SCORE ESTIMATION OF MONOTONE PARTIALLY LINEAR INDEX MODEL

MENGSHAN XU AND TAISUKE OTSU

Abstract. This paper studies semiparametric estimation of partially linear single index models with a monotone link function. Our estimator is an extension of the score-type estimator developed by Balabdaoui, Groeneboom and Hendrickx (2018) for monotone single index models, 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. Furthermore, by introducing an additional smoothing to obtain the efficient score, we propose an asymptotically efficient estimator for the finite-dimensional components. 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 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 quite general. 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.

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

Financial support from the ERC Consolidator Grant (SNP 615882) is gratefully acknowledged (Otsu).

$\min_{\alpha,\beta}[\min_{\psi\in\mathcal{M}}\sum_{i=1}^{n}\{Y_i - X_i'\beta + \psi(Z_i'\alpha)\}^2]$, 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 (2018) (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) developed the score-type estimator for current status models.

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). 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.

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). Section 3 presents some simulation evidence to illustrate the finite sample performance of our estimators.

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 $\mathbb{S} : \mathbb{R}^{d-1} \to \mathcal{S}_{d-1}$ such that for each α in a neighborhood of α_0 on $\mathcal{S}_{d-1}$, there exists a unique $\gamma \in \mathbb{R}^{d-1}$ which satisfies $\alpha = \mathbb{S}(\gamma)$. 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

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

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

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

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

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

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[1]

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

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. 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.

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

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

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

[1] 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 $\bar{\zeta}(\theta_1)\bar{\zeta}(\theta_2) \leq 0$, where $\bar{\zeta}$ is the closure of the image of ζ (so contains its limit points). 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.

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 $R > 0$.*

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, $E[W|Z'\alpha = z'\alpha]$ 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 $z \in \mathcal{Z}$.*

A6: $\mathrm{Cov}[(\beta_0-\beta)'X+Z'(\mathbb{S}(\gamma_0)-\mathbb{S}(\gamma)),\ (\beta_0-\beta)'X+\psi_0(Z'\mathbb{S}(\gamma_0))|Z'\mathbb{S}(\gamma)]$ *is positive definite 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 $\theta \in \mathcal{B}(\theta_0, \delta_0)$. 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 the exponential family. A6 and A7 are to ensure the consistency and existence of limiting variances of the simple score and efficient score estimators, respectively.

Under these assumptions, the asymptotic properties of the 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 $\Pi = B^{-1}T_0\Sigma T'_0(B^{-1})'$ *and* $\Sigma = \mathrm{Var}(V_{X,Z}\epsilon)$.

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. Our result can be considered as an extension of BGH for the monotone PLSI model. 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 (8), (iv) replacing ϵ with the residuals based on our estimator, and (v) replacing the conditional expectations with kernel estimators.

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):

$$\xi_{nh}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} X_i \\ \mathbb{J}(\gamma)' Z_i \hat{\psi}'_{nh,\theta}(Z_i'\mathbb{S}(\gamma)) \end{pmatrix}\{Y_i - X_i'\beta - \hat{\psi}_{n\theta}(Z_i'\mathbb{S}(\gamma))\}, \tag{7}$$

where

$$\hat{\psi}'_{nh,\theta}(u) = \frac{1}{h}\int K\left(\frac{u-x}{h}\right) d\hat{\psi}_{n\theta}(x), \tag{8}$$

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 (7). 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 (8). The condition $h \asymp n^{-1/7}$ is also imposed in BGH.

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

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 $\Pi_E = B_E^{-1} T_0 \Sigma T_0' (B_E^{-1})'$ and $\Sigma = \mathrm{Var}(V_{X,Z,\psi'}\epsilon)$.

If the error term ϵ is homoskedastic, it holds $\Sigma = \sigma^2 \int V_{x,z,\psi'} V'_{x,z,\psi'} dP_0(x,z)$, where $\sigma^2 = \mathrm{Var}(\epsilon)$. 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.

3. Simulation

In this section, we conduct a small simulation study to illustrate the finite sample performance of the proposed estimators. Consider the following data generating process:

$$\begin{aligned} Y &= X\beta_0 + \psi_0(Z'\alpha_0) + \epsilon, \\ \psi_0(u) &= u^3, \quad \beta_0 = 1, \quad \alpha_0 = (1,1)/\sqrt{2} \approx (0.7071, 0.7071), \\ \epsilon &\sim N(0,1), \quad X \sim N(0,1), \quad Z \sim N(0, \mathbb{I}_2), \end{aligned}$$

where $\mathbb{I}_2$ is the 2×2 identity matrix. The sample sizes are $n = 100$, 500, and 1000. The number of Monte Carlo replications is 1000. Tables 1 presents the Monte Carlo averages $(\hat{\mu}_\beta, \hat{\mu}_{\alpha_1}, \hat{\mu}_{\alpha_2})$ and variances $(\hat{\sigma}^2_\beta, \hat{\sigma}^2_{\alpha_1}, \hat{\sigma}^2_{\alpha_2})$ (multiplied by n) of the estimates $(\hat{\beta}, \hat{\alpha}_1, \hat{\alpha}_2)$. 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 (7).

Table 1. Simulation results

Methods	n	$\hat{\mu}_\beta$	$\hat{\mu}_{\alpha_1}$	$\hat{\mu}_{\alpha_2}$	$\hat{\sigma}^2_\beta$	$\hat{\sigma}^2_{\alpha_1}$	$\hat{\sigma}^2_{\alpha_2}$
	100	1.0047	0.7081	0.7032	1.3172	0.2054	0.2072
SSE	500	0.9999	0.7074	0.7065	1.0946	0.1159	0.1163
	1000	1.0000	0.7072	0.7069	1.0725	0.0913	0.0914
	100	1.0062	0.7077	0.7052	1.3528	0.0872	0.0870
ESE	500	0.9998	0.7074	0.7067	1.1092	0.0398	0.0398
	1000	1.0000	0.7073	0.7069	1.0663	0.0315	0.0315

Table 1 shows 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. Overall, the simulation results are encouraging to support our estimation strategy.

Appendix 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 $E[\epsilon|Z] = 0$)

$$\psi_\theta(u) = E[\psi_0(Z'\alpha_0)|Z'\alpha = u] + (\beta_0 - \beta)'E[X|Z'\alpha = u]. \tag{9}$$

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 (9), 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

$$\sum_{j=1}^{k} m_j \sum_{i=n_j}^{n_{j+1}-1} \{y_i - x_i'\beta - \hat{\psi}_{n\theta}(u_i)\} = 0, \tag{10}$$

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

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

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 (10), it holds

$$\int \bar{E}_{n,\hat{\theta}}[W|z'\mathbb{S}(\gamma)]\{y - x'\hat{\beta} - \hat{\psi}_{n\hat{\theta}}(z'\mathbb{S}(\hat{\gamma}))\}d\mathbb{P}_n(x,y,z) = 0, \tag{11}$$

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

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

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 (12) and (13), with probability approaching one:

$$H_B(\varepsilon, \tilde{\mathcal{F}}_a, ||\cdot||_{B,P_0}) \le \frac{C_1}{\varepsilon}, \tag{12}$$

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 (36) below. Also, there exists a constant $C_3 > 0$ such that

$$||\tilde{f}||_{B,P_0} \leq C_3 (\log n) n^{-1/3}, \tag{13}$$

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

$$\begin{aligned}
P\{|II_{a,j}| > An^{-1/2}\} &= P\left\{|II_{a,j}| > An^{-1/2}, \sup_{\theta \in \mathcal{B}(\theta_0,\delta_0)} \sup_{z \in \mathcal{Z}} |\hat{\psi}_{n\theta}(z)| \leq K\right\} + \frac{\nu}{2} \\
&\leq P\left\{||\mathbb{G}_n||_{\mathcal{F}_a} > A, \sup_{\theta \in \mathcal{B}(\theta_0,\delta_0)} \sup_{z \in \mathcal{Z}} |\hat{\psi}_{n\theta}(z)| \leq K\right\} + \frac{\nu}{2} \\
&\leq \frac{E[||\mathbb{G}_n||_{\mathcal{F}_a} \quad | \sup_{\theta \in \mathcal{B}(\theta_0,\delta_0)} \sup_{z \in \mathcal{Z}} |\hat{\psi}_{n\theta}(z)| \leq K]}{A} + \frac{\nu}{2} \\
&= \frac{1}{AC_2 \log n} E[||\mathbb{G}_n||_{\tilde{\mathcal{F}}_a} \quad | \sup_{\theta \in \mathcal{B}(\theta_0,\delta_0)} \sup_{z \in \mathcal{Z}} |\hat{\psi}_{n\theta}(z)| \leq K] + \frac{\nu}{2} \\
&\lesssim \frac{1}{AC_2 \log n} J_n(\delta_n) \left(1 + \frac{J_n(\delta_n)}{\sqrt{n}\delta_n^2}\right) + \frac{\nu}{2} \\
&\lesssim \frac{\log n}{A} (\delta_n + 2B_1^{1/2} \delta_n^{1/2}) \left(1 + \frac{\delta_n + 2B_1^{1/2}\delta_n^{1/2}}{\sqrt{n}\delta_n^2}\right) + \frac{\nu}{2} \\
&\lesssim \frac{1}{A} (\log n)^{3/2} n^{-1/6} \left(1 + \frac{B_2}{(\log n)^{3/2}}\right) + \frac{\nu}{2} \\
&\lesssim \nu,
\end{aligned} \tag{14}$$

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 (36)), 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 (12) 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,

$$II_a = o_p(n^{-1/2}). \tag{15}$$

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

$$\frac{\partial}{\partial \alpha_j} E[\psi_0(Z'\alpha_0)|Z'\alpha = z'\alpha]\bigg|_{\alpha=\alpha_0} = \{z_j - E[Z_j|Z'\alpha = z'\alpha_0]\}\psi_0'(z'\alpha_0), \tag{16}$$

for $j = 1, \ldots, d$. Using an expansion around $\hat{\gamma} = \gamma_0$ with (16) and $E[\psi_0(Z'\mathbb{S}(\gamma_0))|z'\mathbb{S}(\gamma_0)] = \psi_0(z'\mathbb{S}(\gamma_0))$, we have

$$E[\psi_0(Z'\mathbb{S}(\gamma_0))|z'\mathbb{S}(\hat{\gamma})] = \psi_0(z'\mathbb{S}(\gamma_0)) + (\hat{\gamma} - \gamma_0)'\mathbb{J}(\hat{\gamma})'\{z - E[Z|z'\mathbb{S}(\gamma_0)]\}\psi_0'(z'\mathbb{S}(\gamma_0)) + o_p(\hat{\gamma} - \gamma_0). \tag{17}$$

Then we have

$$\begin{aligned}
II_b &= \int V_{II,n}^{x,z} \left\{ \begin{array}{c} (\beta_0 - \hat{\beta})'\{x - E[X|z'\mathbb{S}(\hat{\gamma})]\} \\ +\psi_0(z'\mathbb{S}(\gamma_0)) - E[\psi_0(Z'\alpha_0)|z'\mathbb{S}(\hat{\gamma})] \end{array} \right\} dP_0(x,z) \\
&= \int V_{II,n}^{x,z} \left\{ \begin{array}{c} (\beta_0 - \hat{\beta})'\{x - E[X|z'\mathbb{S}(\hat{\gamma})]\} \\ -(\hat{\gamma} - \gamma_0)'\mathbb{J}(\gamma_0)'\{z - E[Z|z'\mathbb{S}(\gamma_0)]\}\psi_0'(z'\mathbb{S}(\gamma_0)) + o_p(\hat{\gamma} - \gamma_0) \end{array} \right\} dP_0(x,z) \\
&= -\int V_{II,n}^{x,z} \left(\begin{array}{c} x - E[X|z'\mathbb{S}(\gamma_0)] \\ \mathbb{J}(\gamma_0)'\{z - E[Z|z'\mathbb{S}(\gamma_0)]\}\psi_0'(z'\mathbb{S}(\gamma_0)) \end{array} \right)' dP_0(x,z) \left(\begin{array}{c} \hat{\beta} - \beta_0 \\ \hat{\gamma} - \gamma_0 \end{array} \right) + o_p(\hat{\gamma} - \gamma_0) \\
&= o_p(\hat{\theta} - \theta_0), \qquad (18)
\end{aligned}$$

where the first equality follows from $E[\epsilon|X,Z] = 0$ and (9), the second equality follows from (17), 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

$$|E[W|Z'\mathbb{S}(\gamma) = u] - \bar{E}_{n,\theta}[W|Z'\mathbb{S}(\gamma) = u] \le C_0|\psi_\theta(u) - \hat{\psi}_{n\theta}(u)|, \qquad (19)$$

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

$$\begin{aligned}
||II_c|| &= || \int V_{II,n}^{x,z}\{\psi_{\hat{\theta}}(\mathbb{S}(\hat{\gamma})) - \hat{\psi}_{n\hat{\theta}}(z'\mathbb{S}(\hat{\gamma}))\} dP_0(x,z)|| \\
&\lesssim \int \{\psi_{\hat{\theta}}(\mathbb{S}(\hat{\gamma})) - \hat{\psi}_{n\hat{\theta}}(z'\mathbb{S}(\hat{\gamma}))\}^2 dP_0(z) \\
&= O_p((\log)^2 n^{-2/3}) = o_p(n^{-1/2}), \qquad (20)
\end{aligned}$$

uniformly in $\theta \in \mathcal{B}(\theta_0, \delta_0)$, where the second equality follows from Proposition 4 in BGH. Combining (15), (18), and (20), we conclude that

$$II = o_p(n^{-1/2}) + o_p(\hat{\theta} - \theta_0). \qquad (21)$$

Step 3: Decompose I. The term I can be decomposed as

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

In the following steps, we show that

$$\begin{aligned}
T_n I_a &= -T_0 \int V_{x,z} V_{x,z\psi'}' dP_0(x,z) T_0'(\hat{\theta} - \theta_0) + o_p(\hat{\theta} - \theta_0), \qquad (22) \\
T_n I_b &= T_0 \int V_{x,z}\{y - x'\beta_0 - \psi_0(z'\mathbb{S}(\gamma_0)\} d(\mathbb{P}_n - P_0)(x,y,z) \\
&\quad + o_p(\hat{\theta} - \theta_0) + o_p(n^{-1/2}), \qquad (23) \\
I_c &= o_p(n^{-1/2}). \qquad (24)
\end{aligned}$$

Step 4: Show (22).

$$
\begin{aligned}
I_a &= \int V_{I,n}^{x,z}\left\{\begin{array}{c}(\beta_0-\hat\beta)'\{x-E[X|z'\mathbb{S}(\hat\gamma)]\}\\ +\psi_0(z'\mathbb{S}(\gamma_0))-E[\psi_0(Z'\alpha_0)|z'\mathbb{S}(\hat\gamma)]\end{array}\right\}dP_0(x,z)\\
&= \int V_{I,n}^{x,z}\left\{\begin{array}{c}(\beta_0-\hat\beta)\{x-E[X|z'\mathbb{S}(\hat\gamma)]\}\\ -(\hat\gamma-\gamma_0)'\mathbb{J}(\gamma_0)'\{z-E[Z|z'\mathbb{S}(\gamma_0)]\}\psi_0'(z'\mathbb{S}(\gamma_0))+o_p(\hat\gamma-\gamma_0)\end{array}\right\}dP_0(x,z)\\
&= -\int V_{x,z}V_{x,z,\psi'}'dP_0(x,z)T_0'\begin{pmatrix}\hat\beta-\beta_0\\ \hat\gamma-\gamma_0\end{pmatrix}+o_p(\hat\gamma-\gamma_0), \qquad (25)
\end{aligned}
$$

where the the first equality follows from $E[\epsilon|X,Z]=0$ and (9), and some rearrangement, the second equality follows from (17), 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, (22) follows by

$$T_n-T_0=O_p(\hat\gamma-\gamma_0). \qquad (26)$$

Step 5: Show (23). Decompose

$$
\begin{aligned}
T_nI_b &= T_n\int V_{I,n}^{x,z}\{y-x'\hat\beta-\psi_{\hat\theta}(z'\mathbb{S}(\hat\gamma)\}d(\mathbb{P}_n-P_0)(x,y,z)\\
&= (T_n-T_0)\int V_{I,n}^{x,z}\{y-x'\hat\beta-\psi_{\hat\theta}(z'\mathbb{S}(\hat\gamma)\}d(\mathbb{P}_n-P_0)(x,y,z)\\
&\quad +T_0\int V_{I,n}^{x,z}\{x'\beta_0-x'\hat\beta+\psi_0(z'\mathbb{S}(\gamma_0))-\psi_{\hat\theta}(z'\mathbb{S}(\hat\gamma))\}d(\mathbb{P}_n-P_0)(x,y,z)\\
&\quad +T_0\int(V_{I,n}^{x,z}-V_{x,z})\{y-x'\beta_0-\psi_0(z'\mathbb{S}(\gamma_0)\}d(\mathbb{P}_n-P_0)(x,y,z)\\
&\quad +T_0\int V_{x,z}\{y-x'\beta_0-\psi_0(z'\mathbb{S}(\gamma_0)\}d(\mathbb{P}_n-P_0)(x,y,z)\\
&:= (T_n-T_0)I_{b1}+T_0I_{b2}+T_0I_{b3}\\
&\quad +T_0\int V_{x,z}\{y-x'\beta_0-\psi_0(z'\mathbb{S}(\gamma_0)\}d(\mathbb{P}_n-P_0)(x,y,z).
\end{aligned}
$$

First, consider I_{b1}. Note that Lemma 13 BGH-supp and Lemma 2 imply the following (27) and (28):

$$H_B(\varepsilon,\mathcal{F}_{b1},||\cdot||_{B,P_0})\le\frac{C_1}{\varepsilon}, \qquad (27)$$

for some $C_1>0$, where $\mathcal{F}_{b1}$ is defined in (38). Also, there exists a constant $C_2>0$ such that

$$||f||_{B,P_0}\le C_2, \qquad (28)$$

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 (27), (28), and Equation (.2) in BGH. Thus, we have

$$I_{b1} = O_p(n^{-1/2}). \tag{29}$$

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

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

$$\lesssim \frac{1}{A}C_4\eta^{1/2}\left(1 + \frac{C_5(1+\eta^{1/2})}{\sqrt{n}(C'\eta)^{3/2}}C_3\right) + \frac{\nu}{2}, \tag{30}$$

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 (40), 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

$$I_{b2} = o_p(n^{-1/2}). \tag{31}$$

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 (30).) Thus, we have

$$I_{b3} = o_p(n^{-1/2}). \tag{32}$$

Combining (29), (31), and (32) with (26), we obtain (23).

Step 6: Show (24). 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

$$I_{c1} = E\left[E\left[\left(\begin{array}{c} X - E[X|Z'\mathbb{S}(\hat{\gamma})] \\ Z - E[Z|Z'\mathbb{S}(\hat{\gamma})] \end{array}\right)\middle| Z'\mathbb{S}(\hat{\gamma})\right]\{\psi_{\hat{\theta}}(Z'\mathbb{S}(\hat{\gamma})) - \hat{\psi}_{n\hat{\theta}}(Z'\mathbb{S}(\hat{\gamma}))\}\right] = 0. \tag{33}$$

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 (30), and the second inequality follows from the definition of η_n. Thus, we have $I_{c2} = o_p(n^{-1/2})$, and obtain (24).

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

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

A.3. **Lemmas.** In this subsection, we use the following notations:

$$\begin{aligned} \mathcal{M}_{RK} &= \{\text{monotone non-decreasing functions on } [-R, R] \text{ and bounded by } K\}, \\ \mathcal{G}_{RK} &= \{g : g(z) = \psi_\theta(\alpha' z), z \in \mathcal{Z}, (\psi, \theta) \in \mathcal{M}_{RK} \times \mathcal{B}(\theta_0, \delta_0)\}, \\ \mathcal{D}_{RKv} &= \{d : d(z) = g_1(z) - g_2(z),\ (g_1, g_2) \in \mathcal{G}^2_{RK},\ ||d(z)||_{P_0} \le v\}, \\ \mathcal{H}_{RKv} &= \{h : h(\tilde{y}, z) = \tilde{y} d_1(z) - d_2(z),\ (d_1, d_2) \in \mathcal{D}^2_{RKv},\ (\tilde{y}, z) \in \mathbb{R} \times \mathcal{Z}\}. \end{aligned} \tag{34}$$

A.3.1. *Lemma for II_a.* Let W_j be the j-th component of X or Z. Then decompose

$$\begin{aligned} & \{E[W_j | z'\mathbb{S}(\hat{\gamma})] - \bar{E}_{n,\hat{\theta}}[W_j | z'\mathbb{S}(\hat{\gamma})]\}\{y - x\hat{\beta} - \hat{\psi}_{n\hat{\theta}}(z'\mathbb{S}(\hat{\gamma}))\} \\ = & \{E[W_j | z'\mathbb{S}(\hat{\gamma})] - \bar{E}_{n,\hat{\theta}}[W_j | z'\mathbb{S}(\hat{\gamma})]\}\{y - x\hat{\beta}\} \\ & - \{E[W_j | z'\mathbb{S}(\hat{\gamma})] - \bar{E}_{n,\hat{\theta}}[W_j | z'\mathbb{S}(\hat{\gamma})]\}\hat{\psi}_{n\hat{\theta}}(z'\mathbb{S}(\hat{\gamma})) \\ := & \ d_1(z)\{y - x\hat{\beta}\} - d_2(z). \end{aligned} \tag{35}$$

Let

$$\mathcal{F}_a = \big\{f : f(x,y,z) = d_1(z)\{y - x\hat{\beta}\} - d_2(z),\ (x,y,z) \in \mathcal{X} \times \mathbb{R} \times \mathcal{Z}\big\}, \tag{36}$$

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 (19), $||E[W_j|z'\mathbb{S}(\hat{\gamma})] - \bar{E}_{n,\hat{\theta}}[W_j|z'\mathbb{S}(\hat{\gamma})]||_2 \le C_1(\log n)n^{-1/3}$ 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

$$d_1(\cdot) \in \mathcal{D}_{RC_0v_1}, \tag{37}$$

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 (34) yields the conclusion. □

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

$$\mathcal{F}_{b1} = \big\{f : f(w_j, y, z) = \{w_j - E[W_j|z'\mathbb{S}(\hat{\gamma})]\}\{\tilde{y} - \psi_{\hat{\theta}}(z'\mathbb{S}(\hat{\gamma})\},\ (w_j, y, z) \in \mathcal{W}_j \times \mathbb{R} \times \mathcal{Z}\big\}, \tag{38}$$

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 (9), $\psi_{\hat{\theta}}$ 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

$$\begin{aligned} & \{w_j - E[W_j|z'\mathbb{S}(\hat{\gamma})]\}\{y - x'\hat{\beta} - \psi_{\hat{\theta}}(z'\mathbb{S}(\hat{\gamma})\} \\ = \ & w_j\{y - x'\hat{\beta} - \psi_{\hat{\theta}}(z'\mathbb{S}(\hat{\gamma})\} + d_1(z'\mathbb{S}(\hat{\gamma}))(y - x'\hat{\beta}) - d_2(z'\mathbb{S}(\hat{\gamma})). \end{aligned} \tag{39}$$

By b) and c), we have

$$d_1(\cdot) \in \mathcal{D}_{RC_0v_1},$$

for C_0 defined in (37), 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

$$d_2(\cdot) \in \mathcal{D}_{RC_1v_2},$$

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 (34), the second and third terms in (39) satisfy

$$d_1(z'\mathbb{S}(\hat{\gamma}))(y - x'\hat{\beta}) - d_2(z'\mathbb{S}(\hat{\gamma})) \in \mathcal{H}_{RC_1v_1}.$$

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. □

A.3.3. *Lemma for I_{b2}.* Let

$$\mathcal{F}_{b2} = \big\{f : f(w_j, x, z) = \{w_j - E[W_j|z'\mathbb{S}(\hat{\gamma})]\}\{x'\beta_0 - x'\hat{\beta} + \psi_0(z'\mathbb{S}(\gamma_0) - \psi_{\hat{\theta}}(z'\mathbb{S}(\hat{\gamma})\},\ (w_j, x, z) \in \mathcal{W}_j \times \mathcal{X} \times \mathcal{Z}\big\}, \tag{40}$$

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 $\mathcal{F}_{b2}$. □

A.3.4. *Lemma for I_{c2}.* Let

$$\mathcal{F}_{c2} = \big\{f : f(w_j, z) = \{w_j - E[W_j|z'\mathbb{S}(\hat{\gamma})]\}\{\psi_{\hat{\theta}}(z'\mathbb{S}(\hat{\gamma})) - \hat{\psi}_{n\hat{\theta}}(z'\mathbb{S}(\hat{\gamma}))\},\ (w_j, z) \in \mathcal{W}_j \times \mathcal{Z}\big\}, \tag{41}$$

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 (9), $\psi_{\hat{\theta}}$ 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. □

Appendix 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 $u_{n_{j+1}} = \max_i u_{n_i}$)

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

Similar to (19), we have for each $\theta \in \mathcal{B}(\theta_0, \delta_0)$

$$|E[Z|Z'\mathbb{S}(\gamma) = u]\psi'_\theta(u) - \bar{\rho}_{n,\theta}(Z|u)| \le C_0|\psi_\theta(u) - \hat{\psi}_{n\theta}(u)|. \tag{42}$$

Similar to (11), 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

$$
\begin{aligned}
\xi_{nh}(\tilde{\theta}) &= T_n \int V^{x,z}_{I,nh,\psi'}\{y - x'\tilde{\beta} - \hat{\psi}_{n\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma})\}d\mathbb{P}_n(x,y,z) + T_n \int V^{x,z}_{II,n}\{y - x'\tilde{\beta} - \hat{\psi}_{n\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma})\}d\mathbb{P}_n(x,y,z)\\
&:= T_n(I^E + II^E),
\end{aligned} \tag{43}
$$

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

$$
\begin{aligned}
V_{I,nh,\psi'}^{x,z} &= \begin{pmatrix} x - E[X|z'\mathbb{S}(\tilde{\gamma})] \\ z\hat{\psi}'_{nh,\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma})) - E[Z|z'\mathbb{S}(\tilde{\gamma})]\psi'_{\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma})) \end{pmatrix}, \quad V_{I,n,\psi'}^{x,z} = \begin{pmatrix} x - E[X|z'\mathbb{S}(\tilde{\gamma})] \\ [z - E[Z|z'\mathbb{S}(\tilde{\gamma}))]\psi'_{\tilde{\theta}}(z'\mathbb{S}(\tilde{\gamma})) \end{pmatrix}, \\
V_{x,z,\psi'} &= \begin{pmatrix} x - E[X|z'\mathbb{S}(\gamma_0)] \\ [z - E[Z|z'\mathbb{S}(\gamma_0))]\psi'_0(z'S(\gamma_0)) \end{pmatrix}, \\
V_{II,n}^{x,z} &= \begin{pmatrix} E[X|z'\mathbb{S}(\tilde{\gamma})] - \bar{\rho}_{n,\tilde{\theta}}(X|z'\mathbb{S}(\tilde{\gamma})) \\ E[Z|z'\mathbb{S}(\tilde{\gamma})]\psi'_{\theta}(z'\mathbb{S}(\tilde{\gamma}) - \bar{\rho}_{n,\tilde{\theta}}(Z|z'\mathbb{S}(\tilde{\gamma})) \end{pmatrix}.
\end{aligned} \tag{44}
$$

Note: T_n and $V_{II,n}^{x,z}$ 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 (19) and (42), 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

$$
T_n I_a^E = -T_0 \int V_{x,z,\psi'} V'_{x,z\psi'} dP_0(x,z) T'_0(\tilde{\theta} - \theta_0) + o_p(\tilde{\theta} - \theta_0), \tag{45}
$$

$$
\begin{aligned}
T_n I_b^E &= T_0 \int V_{x,z,\psi'}\{y - x'\beta_0 - \psi_0(z'\mathbb{S}(\gamma_0)\}d(\mathbb{P}_n - P_0)(x,y,z) \\
&\quad + o_p(\tilde{\theta} - \theta_0) + o_p(n^{-1/2}),
\end{aligned} \tag{46}
$$

$$
I_c^E = o_p(n^{-1/2}). \tag{47}
$$

Step 4: Show (45). Decompose

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

By a similar argument as in (25), we have

$$I_{a1}^E = -\Big\{\int V_{x,z,\psi'}V'_{x,z,\psi'}dP_0(x,z)\Big\}T'_0(\tilde{\theta} - \theta_0) + o_p(\tilde{\theta} - \theta_0). \tag{48}$$

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 (45).

Step 5: Show (46). Decompose

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

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

$$T_n I_{b1}^E = T_0 \int V_{x,z,\psi'}\{y - x'\beta_0 - \psi_0(z'\mathbb{S}(\gamma_0)\}d(\mathbb{P}_n - P_0)(x,y,z) + o_p(\tilde{\theta} - \theta_0) + o_p(n^{-1/2}). \tag{49}$$

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 (46) is obtained.

Step 6: Show (47). Decompose

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

For I_{c1}^E, note that

$$\begin{aligned}
I_{c1}^E &= \int V_{I,n,\psi'}^{x,z}\{\psi_{\tilde\theta}(z'\mathbb{S}(\tilde\gamma)) - \hat\psi_{n\tilde\theta}(z'\mathbb{S}(\tilde\gamma))\}dP_0(x,y,z)\\
&\quad + \int \begin{pmatrix} 0 \\ z\{\hat\psi'_{nh,\tilde\theta}(z'\mathbb{S}(\tilde\gamma)) - \psi'_{\tilde\theta}(z'\mathbb{S}(\tilde\gamma))\} \end{pmatrix} \{\psi_{\tilde\theta}(z'\mathbb{S}(\tilde\gamma)) - \hat\psi_{n\tilde\theta}(z'\mathbb{S}(\tilde\gamma))\}dP_0(x,y,z)\\
&= \int \begin{pmatrix} 0 \\ E[Z|u]\left\{\frac{1}{h}\int K\left(\frac{u-x}{h}\right)d\hat\psi_{n\tilde\theta}(x) - \psi'_{\tilde\theta}(u)\right\} \end{pmatrix} \{\psi_{\tilde\theta}(u) - \hat\psi_{n\tilde\theta}(u)\}dP_0(u), \quad (50)
\end{aligned}$$

where the last equality follows from a similar argument in (33), 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

$$\begin{aligned}
&\frac{1}{h}\int K\left(\frac{u-x}{h}\right)d\hat\psi_{n\tilde\theta}(x) - \psi'_{\tilde\theta}(u)\\
&= \frac{1}{h}\int K\left(\frac{u-x}{h}\right)d(\hat\psi_{n\tilde\theta}(x) - \psi_{\tilde\theta}(x)) + \frac{1}{h}\int K\left(\frac{u-x}{h}\right)d\psi_{\tilde\theta}(x) - \psi'_{\tilde\theta}(u)\\
&= -\frac{1}{h^2}\int K'\left(\frac{u-x}{h}\right)(\hat\psi_{n\tilde\theta}(x) - \psi_{\tilde\theta}(x))dx + \frac{1}{h}\int K\left(\frac{u-x}{h}\right)d\psi_{\tilde\theta}(x) - \psi'_{\tilde\theta}(u), \quad (51)
\end{aligned}$$

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 (51) into (50), the Cauchy-Schwarz inequality and A9 imply

$$I_{c1}^E = O_p((\log n)^2 n^{-2/3}) \cdot O_p(n^{1/7}) + O_p((\log n)n^{-1/3}) \cdot O_p(n^{-2/7}) = o_p(n^{-1/2}). \quad (52)$$

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 (52), we obtain (47).

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

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

References

[1] Balabdaoui, F., Groeneboom, P. and K. Hendrickx (2018) Score estimation in the monotone single index model, forthcoming in *Scandinavian Journal of Statistics*.

[2] 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.

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

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

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

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

[7] 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.

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

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

[10] 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.

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

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

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

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

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

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

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