = \max_i u_{n_i}$)

$$\bar{\rho}_{n,\theta}(W|u) = \bar{\rho}_{n,\theta}(W|Z'\mathbb{S}(\gamma))$$

$$= \begin{cases} \bar{\rho}_{n,\theta}(X|u) = \begin{cases} E[X|Z'\mathbb{S}(\gamma) = u_{n_j}] & \text{if } \psi_\theta(u) > \hat{\psi}_{n\theta}(u_{n_j}) \text{ for all } u \in (u_{n_j}, u_{n_{j+1}}), \\ E[X|Z'\mathbb{S}(\gamma) = s] & \text{if } \psi_\theta(u) = \hat{\psi}_{n\theta}(s) \text{ for some } s \in (u_{n_j}, u_{n_{j+1}}), \\ E[X|Z'\mathbb{S}(\gamma) = u_{n_{j+1}}] & \text{if } \psi_\theta(u) < \hat{\psi}_{n\theta}(u_{n_j}) \text{ for all } u \in (u_{n_j}, u_{n_{j+1}}), \end{cases} \\ \bar{\rho}_{n,\theta}(Z|u) = \begin{cases} E[Z|Z'\mathbb{S}(\gamma) = u_{n_j}]\psi'_\theta(u_{n_j}) & \text{if } \psi_\theta(u) > \hat{\psi}_{n\theta}(u_{n_j}) \text{ for all } u \in (u_{n_j}, u_{n_{j+1}}), \\ E[Z|Z'\mathbb{S}(\gamma) = s]\psi'_\theta(s) & \text{if } \psi_\theta(u) = \hat{\psi}_{n\theta}(s) \text{ for some } s \in (u_{n_j}, u_{n_{j+1}}), \\ E[Z|Z'\mathbb{S}(\gamma) = u_{n_{j+1}}]\psi'_\theta(u_{n_{j+1}}) & \text{if } \psi_\theta(u) < \hat{\psi}_{n\theta}(u_{n_j}) \text{ for all } u \in (u_{n_j}, u_{n_{j+1}}). \end{cases} \end{cases}$$

Similar to (22), we have for each $\theta \in \mathcal{B}(\theta_0, \delta_0)$

$$|E[Z|Z'\mathbb{S}(\gamma) = u]\psi'_\theta(u) - \bar{\rho}_{n,\theta}(Z|u)| \leq C_0|\psi_\theta(u) - \hat{\psi}_{n\theta}(u)|. \tag{47}$$

Similar to (13), we have

$$\int \bar{\rho}_{n,\tilde{\theta}}(W|z'\mathbb{S}(\gamma))\{y - x'\tilde{\beta} - \hat{\psi}_{n\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma}))\}d\mathbb{P}_n(x, y, z) = 0,$$

for $W = X$ and Z. Thus, $\xi_{nh}(\tilde{\theta})$ can be decomposed as

$$\xi_{nh}(\tilde{\theta}) = T_n \int V^{x,z}_{I,nh,\psi'}\{y - x'\tilde{\beta} - \hat{\psi}_{n\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma})\}d\mathbb{P}_n(x, y, z) + T_n \int V^{x,z}_{II,n}\{y - x'\tilde{\beta} - \hat{\psi}_{n\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma})\}d\mathbb{P}_n(x, y, z)$$

$$:= T_n(I^E + II^E), \tag{48}$$

where $T_n = \begin{bmatrix} \mathbb{I}_k & 0 \\ 0 & \mathbb{J}(\tilde{\gamma})' \end{bmatrix}$,

$$V^{x,z}_{I,nh,\psi'} = \begin{pmatrix} x - E[X|z'\mathbb{S}(\tilde{\gamma})] \\ z\hat{\psi}'_{nh,\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma})) - E[Z|z'\mathbb{S}(\tilde{\gamma})]\psi'_{\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma})) \end{pmatrix}, \quad V^{x,z}_{I,n,\psi'} = \begin{pmatrix} x - E[X|z'\mathbb{S}(\tilde{\gamma})] \\ [z - E[Z|z'\mathbb{S}(\tilde{\gamma})]]\psi'_{\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma})) \end{pmatrix},$$

$$V_{x,z,\psi'} = \begin{pmatrix} x - E[X|z'\mathbb{S}(\gamma_0)] \\ [z - E[Z|z'\mathbb{S}(\gamma_0))]\psi'_0(z'S(\gamma_0)) \end{pmatrix}, \tag{49}$$

$$V^{x,z}_{II,n} = \begin{pmatrix} E[X|z'\mathbb{S}(\tilde{\gamma})] - \bar{\rho}_{n,\tilde{\theta}}(X|z'\mathbb{S}(\tilde{\gamma})) \\ E[Z|z'\mathbb{S}(\tilde{\gamma})]\psi'_\theta(z'\mathbb{S}(\tilde{\gamma}) - \bar{\rho}_{n,\tilde{\theta}}(Z|z'\mathbb{S}(\tilde{\gamma})) \end{pmatrix}.$$

Note: T_n and $V^{x,z}_{II,n}$ are redefined for $\tilde{\theta}$ in Appendix B.

Step 2: Show $II^E = o_p(n^{-1/2}) + o_p(\tilde{\theta} - \theta_0)$

Decompose

$$
\begin{aligned}
II^E &= \int V_{II,n}^{x,z}\{y - x'\tilde{\beta} - \hat{\psi}_{n\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma}))\}d(\mathbb{P}_n - P_0)(x,y,z) \\
&\quad + \int V_{II,n}^{x,z}\{y - x'\tilde{\beta} - \psi_{\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma}))\}dP_0(x,y,z) + \int V_{II,n}^{x,z}\{\psi_{\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma})) - \hat{\psi}_{n\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma}))\}dP_0(x,y,z) \\
&:= II_a^E + II_b^E + II_c^E.
\end{aligned}
$$

First, we consider II_a^E. By A8, $\psi_\theta'(z'\mathbb{S}(\tilde{\gamma}))$ is uniformly bounded with a bounded total variation. Therefore, $E[Z|z'\mathbb{S}(\tilde{\gamma})]\psi_\theta'(z'\mathbb{S}(\tilde{\gamma}))$ is also uniformly bounded with a bounded total variation, and all the arguments in Step 2 of Appendix A.2 can be applied to show $II_a^E = o_p(n^{-1/2})$.

Next, we consider II_b^E. For the redefined $V_{II,n}^{x,z}$, we still have $\int V_{II,n}^{x,z}dP_0(x,z) = o_p(1)$ and boundedness of the functions $x - E[X|z'\mathbb{S}(\gamma_0)]$ and $\mathbb{J}(\gamma_0)'\{\{z - E[Z|z'\mathbb{S}(\gamma_0)]\}\psi_0'(z'\mathbb{S}(\gamma_0))\}$. Thus the same argument as in in Step 2 of Appendix A.2 yields $II_b^E = o_p(\tilde{\theta} - \theta_0)$.

Finally, we consider II_c^E. By (22) and (47), the same argument in Step 2 of Appendix A.2 implies $II_c^E = o_p(n^{-1/2})$. Combining these results, we obtain $II^E = o_p(n^{-1/2}) + o_p(\tilde{\theta} - \theta_0)$.

Step 3: Decompose I^E

Note that

$$
\begin{aligned}
I^E &= T_n \int V_{I,nh,\psi'}^{x,z}\{y - x'\tilde{\beta} - \hat{\psi}_{n\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma})\}d\mathbb{P}_n(x,y,z) \\
&= \int V_{I,nh,\psi'}^{x,z}\{y - x'\tilde{\beta} - \psi_{\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma})\}dP_0(x,y,z) + \int V_{I,nh,\psi'}^{x,z}\{y - x'\tilde{\beta} - \psi_{\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma})\}d(\mathbb{P}_n - P_0)(x,y,z) \\
&\quad + \int V_{I,nh,\psi'}^{x,z}\{\psi_{\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma})) - \hat{\psi}_{n\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma}))\}d\mathbb{P}_n(x,y,z) \\
&:= I_a^E + I_b^E + I_c^E
\end{aligned}
$$

In the following steps, we show that

$$
\begin{aligned}
T_n I_a^E &= -T_0 \int V_{x,z,\psi'}V_{x,z\psi'}' dP_0(x,z)T_0'(\tilde{\theta} - \theta_0) + o_p(\tilde{\theta} - \theta_0), &&(50) \\
T_n I_b^E &= T_0 \int V_{x,z,\psi'}\{y - x'\beta_0 - \psi_0(z'\mathbb{S}(\gamma_0)\}d(\mathbb{P}_n - P_0)(x,y,z) \\
&\quad + o_p(\tilde{\theta} - \theta_0) + o_p(n^{-1/2}), &&(51) \\
I_c^E &= o_p(n^{-1/2}). &&(52)
\end{aligned}
$$

Step 4: Show (50)

Decompose

$$
\begin{aligned}
I_a^E &= \int V_{I,n,\psi'}^{x,z}\{y - x'\tilde{\beta} - \psi_{\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma})\}dP_0(x,y,z) \\
&\quad + \int \begin{pmatrix} 0 \\ z[\hat{\psi}'_{nh,\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma})) - \psi'_{\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma}))] \end{pmatrix}\{y - x'\tilde{\beta} - \psi_{\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma})\}dP_0(x,y,z) \\
&:= I_{a1}^E + I_{a2}^E.
\end{aligned}
$$

By a similar argument as in (29), we have

$$
I_{a1}^E = -\left\{ \int V_{x,z,\psi'} V'_{x,z,\psi'} dP_0(x,z) \right\} T'_0(\tilde{\theta} - \theta_0) + o_p(\tilde{\theta} - \theta_0). \tag{53}
$$

and

$$
I_{a2}^E = -\left\{ \int \begin{pmatrix} 0 \\ z\{\hat{\psi}'_{nh,\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma})) - \psi'_{\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma}))\} \end{pmatrix} V'_{x,z,\psi'} dP_0(x,z) \right\} T'_0(\tilde{\theta} - \theta_0) + o_p(\tilde{\theta} - \theta_0).
$$

From $\hat{\psi}'_{nh,\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma})) - \psi'_{\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma})) = o_p(1)$, $V'_{x,z,\psi'} = O_p(1)$, and the compact supports of x and z, it holds $I_{a2}^E = o_p(\tilde{\theta} - \theta_0)$. Thus, we obtain (50).

Step 5: Show (51)

Decompose

$$
\begin{aligned}
T_n I_b^E &= T_n \int V_{I,n,\psi'}^{x,z}\{y - x'\tilde{\beta} - \psi_{\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma})\}d(\mathbb{P}_n - P_0)(x,y,z) \\
&\quad + T_n \int \begin{pmatrix} 0 \\ z\{\tilde{\psi}'_{nh,\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma})) - \psi'_{\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma}))\} \end{pmatrix}\{y - x'\tilde{\beta} - \psi_{\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma})\}d(\mathbb{P}_n - P_0)(x,y,z) \\
&:= T_n I_{b1}^E + T_n I_{b2}^E.
\end{aligned}
$$

By similar steps as in Step 5 of Appendix A.2 combined with A8, we can derive

$$
T_n I_{b1}^E = T_0 \int V_{x,z,\psi'}\{y - x'\beta_0 - \psi_0(z'\mathbb{S}(\gamma_0)\}d(\mathbb{P}_n - P_0)(x,y,z) + o_p(\tilde{\theta} - \theta_0) + o_p(n^{-1/2}). \tag{54}
$$

By Lemma 23 in BGH-supp, the analysis for $T_n I_{b2}^E$ is similar to the one for I_{b3} in Step 5 of Appendix A.2. Therefore, we have $T_n I_{b2}^E = o_p(n^{-1/2})$, and (51) is obtained.

Step 6: Show (52)

Decompose

$$
\begin{aligned}
I_c^E &= \int V_{I,nh,\psi'}^{x,z}\{\psi_{\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma})) - \hat{\psi}_{n\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma}))\}dP_0(x,y,z) \\
&\quad + \int V_{I,nh,\psi'}^{x,z}\{\psi_{\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma})) - \hat{\psi}_{n\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma}))\}d(\mathbb{P}_n - P_0)(x,y,z) \\
&= I_{c1}^E + I_{c2}^E.
\end{aligned}
$$

For I_{c1}^E, note that

$$
\begin{aligned}
I_{c1}^E &= \int V_{I,n,\psi'}^{x,z}\{\psi_{\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma})) - \hat{\psi}_{n\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma}))\}dP_0(x,y,z) \\
&\quad + \int \begin{pmatrix} 0 \\ z\{\hat{\psi}'_{nh,\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma})) - \psi'_{\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma}))\} \end{pmatrix} \{\psi_{\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma})) - \hat{\psi}_{n\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma}))\}dP_0(x,y,z) \\
&= \int \begin{pmatrix} 0 \\ E[Z|u]\left\{\frac{1}{h}\int K\left(\frac{u-x}{h}\right)d\hat{\psi}_{n\tilde{\theta}}(x) - \psi'_{\tilde{\theta}}(u)\right\} \end{pmatrix} \{\psi_{\tilde{\theta}}(u) - \hat{\psi}_{n\tilde{\theta}}(u)\}dP_0(u), \quad (55)
\end{aligned}
$$

where the last equality follows from a similar argument in (37), a change of variables $u = z'\mathbb{S}(\tilde{\gamma})$, and the definition of $\hat{\psi}_{nh,\tilde{\theta}}(u)$. We know $E[Z|u] = O(1)$ and $\int\{\psi_{\tilde{\theta}}(u) - \hat{\psi}_{n\tilde{\theta}}(u)\}^2 dP_0(u) = O_p((\log n)^2 n^{-2/3})$ by Proposition 4 in BGH. Also note that

$$
\begin{aligned}
&\frac{1}{h}\int K\left(\frac{u-x}{h}\right)d\hat{\psi}_{n\tilde{\theta}}(x) - \psi'_{\tilde{\theta}}(u) \\
&= \frac{1}{h}\int K\left(\frac{u-x}{h}\right)d(\hat{\psi}_{n\tilde{\theta}}(x) - \psi_{\tilde{\theta}}(x)) + \frac{1}{h}\int K\left(\frac{u-x}{h}\right)d\psi_{\tilde{\theta}}(x) - \psi'_{\tilde{\theta}}(u) \\
&= -\frac{1}{h^2}\int K'\left(\frac{u-x}{h}\right)(\hat{\psi}_{n\tilde{\theta}}(x) - \psi_{\tilde{\theta}}(x))dx + \frac{1}{h}\int K\left(\frac{u-x}{h}\right)d\psi_{\tilde{\theta}}(x) - \psi'_{\tilde{\theta}}(u), \quad (56)
\end{aligned}
$$

where the second equality follows from integration by parts and A9. With small h, $\frac{1}{h^2}\int K'\left(\frac{u-x}{h}\right)(\hat{\psi}_{n\tilde{\theta}}(x) - \psi_{\tilde{\theta}}(x))dx \sim \frac{1}{h}(\hat{\psi}_{n\tilde{\theta}}(u) - \psi_{\tilde{\theta}}(u))$. And $\frac{1}{h}\int K\left(\frac{u-x}{h}\right)d\psi_{\tilde{\theta}}(x) - \psi'_{\tilde{\theta}}(u)$ is a typical bias term of a kernel estimator, which is of order h^2 by A9. Plugging (56) into (55), the Cauchy-Schwarz inequality and A9 imply

$$
I_{c1}^E = O_p((\log n)^2 n^{-2/3}) \cdot O_p(n^{1/7}) + O_p((\log n)n^{-1/3}) \cdot O_p(n^{-2/7}) = o_p(n^{-1/2}). \quad (57)
$$

For I_{c2}^E, A8 and Lemma 23 in BGH-supp imply that both $z\hat{\psi}'_{nh,\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma}))$ and $E[Z|z'\mathbb{S}(\tilde{\gamma})]\psi'_{\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma}))$ are bounded with finite total variation. By a similar argument to Step 6 of Appendix A.2, we have $I_{c2}^E = o_p(n^{-1/2})$. Combined with (57), we obtain (52).

Step 7: Conclusion

From Steps 1-6 above, we obtain

$$
\begin{aligned}
0 &= \xi_{nh}(\tilde{\theta}) \\
&= -T_0 \int V_{x,z,\psi'} V'_{x,z,\psi'} dP_0(x,z) T'_0(\tilde{\theta} - \theta_0) \\
&\quad + T_0 \int V_{x,z,\psi'}\{y - x'\beta_0 - \psi_0(z'\mathbb{S}(\gamma_0))\} d(\mathbb{P}_n - P_0)(x,y,z) + o_p(\tilde{\theta} - \theta_0) + o_p(n^{-1/2}).
\end{aligned}
$$

With B_E defined in A7, the central limit theorem implies

$$
\begin{aligned}
\sqrt{n}(\tilde{\theta} - \theta_0) &= \sqrt{n} B_E^{-1} T_0 \int V_{x,z,\psi'}\{y - x'\beta_0 - \psi_0(z'\mathbb{S}(\gamma_0))\} d(\mathbb{P}_n - P_0)(x,y,z) \\
&\quad + o_p(1 + \sqrt{n}(\tilde{\theta} - \theta_0)) \\
&\stackrel{d}{\to} N(0, \Pi_E).
\end{aligned}
$$

C Proof of Theorem 3

Here we adapt the relevant proof in Groeneboom and Hendrickx (2017) (hereafter GH) to the monotone partially linear single index model. Let $\phi_n^*(\cdot)$ be the score function in the bootstrap sample. By definition (6),

$$
\phi_n^*(\hat{\theta}^*) = \int \begin{pmatrix} x \\ \mathbb{J}(\hat{\gamma}^*)'z \end{pmatrix} \{y - x'\hat{\beta}^* - \hat{\psi}^*_{n\hat{\theta}^*}(z'\mathbb{S}(\hat{\gamma}^*))\} d\hat{\mathbb{P}}_n(x,y,z),
$$

where $\hat{\mathbb{P}}_n$ is the empirical measure. Suppose

$$
\begin{aligned}
\phi_n^*(\hat{\theta}^*) &= -B(\hat{\theta}^* - \theta_0) + T_0 \int V_{x,z}\{y - x'\beta_0 - \psi_0(z'\mathbb{S}(\gamma_0))\} d(\hat{\mathbb{P}}_n - \mathbb{P}_n)(x,y,z) \qquad (58) \\
&\quad + T_0 \int V_{x,z}\{y - x'\beta_0 - \psi_0(z'\mathbb{S}(\gamma_0))\} d(\mathbb{P}_n - P_0)(x,y,z) + o_{P_M}(n^{-1/2} + (\hat{\theta}^* - \theta_0)).
\end{aligned}
$$

where P_M is defined in p. 3450 of GH. Then with $\phi_n^*(\hat{\theta}^*) = 0$ and (38), we have

$$
\begin{aligned}
\sqrt{n}(\hat{\theta}^* - \hat{\theta}) &= \sqrt{n} B^{-1} T_0 \int V_{x,z}\{y - x'\beta_0 - \psi_0(z'\mathbb{S}(\gamma_0))\} d(\hat{\mathbb{P}}_n - \mathbb{P}_n)(x,y,z) + o_{P_M}(1 + \sqrt{n}(\hat{\theta}^* - \theta_0)) \\
&\stackrel{d}{\to} N(0, \Pi),
\end{aligned}
$$

and the conclusion follows by Theorem 1.

It remains to prove (58). Similarly to Proposition 4 in BGH and (6.21) in GH, we can obtain the L^2-rate as

$$
\sup_\theta \int \{\hat{\psi}^*_{n\theta}(z'\mathbb{S}(\gamma)) - \psi_\theta(z'\mathbb{S}(\gamma))\}^2 d\hat{\mathbb{P}}_n(x,y,z) = O_{P_M}(n^{-2/3}).
$$

Define

$$T_n^* = \begin{bmatrix} \mathbb{I}_k & 0 \\ 0 & \mathbb{J}(\hat{\gamma})^{*\prime} \end{bmatrix}, \quad V_{I^*,n}^{x,z} = \begin{pmatrix} x - E[X|z'\mathbb{S}(\hat{\gamma}^*)] \\ z - E[Z|z'\mathbb{S}(\hat{\gamma}^*)] \end{pmatrix}, \quad V_{II^*,n}^{x,z} = \begin{pmatrix} E[X|z'\mathbb{S}(\hat{\gamma}^*)] - \bar{E}_{n,\hat{\theta}}[X|z'\mathbb{S}(\hat{\gamma}^*)] \\ E[Z|z'\mathbb{S}(\hat{\gamma}^*)] - \bar{E}_{n,\hat{\theta}}[Z|z'\mathbb{S}(\hat{\gamma}^*)] \end{pmatrix},$$

where $\bar{E}_{n,\theta}^*[W|u]$ is similarly defined as in (12). With similar arguments in Steps 1 and 2 in Section (A.2), we can show that

$$\phi_n^*(\hat{\theta}^*) = T_n^* \int V_{I^*,n}^{x,z}\{y - x'\hat{\beta}^* - \hat{\psi}^*_{n\hat{\theta}^*}(z'\mathbb{S}(\hat{\gamma}^*))\}d\hat{\mathbb{P}}_n(x,y,z) + o_{P_M}(n^{-1/2} + (\hat{\theta}^* - \theta_0)). \quad (59)$$

For the first term of (59),

$$T_n^* \int V_{I^*,n}^{x,z}\{y - x'\hat{\beta}^* - \hat{\psi}^*_{n\hat{\theta}^*}(z'\mathbb{S}(\hat{\gamma}^*))\}d\hat{\mathbb{P}}_n(x,y,z)$$

$$= T_n^* \int V_{I^*,n}^{x,z}\{y - x'\hat{\beta}^* - \hat{\psi}^*_{n\hat{\theta}^*}(z'\mathbb{S}(\hat{\gamma}^*))\}d(\hat{\mathbb{P}}_n - \mathbb{P}_n)(x,y,z)$$

$$+ T_n^* \int V_{I^*,n}^{x,z}\{y - x'\hat{\beta}^* - \hat{\psi}^*_{n\hat{\theta}^*}(z'\mathbb{S}(\hat{\gamma}^*))\}d\mathbb{P}_n(x,y,z)$$

$$:= T_n^*I^* + T_n^*II^*.$$

$T_n^*I^*$ is the bootstrap version of T_nI_b in (27). Therefore, with a similar arguments in Step 5 of Section (A.2), we have

$$T_n^*I^* = T_0 \int V_{x,z}\{y - x'\beta_0 - \psi_0(z'\mathbb{S}(\gamma_0)\}d(\hat{\mathbb{P}}_n - \mathbb{P}_n)(x,y,z) + o_{P_M}(n^{-1/2} + (\hat{\theta}^* - \theta_0)). \quad (60)$$

$T_n^*II^*$ is actually the first item of (14), T_nI, evaluated at $\hat{\theta}^*$. It can be decomposed as in (25). With similar argument from Step 3 to Step 6 in Section (A.2), we have

$$T_n^*II^* = -T_0 \int V_{x,z}V'_{x,z,\psi'}dP_0(x,z)T_0'(\hat{\theta}^* - \theta_0)$$

$$+ T_0 \int V_{x,z}\{y - x'\beta_0 - \psi_0(z'\mathbb{S}(\gamma_0)\}d(\mathbb{P}_n - P_0)(x,y,z) + o_P(n^{-1/2} + (\hat{\theta}^* - \theta_0))$$

$$= -B(\hat{\theta}^* - \theta_0) + T_0 \int V_{x,z}\{y - x'\beta_0 - \psi_0(z'\mathbb{S}(\gamma_0)\}d(\mathbb{P}_n - P_0)(x,y,z) \quad (61)$$

$$+ o_{P_M}(n^{-1/2} + (\hat{\theta}^* - \theta_0)),$$

where the last equality follows from the definition of B and the fact that any item of order $o_P(n^{-1/2} + (\hat{\theta}^* - \theta_0))$ will be of order $o_{P_M}(n^{-1/2} + (\hat{\theta}^* - \theta_0))$.

Combining (59), (60), and (61), we have (58).

References

[1] Balabdaoui, F., Durot, C. and Jankowski, H., (2019). Least squares estimation in the monotone single index model. *Bernoulli*, 25(4B), pp.3276-3310.

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

[3] Carroll, R. J., Fan, J., Gijbels, I. and M. P. Wand (1997) Generalized partially linear single-index models, *Journal of the American Statistical Association*, 92, 477-489.

[4] Cheng, G. (2009) Semiparametric additive isotonic regression, *Journal of Statistical Planning and Inference*, 139, 1980-1991.

[5] Groeneboom, P. and G. Jongbloed (2014) Nonparametric Estimation under Shape Constraints, *Cambridge University Press*.

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

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

[8] Groeneboom, P. (2018). The Lagrange approach in the monotone single index model. *arXiv preprint arXiv:1812.01380*.

[9] Huang, J. (2002) A note on estimating a partly linear model under monotonicity constraints, *Journal of Statistical Planning and Inference*, 107, 343-351.

[10] Ma, Y. and L. Zhu (2013) Doubly robust and efficient estimators for heteroscedastic partially linear single-index models allowing high dimensional covariates, *Journal of the Royal Statistical Society*, B, 75, 305-322.

[11] van der Vaart, A. W. (2000) *Asymptotic Statistics*, Cambridge university press.

[12] van der Vaart, A. W. and J. A. Wellner (1996) *Weak Convergence and Empirical Processes*, Springer.

[13] Wang, J.-L., Xue, L., Zhu, L. and Y. S. Chong (2010) Estimation for a partial-linear single-index model, *Annals of statistics*, 38, 246-274.

[14] Xia, Y. and W. Härdle (2006) Semi-parametric estimation of partially linear single-index models, *Journal of Multivariate Analysis*, 97, 1162-1184.

[15] Xia, Y., Tong, H. and W. Li (1999) On extended partially linear single-index models, *Biometrika*, 86, 831-842.

[16] Yu, K. (2014) On partial linear additive isotonic regression, *Journal of the Korean Statistical Society*, 43, 11–17.

[17] Yu, Y. and D. Ruppert (2002) Penalized spline estimation for partially linear single-index models, *Journal of the American Statistical Association*, 97, 1042-1054.